ORIGINAL RESEARCH

Whole-body MRI versus ^{18}F-FDG PET/CT for pretherapeutic assessment and staging of lymphoma: a meta-analysis

Danyang Wang[1]
Yanlei Huo[1]
Suyun Chen[1]
Hui Wang[1]
Yingli Ding[2]
Xiaochun Zhu[3]
Chao Ma[1,4]

[1]Department of Nuclear Medicine, Affiliated XinHua Hospital of Shanghai Jiaotong University School of Medicine, Shanghai, China; [2]Department of Nuclear Medicine, Affiliated Third People's Hospital of Shanghai Jiaotong University School of Medicine, Shanghai, China; [3]Department of Nuclear Medicine, Affiliated Ninth People's Hospital of Shanghai Jiaotong University School of Medicine, Shanghai, China; [4]Department of Nuclear Medicine, Tenth People's Hospital of Tongji University, Shanghai, China

Purpose: ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT) is the reference standard in staging of ^{18}F-FDG-avid lymphomas; however, there is no recommended functional imaging modality for indolent lymphomas. Therefore, we aimed to compare the performance of whole-body magnetic resonance imaging (WB-MRI) with that of ^{18}F-FDG PET/CT for lesion detection and initial staging in patients with aggressive or indolent lymphoma.

Materials and methods: We searched the MEDLINE, EMBASE, and CENTRAL databases for studies that compared WB-MRI with ^{18}F-FDG PET/CT for lymphoma staging or lesion detection. The methodological quality of the studies was assessed using version 2 of the "Quality Assessment of Diagnostic Accuracy Studies" tool. The pooled staging accuracy (μ) of WB-MRI and ^{18}F-FDG PET/CT for initial staging and for assessing possible heterogeneity (χ^2) across studies were calculated using commercially available software.

Results: Eight studies comprising 338 patients were included. In terms of staging, the meta-analytic staging accuracies of WB-MRI and ^{18}F-FDG PET/CT for Hodgkin lymphoma and aggressive non-Hodgkin lymphoma (NHL) were 98% (95% CI, 94%–100%) and 98% (95% CI, 94%–100%), respectively. The pooled staging accuracy of ^{18}F-FDG PET/CT dropped to 87% (95% CI, 72%–97%) for staging in patients with indolent lymphoma, whereas that of WB-MRI remained 96% (95% CI, 91%–100%). Subgroup analysis indicated an even lower staging accuracy of ^{18}F-FDG PET/CT for staging of less FDG-avid indolent NHLs (60%; 95% CI, 23%–92%), in contrast to the superior performance of WB-MRI (98%; 95% CI, 88%–100%).

Conclusion: WB-MRI is a promising radiation-free imaging technique that may serve as a viable alternative to ^{18}F-FDG PET/CT for staging of ^{18}FDG-avid lymphomas, where ^{18}F-FDG PET/CT remains the standard of care. Additionally, WB-MRI seems a less histology-dependent functional imaging test than ^{18}F-FDG PET/CT and may be the imaging test of choice for staging of indolent NHLs with low ^{18}F-FDG avidity.

Keywords: lymphoma, meta-analysis, whole-body diffusion-weighted MRI, ^{18}F-FDG PET/CT, staging

Introduction

Malignant lymphoma is the most common primary hematopoietic malignancy, which is also considered as one of the most curable forms of cancer. The two main categories of lymphomas are Hodgkin lymphoma (HL) and non-Hodgkin lymphomas (NHLs). Accurate staging at the initial diagnosis of lymphoma is crucial to appropriate treatment planning and prognostication. The current staging system for malignant lymphoma, the Cotswolds modification of the Ann Arbor staging system, is applied to both HL and

Correspondence: Chao Ma
Department of Nuclear Medicine, Affiliated XinHua Hospital of Shanghai Jiaotong University School of Medicine, 1665 Kongjiang Road, Yangpu District, Shanghai 200092, China
Tel +86 21 2507 8593
Email mc_7419@hotmail.com

NHL. Staging according to the Ann Arbor system is based on the number of lesions and the extent of nodal or extra-nodal involvement, whereas the Cotswolds modification put forward tumor burden as a complementary factor. Therefore, several imaging modalities have been used for noninvasive staging of lymphoma, including computed tomography (CT) and ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT). Compared with CT, ^{18}F-FDG PET/CT is more sensitive in the detection of nodal and extra-nodal lymphomatous involvement, for it can provide both meta-bolic and morphologic information.[1] However, the diagnostic accuracy of ^{18}F-FDG PET/CT depends on ^{18}F-FDG avidities of tumor lesions which vary for different histologic subtypes.[2,3] ^{18}F-FDG PET/CT has been proven to be highly sensitive and regarded as the standard method in staging of HL, aggressive NHLs, and follicular lymphoma (FL),[1] but its diagnostic accu-racy is unstable in staging of indolent lymphomas.[4–7] Besides, the need for initial staging, restaging, and therapy monitoring requires multiple ^{18}F-FDG PET/CT examinations for a single patient, which lead to concerns of radiation-induced secondary malignancies, especially in younger patients.[8–10]

In recent years, whole-body magnetic resonance imaging (WB-MRI) has emerged as a radiation-free alternative to ^{18}F-FDG PET/CT for obtaining both anatomic and functional images. Conventional WB-MRI comprises a T1-weighted sequence, a T2-weighted sequence, and fat-suppression tech-niques such as short-tau inversion recovery. This sequence has been reported to be particularly effective for detecting bone marrow metastases, but its use is restricted by its limited ability to detect small lesions in the abdomen and pelvis.[11–13] Diffusion-weighted imaging is being increasingly applied in the evaluation of patients with malignant disease.[14] Through its ability to measure the Brownian motion of water molecules in tissue, diffusion-weighted imaging can be used to detect tumors with restricted diffusion and provide superior visual-ization of lymph nodes.[15] Even though ^{18}F-FDG PET/CT is the current reference standard in the staging of HL and high-grade lymphomas, there is hitherto no recommended functional imaging modality for staging of some indolent histologic subtypes. Therefore, in order to determine the diagnostic value of WB-MRI for initial lymphoma staging, especially for indolent subtypes, this study systematically compared the diagnostic accuracy of WB-MRI with that of ^{18}F-FDG PET/CT in staging of aggressive and indolent lymphomas.

Materials and methods
Literature search
We searched the MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials (CENTRAL) databases for studies comparing the diagnostic accuracy of WB-MRI and ^{18}F-FDG PET/CT in lymphoma. The database search was without language or date limits. The search terms were designed to capture the concepts of lymphoma, WB-MRI, and diffusion-weighted imaging. The search was performed over the span of 2 months and terminated in June 2016. The detailed electronic search strategies applied to each database are shown in Table S1.

Study selection
Studies evaluating the diagnostic accuracy of WB-MRI using ^{18}F-FDG PET/CT as the comparator test in patients with histologically proven lymphoma were included. Studies were excluded if they 1) lacked a diffusion-weighted sequence, 2) included previously reported cases, 3) reported only on the detection of bone marrow involvement, 4) focused only on restaging or response assessment, 5) used ^{18}F-FDG PET/CT as the only reference standard, 6) had a sample size smaller than 10, or 7) included patients with other malignancies or hematologic disorders.

Two reviewers (YLD and YLH) separately screened the titles and abstract of every record retrieved, and the full texts were obtained for potentially eligible studies. YLD and YLH then independently applied the inclusion and exclusion crite-ria to the full text and together resolved any disagreement.

Methodological quality assessment
The methodological quality of the studies was assessed using version 2 of the "Quality Assessment of Diagnostic Accuracy Studies" tool. The tool comprises four domains: patient selec-tion, index test, reference standard, and flow and timing.[16] Two reviewers (DYW and XCZ) independently applied the signaling questions for each domain to the content of the studies and reached a consensus on adding or omitting certain signaling questions to develop a guide specifically for quality assessment. In accord with the guide, each domain was assessed in terms of risk of bias, and the first three domains were assessed in terms of concerns about applicability. The bias risk or applicability concerns were described as low, high, or unclear.

Reference standard
The pathologic diagnosis was regarded as the most reliable evidence of tumor involvement. However, when patho-logic evidence was inadequate for disease assessment, the reference standard of staging was based on the integrated interpretation of ^{18}F-FDG PET/CT and WB-MRI images, with the confirmation of clinical and imaging follow-up or bone marrow biopsy.

Staging was performed according to the Cotswolds modification of the Ann Arbor system.

Data extraction and statistical analysis

Data of all included studies were independently extracted by two reviewers (DYW and YLH) and any disagreement was resolved through discussion. The data extracted from the articles included: authors, date, study type, population description, sample size and characteristics, ^{18}F-FDG PET/CT and WB-MRI protocols, image interpretation methods, reference standards, sensitivities and specificities of ^{18}F-FDG PET/CT or WB-MRI for lesion detection, stages according to ^{18}F-FDG PET/CT or WB-MRI findings, final stages according to histopathology or imaging follow-ups.

The agreements between ^{18}F-FDG PET/CT and WB-MRI for lesion detection were calculated with Cohen's kappa statistic, defined as poor ($\kappa \leq 0.4$), moderate ($0.4 < \kappa \leq 0.6$), good ($0.6 < \kappa \leq 0.8$), and excellent ($0.8 < \kappa \leq 1$) agreement. To access the diagnostic accuracy of initial staging, we split the study populations into two groups according to pathologic subtypes, including HL or aggressive NHL groups and indolent NHL groups. The staging accuracy estimates were defined as the proportion of the patients who were correctly staged by ^{18}F-FDG PET/CT or WB-MRI in each group. In order to stabilize the variance, we calculated the staging accuracies with Freeman–Tukey double arcsine transformation[17] before calculating the pooled estimates. The 95% CI of these estimates were calculated in accordance with the Wilson score CI method.[18] Statistical heterogeneity was assessed by visually inspecting the forest plots and using the χ^2-test (with a p-value of 0.10 indicating statistical significance). The I^2 statistic was used to measure the degree of inconsistency across studies, with I^2 values of 25%, 50%, and 75% representing low, moderate, and high substantial heterogeneity.[19,20] We used random-effects model to calculate meta-analytic estimates (μ) because moderately high inconsistency was found in one group ($I^2 = 49.8\%$). Publication bias was weighed using Egger's test, with a p-value of 0.05 indicating statistical significance.[21] All data were analyzed using Stata (version 14.0).

Results
Study selection

We identified a total of 3,453 records through database searches (845 articles from MEDLINE, 2,598 from EMBASE, and 10 from CENTRAL). Figure 1 shows the flow of studies through the screening and selection process. After primary screening of the titles and abstracts and removal of duplicates, 69 articles were selected for full-text analysis. Two of these articles did not provide sufficient information: one did not provide the histologic subtype of each patient;[22] the other had incomplete follow-up data and we could not determine the standard stage.[23] Of the remaining 67 articles, 8 met our inclusion criteria. Table 1 summarizes the main features of the included studies.

Methodological quality of studies

Figure 2 presents our risk of bias and applicability concerns ratings for each study, and Figure 3 summarizes the overall methodological quality of the studies.

In the participant-selection domain, the risk of bias in five studies was judged to be unclear because it was not clear whether a consecutive sample of patients had been enrolled.[25,30] In the index-test domain, all studies were considered to be at low risk of bias. In the reference-standard domain, the study by Ferrari et al[5] was at high risk of bias because the reference standard had not been clearly described. We considered the remaining eight studies to be at low risk of bias. In the flow-and-timing domain, all studies were considered to be at low risk of bias.

In our assessment of applicability concerns, we were highly concerned that in one study[5] the target condition as defined by the reference standard did not match the review question.

Agreement between WB-MRI and ^{18}F-FDG PET/CT for lesion detection

Five studies reported the results on the comparison between WB-MRI and ^{18}F-FDG PET/CT on a per-lesion basis.[5,24–26,29] Among them, three studies analyzed the agreement between WB-MRI and ^{18}F-FDG PET/CT for the detection of both nodal and extra-nodal lesions, and two studies only provided the information of nodal regions. All five studies reported excellent agreement between the two imaging modalities (Table 2).

Separated staging accuracy of WB-MRI and ^{18}F-FDG PET/CT for initial lymphoma staging

Of the eight studies reported on the performance of WB-MRI and ^{18}F-FDG PET/CT for initial lymphoma staging, two included data for HL and aggressive NHL patients,[26,29] and the other six included mixed histologic subtypes of aggressive and indolent lymphomas.[5,24,25,27,28,30] Table 3 summarized the staging accuracies of WB-MRI and
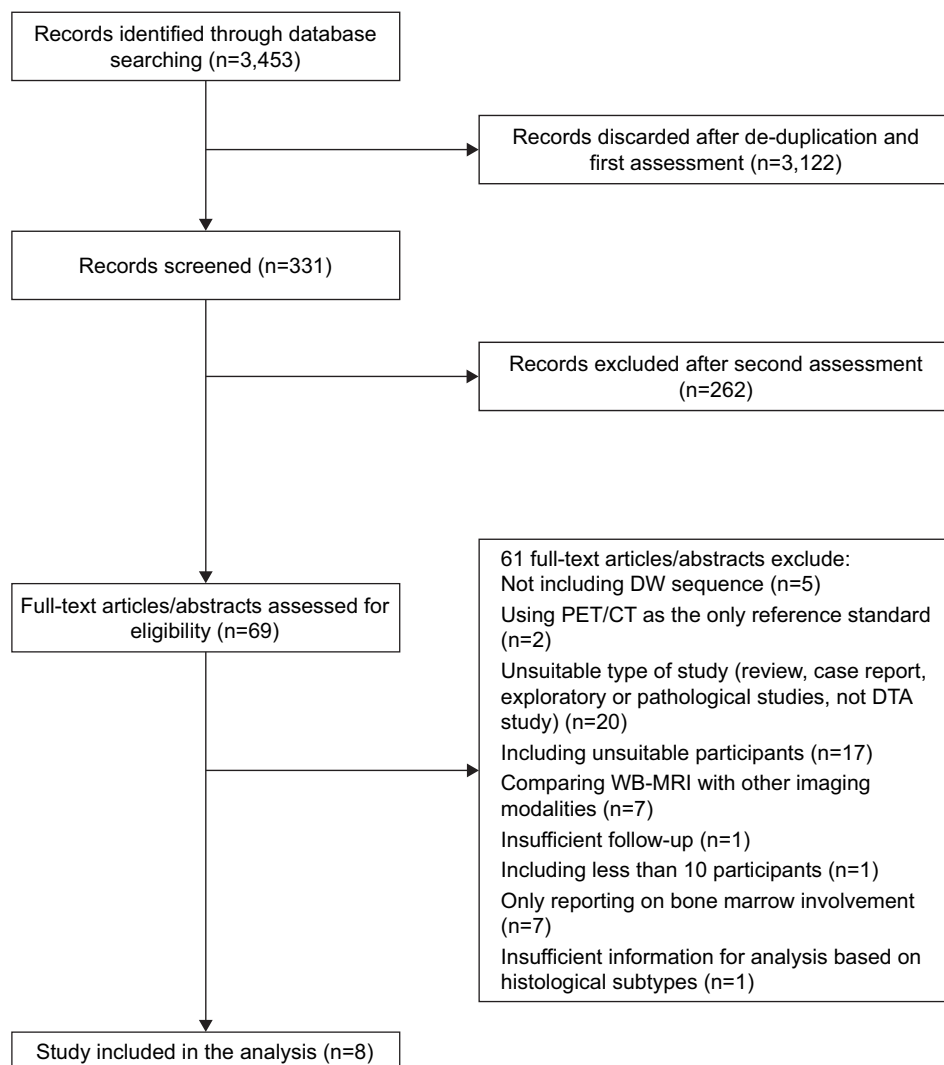
```
┌─────────────────────────────────┐
│ Records identified through       │
│ database searching (n=3,453)     │
└─────────────────────────────────┘
              │
              │────────────────────►┌─────────────────────────────────┐
              │                     │ Records discarded after          │
              │                     │ de-duplication and first         │
              │                     │ assessment (n=3,122)             │
              ▼                     └─────────────────────────────────┘
┌─────────────────────────────────┐
│ Records screened (n=331)         │
└─────────────────────────────────┘
              │
              │────────────────────►┌─────────────────────────────────┐
              │                     │ Records excluded after second    │
              │                     │ assessment (n=262)               │
              ▼                     └─────────────────────────────────┘
┌─────────────────────────────────┐
│ Full-text articles/abstracts     │
│ assessed for eligibility (n=69)  │
└─────────────────────────────────┘
```

61 full-text articles/abstracts exclude:
Not including DW sequence (n=5)
Using PET/CT as the only reference standard (n=2)
Unsuitable type of study (review, case report, exploratory or pathological studies, not DTA study) (n=20)
Including unsuitable participants (n=17)
Comparing WB-MRI with other imaging modalities (n=7)
Insufficient follow-up (n=1)
Including less than 10 participants (n=1)
Only reporting on bone marrow involvement (n=7)
Insufficient information for analysis based on histological subtypes (n=1)

```
┌─────────────────────────────────┐
│ Study included in the analysis   │
│ (n=8)                            │
└─────────────────────────────────┘
```

Figure 1 Flow diagram of study selection.
Abbreviations: DTA, diagnostic test accuracy; DW, diffusion-weighted; WB-MRI, whole-body magnetic resonance imaging.

^{18}F-FDG PET/CT in both groups. In staging of HL and aggressive NHLs, both WB-MRI and ^{18}F-FDG PET/CT were reported to be highly sensitive with the staging accuracy of 90%–100%, except that one study[25] reported a staging accuracy of 75% for staging with ^{18}F-FDG PET/CT. In indolent NHL group, the staging accuracies of WB-MRI were higher than that of ^{18}F-FDG PET/CT in two studies.[28,30]

Table 1 Main features of studies included in meta-analysis

Study, year	Study type	Patients (n)	Mean age and range (years)	Sex, M/F	Lymphoma subtype
Albano et al,[24] 2016	Cohort, prospective	68	42, 15–86	37/31	HL, DLBCL, FL, MCL
Stecco et al,[25] 2015	Cohort, retrospective	17	63.1, 34–82	12/5	DLBCL, FL, MALTL
Ferrari et al,[5] 2014	Cohort, prospective	27	40.6, 23–78	15/12	HL, DLBCL, MCL, FL, MZL, SLL
Stéphane et al,[26] 2013	Cohort, prospective	23	51, 18–84	11/12	HL, DLBCL, BL
Gu et al,[27] 2011	Cohort, prospective	17	20–80	11/6	HL, FL, DLBCL, NKTCL, AITCL
Mayerhoefer et al,[28] 2014	Cohort, prospective	140	57.6, 19–88	70/70	HL, MALTL, DLBCL, FL, nMZL, MCL, SLL/CLL
Lin et al,[29] 2010	Cohort, prospective	15	48, 23–79	9/6	DLBCL
Abdulqadhr et al,[30] 2011	Cohort, prospective	31	47.2, 18–78	20/11	HL, DLBCL, ALCL, AITCL, PMBL

Abbreviations: AITCL, angioimmunoblastic T-cell lymphoma; ALCL, anaplastic large cell lymphoma; CLL, chronic lymphocytic leukemia; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; MALTL, mucosa-associated lymphoid tissue lymphoma; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; NKTCL, natural killer T-cell lymphoma; nMZL, nodal marginal zone lymphoma; NR, not reported; PMBL, primary mediastinal B-cell lymphoma; SLL, small lymphocytic lymphoma; BL, Burkitt lymphoma; HL, Hodgkin lymphoma.
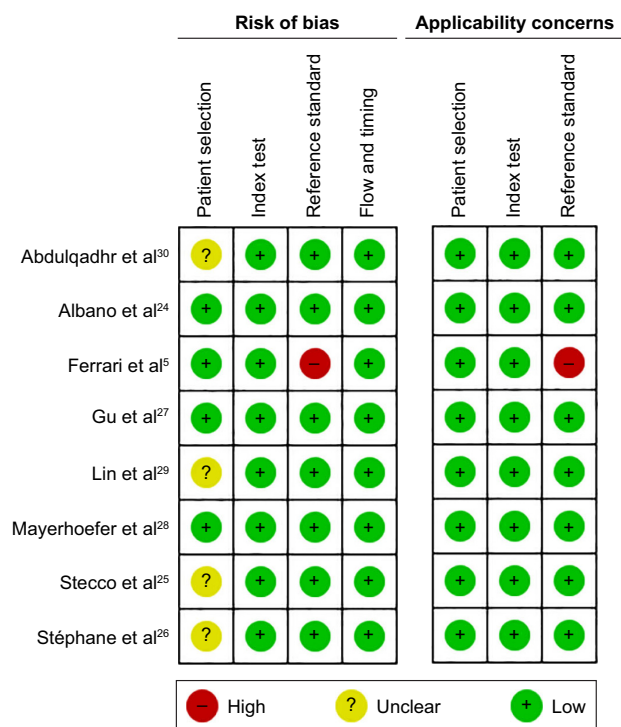
Figure 2 Summary of risk of bias and applicability concerns: reviewers' judgments about each domain for each included study.

Assessment of heterogeneity and publication bias

When analyzing the staging accuracy of WB-MRI, the studies in both the HL and aggressive NHL group and the indolent NHL group showed good homogeneity, according to the χ^2-test (p=0.20 and 0.64, respectively). In the analyses of 18F-FDG PET/CT, no statistically significant heterogeneity was found in the HL and aggressive NHL group (p=0.45; I^2=0.00%). However, moderately high heterogeneity was observed in the indolent NHL group (I^2=49.78%), so we used random effect models for the calculation of pooled estimates.

For both ^{18}F-FDG PET/CT and WB-MRI, Egger's test showed no indications of small-study effects and publication bias in the HL and aggressive NHL group (p=0.099 and 0.659, respectively) and the indolent lymphoma group (p=0.968 and 0.890, respectively). We did not perform a

funnel plot because this meta-analysis included a limited number (<10) of relevant studies.

Pooled staging accuracy of WB-MRI and ^{18}F-FDG PET/CT for initial lymphoma staging

Table 4 lists the pooled staging accuracies derived from the meta-analysis, and Figures 4 and 5 present forest plots of the individual and meta-analytic staging accuracies of the two imaging modalities. In terms of staging, the meta-analytic staging accuracies of WB-MRI and ^{18}F-FDG PET/CT for HL and aggressive NHL were 98% (95% CI, 94%–100%) and 98% (95% CI, 94%–100%), respectively. The pooled staging accuracies of ^{18}F-FDG PET/CT dropped to 87% (95% CI, 72%–97%) for indolent lymphoma, whereas that of WB-MRI remained at 96% (95% CI, 91%–100%).

Subgroup analysis of indolent NHLs in accordance with ^{18}F-FDG avidity

The diagnostic accuracy of ^{18}F-FDG PET/CT relies on the ^{18}F-FDG avidity of lymphoma lesions. The indolent series comprised a heterogeneous group of lymphoma subtypes with variable FDG avidities, which may potentially contribute to the heterogeneity in the indolent group. Thereby, a subgroup analysis was conducted and the patients with indolent NHLs were assigned into two subgroups in view of the FDG avidity: group A, histological subtypes with generally high FDG avidities, including FL and nodal marginal zone lymphoma (nMZL); group B, subtypes in which FDG-avid lesions were less common, including small lymphocytic lymphoma/chronic lymphocytic leukemia (SLL/CLL) and mucosa-associated lymphoid tissue lymphoma (MALTL). The meta-analytic staging accuracies of ^{18}F-FDG PET/CT in group A and group B were 99% (95% CI, 93%–100%) and 60% (95% CI, 23%–92%), respectively. Significant difference was found between the above two estimates according to the test for heterogeneity between subgroups (p<0.001). In contrast to ^{18}F-FDG PET/CT, the meta-analytic staging accuracies of WB-MRI in group A (99%; 95% CI,
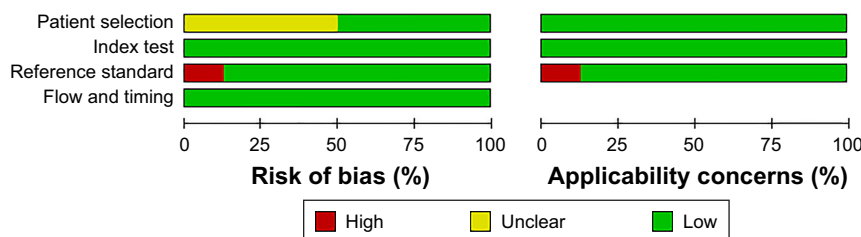

Figure 3 Risk of bias and applicability concerns: reviewers' judgments about each domain, presented as percentages across included studies.

Table 2 Agreement between WB-MRI and ^{18}F-FDG PET/CT for nodal and extra-nodal lymphoma involvement

Study	Nodal		Extra-nodal	
	Kappa value (95% CI)	Agreement	Kappa value (95% CI)	Agreement
Lin et al[29]	0.84 (0.77–0.91)	Excellent	1.00 (1.00–1.00)	Excellent
Stéphane et al[26]	0.92 (0.88–0.96)	Excellent	0.93 (0.85–1.00)	Excellent
Ferrari et al[5]	0.82 (0.73–0.90)	Excellent	0.83 (0.61–1.00)	Excellent
Stecco et al[25]	0.87 (0.82–0.92)	Excellent	NR	NR
Albano et al[24]	0.95 (0.93–0.97)	Excellent	NR	NR

Abbreviations: ^{18}F-FDG, ^{18}F-fluorodeoxyglucose; PET/CT, positron emission tomography/computed tomography; NR, not reported; WB-MRI, whole-body MRI.

93%–100%) and group B (98%; 95% CI, 88%–100%) were comparable (*p*=0.35).

Discussion

This systematic review included eight studies, five of which[5,24–26,29] reported the agreement between WB-MRI and ^{18}F-FDG PET/CT using a region-based scoring system. Regarding lesion detection, excellent agreement between WB-MRI and ^{18}F-FDG PET/CT was reported. We did not perform the meta-analysis on a per-lesion basis because the studies differed widely in design, with several potential sources of heterogeneity. The largest source of variability between studies was most likely the application and interpretation of WB-MRI. In accordance with the stated objectives

Table 3 WB-MRI vs ^{18}F-FDG PET/CT for initial lymphoma staging: variables extracted by groups

Study	Patients (n)	Staging accuracy of WB-MRI	Staging accuracy of PET/CT	ADC calculated
HL and aggressive NHL				
Abdulqadhr et al[30]	26	1.00	1.00	No
Lin et al[29]	15	1.00	0.93	Yes
Gu et al[27]	11	0.91	1.00	No
Stéphane et al[26]	23	1.00	1.00	Yes
Ferrari et al[5]	20	1.00	0.95	Yes
Mayerhoefer et al[28]	62	0.90	0.94	Yes
Stecco et al[25]	4	1.00	0.75	No
Albano et al[24]	58	0.91	0.95	Yes
Indolent NHL				
Abdulqadhr et al[30]	5	1.00	0.40	No
Gu et al[27]	6	0.83	0.83	No
Ferrari et al[5]	7	1.00	1.00	Yes
Mayerhoefer et al[28]	78	0.94	0.79	Yes
Stecco et al[25]	13	0.85	0.85	No
Albano et al[24]	10	1.00	1.00	Yes

Abbreviations: ^{18}F-FDG, ^{18}F-fluorodeoxyglucose; PET/CT, positron emission tomography/computed tomography; ADC, apparent diffusion coefficient; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; WB-MRI, whole-body magnetic resonance imaging.

of the studies, they differed in the sequence, acquisition plane, and slice thickness used and in the body area imaged (Table 5). Additionally, in the process of image interpretation, the criteria for identifying potentially involved nodal or extra-nodal sites also differed across studies, and the cutoffs for lymph node size were not standardized as well. Similarly, the differences in reference standard would also contribute to heterogeneity.

In the evaluation of staging accuracy, the patients were categorized by histologic subtype and were divided into two groups, including with HL or aggressive NHL group and indolent NHL group. WB-MRI has been previously shown to have high accuracy for initial lymphoma staging, compared with ^{18}F-FDG PET/CT, by meta-analysis.[32] But the value of ^{18}F-FDG PET/CT for lymphoma staging depends on the ^{18}F-FDG avidities of the lesions, which vary for different histologic subtypes. According to the literature, HL and most forms of aggressive NHL – such as diffuse large B-cell lymphoma, Burkitt lymphoma, mantle cell lymphoma (MCL), natural killer T-cell lymphoma, and angioimmunoblastic T-cell lymphoma – are ^{18}F-FDG-avid in more than 95% of patients and thus can be easily detected by ^{18}F-FDG PET/CT.[2,3] Indolent lymphomas, however, demonstrate a variety of ^{18}F-FDG avidities. FL, the most common form of indolent NHLs, discloses positive ^{18}F-FDG uptake in 95% of patients,[2] but the uptake is less than that in high-grade lymphomas.[33] nMZL is highly ^{18}F-FDG-avid, whereas extra-nMZL and SLL show a much lower ^{18}F-FDG avidity of 55% and 83%, respectively.[2] Moderate ^{18}F-FDG avidity was observed in patients with MALTL, ranging from 54% to 82%.[2,34,35] Therefore, to determine the role of WB-MRI in the management of different lymphoma subtypes, we compared the staging accuracy of WB-MRI with that of ^{18}F-FDG PET/CT for lymphomas subtypes with high ^{18}F-FDG avidity versus those for which variable ^{18}F-FDG avidity has been reported.

In our study, the pooled staging accuracies of WB-MRI and ^{18}F-FDG PET/CT were high and similar for initial staging

Table 4 Pooled staging accuracies: WB-MRI vs ^{18}F-FDG PET/CT

Series	Modality	μ^a	95% CI[b]	χ^{2c}	df[d]	p-value	I[2e]
HL and aggressive NHL	WB-MRI	0.98	0.94–1.00	9.78	7	0.20	28.42%
	^{18}F-FDG PET/CT	0.98	0.94–1.00	3.42	5	0.45	0.00%
Indolent NHL	WB-MRI	0.96	0.91–1.00	6.79	7	0.64	0.00%
	^{18}F-FDG PET/CT	0.87	0.72–0.97	9.96	5	0.08	49.78%

Notes: [a]μ is meta-analysis derived pooled staging accuracy across all groups. [b]95% CIs are for μ. [c]χ^2 is measure of homogeneity. [d]df indicates degrees of freedom. [e]I[2] indicates variation in μ attributable to heterogeneity.
Abbreviations: ^{18}F-FDG, ^{18}F-fluorodeoxyglucose; PET/CT, positron emission tomography/computed tomography; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; WB-MRI, whole-body magnetic resonance imaging.

and lesion detection in HL and aggressive NHL; 219 patients with these types of lymphoma underwent both WB-MRI and ^{18}F-FDG PET/CT. WB-MRI stage agreed with the reference standard in 207 patients, and ^{18}F-FDG PET/CT in 209 patients. As the present recommended standard imaging method for ^{18}F-FDG-avid lymphoma, ^{18}F-FDG PET/CT has several disadvantages. Radiation, for one, is of most concern. Since patients will require numerous examinations over the course of their treatment and postremission surveillance, the risk of secondary malignancies will increase, especially in younger patients.[36]

Even though more than one study has reported that breathing-motion artifacts degenerated images and led to false-negative results in corresponding nodal regions,[28,30] our results showed that, in terms of staging, diffusion-weighted MRI performed as well as ^{18}F-FDG PET/CT, with an overall staging accuracy of 98% (95% CI, 94%–100%). Therefore, WB-MRI may serve as a viable alternative for initial staging in patients with HL and aggressive NHL.

In total, 119 patients with indolent NHL were included in our meta-analysis. With regard to staging, WB-MRI agreed
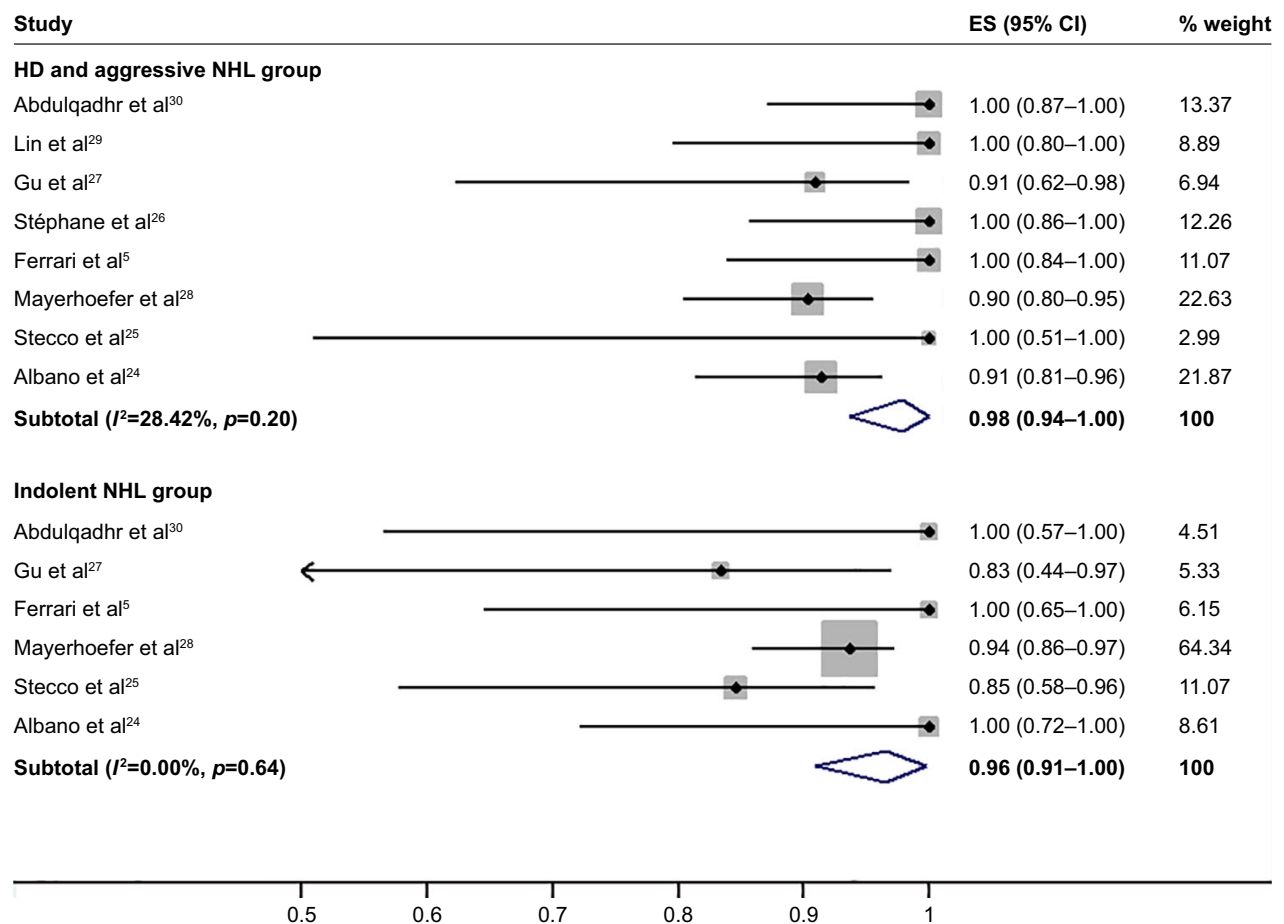

Figure 4 Individual and meta-analytic (μ) sensitivity of WB-MRI vs ^{18}F-FDG PET/CT for initial staging in patients with HD and aggressive NHL. 95% CIs are indicated by horizontal bars.
Abbreviations: ^{18}F-FDG, ^{18}F-fluorodeoxyglucose; PET/CT, positron emission tomography/computed tomography; ES, effect size; HD, Hodgkin disease; NHL, non-Hodgkin lymphoma; WB-MRI, whole-body MRI.
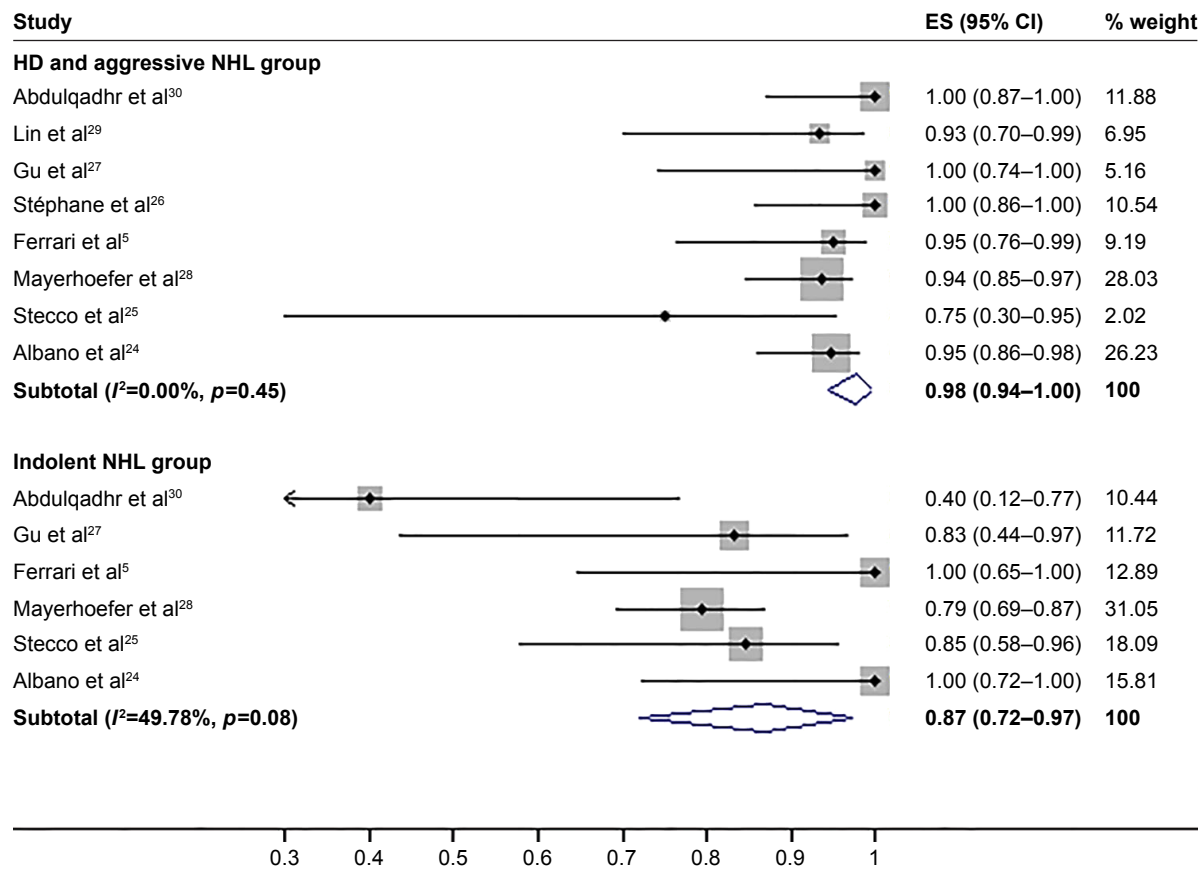
Study			ES (95% CI)	% weight
HD and aggressive NHL group				
Abdulqadhr et al[30]			1.00 (0.87–1.00)	11.88
Lin et al[29]			0.93 (0.70–0.99)	6.95
Gu et al[27]			1.00 (0.74–1.00)	5.16
Stéphane et al[26]			1.00 (0.86–1.00)	10.54
Ferrari et al[5]			0.95 (0.76–0.99)	9.19
Mayerhoefer et al[28]			0.94 (0.85–0.97)	28.03
Stecco et al[25]			0.75 (0.30–0.95)	2.02
Albano et al[24]			0.95 (0.86–0.98)	26.23
Subtotal (I^2=0.00%, p=0.45)			**0.98 (0.94–1.00)**	**100**
Indolent NHL group				
Abdulqadhr et al[30]			0.40 (0.12–0.77)	10.44
Gu et al[27]			0.83 (0.44–0.97)	11.72
Ferrari et al[5]			1.00 (0.65–1.00)	12.89
Mayerhoefer et al[28]			0.79 (0.69–0.87)	31.05
Stecco et al[25]			0.85 (0.58–0.96)	18.09
Albano et al[24]			1.00 (0.72–1.00)	15.81
Subtotal (I^2=49.78%, p=0.08)			**0.87 (0.72–0.97)**	**100**

```
0.3  0.4  0.5  0.6  0.7  0.8  0.9  1
```

Figure 5 Individual and meta-analytic (μ) sensitivity of WB-MRI vs ^{18}F-FDG PET/CT for initial staging in patients with indolent NHL. 95% CIs are indicated by horizontal bars.
Abbreviations: ^{18}F-FDG, ^{18}F-fluorodeoxyglucose; PET/CT, positron emission tomography/computed tomography; ES, effect size; HD, Hodgkin disease; NHL, non-Hodgkin lymphoma; WB-MRI, whole-body MRI.

Table 5 WB-MRI protocols of included studies

Study	Protocol		Parameters		Extent
	Sequence	Plane	S,[a] mm	b value, s/mm^2	
Abdulqadhr et al[30]	T1WI, T2W-STIR	Coronal	6	–	NR
	DWIBS	Axial	6	0 and 1,000	
Lin et al[29]	DWI	Axial	5	50, 400, and 800	Head, neck, chest, abdomen, pelvis, proximal thigh
Gu et al[27]	T2WI, T2W SPAIR	Axial	5	–	Head to distal thigh
	DWI	Axial	6	0 and 1,000	
Stéphane et al[26]	T1WI, STIR	NR	NR	NR	Head to thigh
	DWIBS	Axial	4	0 and 1,000	
Ferrari et al[5]	T1WI, STIR	Coronal	6	–	Head to thigh
	DWIBS	Axial	6	0, 500, and 1,000	
Mayerhoefer et al[28]	T1W-TSE, FLASH	NR	NR	NR	Vertex to upper thigh
	DWI	Axial	5	50 and 1,000	
Klenk et al[31]	T1W-SPGR	Axial	4–5	–	NR
	DWI	Axial	4–5	0 and 600	
Stecco et al[25]	T1W-TSE, T2W-STIR	Axial	6	–	Head to feet
	DWIBS	Axial	6	0 and 1,000	
Albano et al[24]	T1WI, STIR	Coronal	4–5	–	Vertex to toes
	DWI	Axial	4–5	0 and 800	

Notes: [a]S indicates slice thickness. b value is a parameter of diffusion weighted sequence. "–" indicates not applicable.
Abbreviations: DWI, diffusion-weighted imaging; DWIBS, diffusion-weighted wholebody imaging with background body signal suppression; FLASH, fast low-angle shot; NR, not reported; SPAIR, spectral attenuated inversion recovery; SPGR, spoiled gradient; STIR, short-tau inversion recovery; T1W, T1-weighted imaging; T2W, T2-weighted imaging; TSE, turbo spin echo; WB-MRI, whole-body MRI.

with the reference standard in 111 patients, and ^{18}F-FDG PET/CT, in 97 patients. Moderately high heterogeneity was found in the indolent NHL group for staging with ^{18}F-FDG PET/CT. It may due in part to the differences in the distribution of histological subtypes of the study population which consequently resulted in the diversity of ^{18}F-FDG avidity among the studies. Compared with diagnostic performance in HL and aggressive NHLs, the meta-analytic staging accuracy of WB-MRI in the indolent series (96%; 95% CI, 91%–100%) remained high, whereas that of ^{18}F-FDG PET/CT (87%; 95% CI, 72%–97%) decreased significantly. Moderately high heterogeneity was observed in the indolent NHL group; thereby, a subgroup analysis was performed in view of ^{18}F-FDG avidity. The results of the subgroup analysis revealed that ^{18}F-FDG PET/CT remained highly sensitive in staging of ^{18}F-FDG avid indolent lymphomas, and the less favorable staging accuracy of ^{18}F-FDG PET/CT for indolent series was primarily caused by the inclusion of less ^{18}F-FDG avid lymphoma subtypes (ie, SLL/CLL and MALTL) in the indolent group. In contrast, the staging performance of WB-MRI was stable and excellent, regardless of the variety of histological subtypes. These results indicate that WB-MRI, as a functional imaging method, may be less histology-dependent than ^{18}F-FDG PET/CT, as is consistent with a previous study.[28] Such results may be explained by the hypothesis that high cellular density, which can be detected by diffusion-weighted imaging, is a more general feature of lymphoma than elevated glucose metabolism. For many types of indolent lymphomas, such as MALTL and SLL/CLL, no functional imaging modality has yet been recommended for staging or for treatment response assessment. Our results showed that WB-MRI may be the imaging method of choice for staging of these subtypes of indolent lymphomas.

Our meta-analysis had several limitations. First, we did not perform subgroup analyses according to every NHL subtype because the studies included a wide spectrum of subtypes, some of which (eg, small lymphocytic lymphoma/chronic lymphocytic leukemia, MZL, and MCL) were represented by only a few patients. Larger sample sizes are essential for evaluating the staging performance of WB-MRI. Second, the differences in WB-MRI protocols (Table 5) and image interpretation methods may have contributed to a large part of the heterogeneity. Even though WB-MRI seems a promising and radiation-free imaging method for initial staging of lymphoma patients, there is no agreement concerning the sequence protocol applied, especially for the choice of *b* values (a parameter of diffusion weighted sequence), which may compromise the reproducibility of the results

among different studies included in the meta-analysis.[37] Indeed, lack of reproducibility may also be one of the major obstacles against the demonstration of the potential merits of WB-MRI for lymphoma staging, which calls for multicenter prospective studies to address the degree to which the diagnostic performance of WB-MRI depends on the sequence protocol or the method of acquisition. Third, the differences in reference standards may have increased the clinical heterogeneity. Only one study used histopathology aided by clinical and imaging follow-up as the standard of reference.[31] Although a pathologic diagnosis is regarded as the most reliable evidence of tumor involvement, invasive exploration of all potential sites for pathologic analysis is ethically and practically infeasible since lymphoma is often diffuse. On that account, most studies based the reference standard on the integrated interpretation of ^{18}F-FDG PET/CT and WB-MRI images, and confirmed the stage by clinical and imaging follow-up or bone marrow biopsy, especially when the results of WB-MRI and ^{18}F-FDG PET/CT were discordant.[26–30] Inclusion of an index test (WB-MRI) as part of the reference standard may increase the risk of incorporation bias and, thus, lead to an overestimation of diagnostic accuracy.

Conclusion

WB-MRI is a promising radiation-free imaging technique that may serve as a viable alternative to ^{18}F-FDG PET/CT for staging of [18]FDG-avid lymphomas, where ^{18}F-FDG PET/CT remains the standard of care. Additionally, WB-MRI seems a less histology-dependent functional imaging test than ^{18}F-FDG PET/CT and may be the imaging test of choice for staging of indolent lymphomas with low ^{18}F-FDG avidity. Larger-scaled or multicenter prospective studies are needed to further confirm the usefulness of WB-MRI for staging of lymphoma.

Acknowledgments

The abstract of this paper was presented at the Society of Nuclear Medicine and Molecular Imaging 2017 Annual Meeting as a poster presentation with interim findings. The poster's abstract was published in *The Journal of Nuclear Medicine* at: http://jnm.snmjournals.org/content/58/supplement_1/636.short?58%2Fsupplement_1%2F636=&cited-by=yes&legid=jnumed.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Cronin CG, Swords R, Truong MT, et al. Clinical utility of PET/CT in lymphoma. *AJR Am J Roentgenol*. 2010;194(1):W91–W103.

2. Weiler-Sagie M, Bushelev O, Epelbaum R, et al. (18)F-FDG avidity in lymphoma readdressed: a study of 766 patients. *J Nucl Med*. 2010; 51(1):25–30.

3. Chan WK, Au WY, Wong CY, et al. Metabolic activity measured by F-18 FDG PET in natural killer-cell lymphoma compared to aggressive B- and T-cell lymphomas. *Clin Nucl Med*. 2010;35(8):571–575.

4. Jerusalem G, Beguin Y, Najjar F, et al. Positron emission tomography (PET) with 18F-fluorodeoxyglucose (18F-FDG) for the staging of low-grade non-Hodgkin's lymphoma (NHL). *Ann Oncol*. 2001;12(6): 825–830.

5. Ferrari C, Minoia C, Asabella AN, et al. Whole body magnetic resonance with diffusion weighted sequence with body signal suppression compared to (18)F-FDG PET/CT in newly diagnosed lymphoma. *Hell J Nucl Med*. 2014;17 Suppl 1:40–49.

6. Cheson BD. Role of functional imaging in the management of lymphoma. *J Clin Oncol*. 2011;29(14):1844–1854.

7. Luminari S, Biasoli I, Arcaini L, et al. The use of FDG-PET in the initial staging of 142 patients with follicular lymphoma: a retrospective study from the FOLL05 randomized trial of the Fondazione Italiana Linfomi. *Ann Oncol*. 2013;24(8):2108–2112.

8. Huang B, Law MW, Khong PL. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. *Radiology*. 2009;251(1): 166–174.

9. Robbins E. Radiation risks from imaging studies in children with cancer. *Pediatr Blood Cancer*. 2008;51(4):453–457.

10. Kleinerman RA. Cancer risks following diagnostic and therapeutic radiation exposure in children. *Pediatr Radiol*. 2006;36 Suppl 2: 121–125.

11. Kellenberger CJ, Miller SF, Khan M, Gilday DL, Weitzman S, Babyn PS. Initial experience with FSE STIR whole-body MR imaging for staging lymphoma in children. *Eur Radiol*. 2004;14(10):1829–1841.

12. Brennan DD, Gleeson T, Coate LE, Cronin C, Carney D, Eustace SJ. A comparison of whole-body MRI and CT for the staging of lymphoma. *AJR Am J Roentgenol*. 2005;185(3):711–716.

13. Lauenstein TC, Goehde SC, Herborn CU, et al. Whole-body MR imaging: evaluation of patients for metastases. *Radiology*. 2004;233(1): 139–148.

14. Kwee TC, Takahara T, Ochiai R, Nievelstein RA, Luijten PR. Diffusion-weighted whole-body imaging with background body signal suppression (DWIBS): features and potential applications in oncology. *Eur Radiol*. 2008;18(9):1937–1952.

15. Taylor DG, Bushell MC. The spatial mapping of translational diffusion coefficients by the NMR imaging technique. *Phys Med Biol*. 1985; 30(4):345–349.

16. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536.

17. Freeman MF, Tukey JW. Transformations related to the angular and the square root. *Ann Math Stat*. 1950;21(4):607–611.

18. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Ass*. 1927;22(158):209–212.

19. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539–1558.

20. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–560.

21. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629–634.

22. Littooij AS, Kwee TC, Barber I, et al. Whole-body MRI for initial staging of paediatric lymphoma: prospective comparison to an FDG-PET/CT-based reference standard. *Eur Radiol*. 2014;24(5):1153–1165.

23. van Ufford HM, Kwee TC, Beek FJ, et al. Newly diagnosed lymphoma: initial results with whole-body T1-weighted, STIR, and diffusion-weighted MRI compared with 18F-FDG PET/CT. *AJR Am J Roentgenol*. 2011;196(3):662–669.

24. Albano D, Patti C, La Grutta L, et al. Comparison between whole-body MRI with diffusion-weighted imaging and PET/CT in staging newly diagnosed FDG-avid lymphomas. *Eur J Radiol*. 2016;85(2):313–318.

25. Stecco A, Buemi F, Quagliozzi M, et al. Staging of primary abdominal lymphomas: comparison of whole-body mri with diffusion-weighted imaging and (18)F-FDG-PET/CT. *Gastroenterol Res Pract*. 2015; 2015:104794.

26. Stéphane V, Samuel B, Vincent D, et al. Comparison of PET-CT and magnetic resonance diffusion weighted imaging with body suppression (DWIBS) for initial staging of malignant lymphomas. *Eur J Radiol*. 2013;82(11):2011–2017.

27. Gu J, Chan T, Zhang J, Leung AY, Kwong YL, Khong PL. Whole-body diffusion-weighted imaging: the added value to whole-body MRI at initial diagnosis of lymphoma. *AJR Am J Roentgenol*. 2011;197(3): W384–W391.

28. Mayerhoefer ME, Karanikas G, Kletter K, et al. Evaluation of diffusion-weighted MRI for pretherapeutic assessment and staging of lymphoma: results of a prospective study in 140 patients. *Clin Cancer Res*. 2014;20(11):2984–2993.

29. Lin C, Luciani A, Itti E, et al. Whole-body diffusion-weighted magnetic resonance imaging with apparent diffusion coefficient mapping for staging patients with diffuse large B-cell lymphoma. *Eur Radiol*. 2010; 20(8):2027–2038.

30. Abdulqadhr G, Molin D, Astrom G, et al. Whole-body diffusion-weighted imaging compared with FDG-PET/CT in staging of lymphoma patients. *Acta Radiol*. 2011;52(2):173–180.

31. Klenk C, Gawande R, Uslu L, et al. Ionising radiation-free whole-body MRI versus (18)F-fluorodeoxyglucose PET/CT scans for children and young adults with cancer: a prospective, non-randomised, single-centre study. *Lancet Oncol*. 2014;15(3):275–285.

32. Regacini R, Puchnick A, Shigueoka DC, Iared W, Lederman HM. Whole-body diffusion-weighted magnetic resonance imaging versus FDG-PET/CT for initial lymphoma staging: systematic review on diagnostic test accuracy studies. *Sao Paulo Med J*. 2015;133(2):141–150.

33. Schoder H, Noy A, Gonen M, et al. Intensity of 18fluorodeoxyglucose uptake in positron emission tomography distinguishes between indolent and aggressive non-Hodgkin's lymphoma. *J Clin Oncol*. 2005;23(21): 4643–4651.

34. Hoffmann M, Wöhrer S, Becherer A, et al. 18F-Fluoro-deoxy-glucose positron emission tomography in lymphoma of mucosa-associated lymphoid tissue: histology makes the difference. *Ann Oncol*. 2006;17(12): 1761–1765.

35. Radan L, Fischer D, Bar-Shalom R, et al. FDG avidity and PET/CT patterns in primary gastric lymphoma. *Eur J Nucl Med Mol Imaging*. 2008;35(8):1424–1430.

36. Brenner DJ, Elliston CD. Estimated radiation risks potentially associated with full-body CT screening. *Radiology*. 2004;232(3):735–738.

37. Lin C, Luciani A, Itti E, et al. Whole-body diffusion magnetic resonance imaging in the assessment of lymphoma. *Cancer Imaging*. 2012; 12(2):403–408.

Supplementary material

Table S1 Search strategy

Source	Search strategy	Hits retrieved
1. MEDLINE In-process and other nonindexed citations and MEDLINE 1946–present (Ovid SP)	1. "Positron emission tomography".ti,ab. 2. exp Positron-emission Tomography/ 3. "FDG-PET".ti,ab. 4. PET.ti,ab. 5. ("18F-FDG" or "FDG uptake").ti,ab. 6. fluodeoxyglucose*.ti,ab. 7. fluorodeoxyglucose*.ti,ab. 8. or/1-7 9. exp Magnetic Resonance Imaging/ 10. (MRI or MRIs).ti,ab. 11. (("magnetic resonance" or MR) adj2 (imag* or tomograph* or scan*)).ti,ab. 12. ((WB or "whole body") adj2 ("magnetic resonance" or MR or MRI)).ti,ab. 13. "diffusion weighted imag*".ti,ab. 14. DWI.ti,ab. 15. ((DW or "diffusion weighted") adj2 ("magnetic resonance" or MR or MRI)).ti,ab. 16. or/9-15 17. exp lymphoma/ 18. lymphom*.ti,ab. 19. (Hodgkin or Hodgkins or Hodgkin's).ti,ab. 20. ("non-Hodgkin" or "non-Hodgkins" or "non-Hodgkin's").ti,ab. 21. (HD or HL).ti,ab. 22. (NHD or NHL).ti,ab. 23. (AITCL or CLL or DLBCL or FL or MALTL or MCL or MZL or NKTCL or nMZL or PMBL or SLL).ti,ab. 24. or/17-23 25. 8 and 16 and 24	December 2016: 845
2. EMBASE	#1 'positron emission tomography'/exp #2 'positron emission tomography':ab,ti #3 pet:ab,ti #4 'fdg pet':ab,ti #5 '18f-fdg':ab,ti OR 'fdg uptake':ab,ti #6 fluodeoxyglucose*:ab,ti #7 fluorodeoxyglucose*:ab,ti #8 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 #9 'magnetic resonance imaging'/exp #10 'magnetic resonance imaging':ab,ti #11 mri:ab,ti OR mris:ab,ti #12 (('magnetic resonance' OR mr) NEAR/2 (imag* OR tomograph* OR scan*)):ab,ti #13 ((wb OR 'whole body') NEAR/2 ('magnetic resonance' OR mr OR mri)):ab,ti #14 'diffusion weighted imag*':ab,ti #15 dwi:ab,ti #16 ((dw OR 'diffusion weighted') NEAR/2 ('magnetic resonance' OR mr OR mri)):ab,ti #17 #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 #18 'lymphoma'/exp #19 lymphoma:ab,ti #20 hodgkin*:ab,ti #21 'non hodgkin*':ab,ti #22 hd:ab,ti OR hl:ab,ti #23 nhd:ab,ti OR nhl:ab,ti #24 aitcl:ab,ti OR cll:ab,ti OR dlbcl:ab,ti OR fl:ab,ti OR maltl:ab,ti OR mcl:ab,ti OR mzl:ab,ti OR nktcl:ab,ti OR nmzl:ab,ti OR pmbl:ab,ti OR sll:ab,ti #25 #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 #26 #8 AND #17 AND #25	December 2016: 2,598

(*Continued*)

Table S1 (*Continued*)

Source	Search strategy	Hits retrieved
3. CENTRAL	1. MeSH descriptor Lymphoma explode all trees	December 2016: 10
	2. MeSH descriptor Hodgkin Disease explode all trees	
	3. (hodgkin* or hogkin* or hodkin* or hodgin*):ti,ab,kw	
	4. ("non hodgkin*" or "non hogkin*" or "non hodkin*" or "non hodgin*"):ti,ab,kw	
	5. (AITCL or CLL or DLBCL or FL or MALTL or MCL or MZL or NKTCL or nMZL or PMBL or SLL):ti,ab,kw	
	6. (NHD or NHL):ti,ab,kw	
	7. (HD or HL):ti,ab,kw	
	8. (#1 or #2 or #3 or #4 or #5 or #6 or #7)	
	9. MeSH descriptor Positron-Emission Tomography explode all trees	
	10. MeSH descriptor Tomography, Emission-Computed explode all trees	
	11. (pet* or petscan* or (Positron* and emission*) or (Positron* and tomography*)):ti,ab,kw	
	12. (pet* and (deoxy* or fluor* or 18fluor* or fdg* or 18fdg* or fludeoxy*)):ti,ab,kw	
	13. (pet* or petscan*):ti,ab,kw	
	14. (tomograph* or tomographs* or tomographic* or tomography* or tomographies*):ti,ab,kw	
	15. emission*:ti,ab,kw	
	16. (#14 and #15)	
	17. (#9 or #10 or #11 or #12 or #13 or #16)	
	18. MeSH descriptor Magnetic Resonance Imaging explode all trees	
	19. "diffusion weighted imag*":ti,ab,kw	
	20. DWI:ti,ab,kw	
	21. (DW or "diffusion weighted"):ti,ab	
	22. ("magnetic resonance" or MR or MRI):ti,ab	
	23. (#21 and #22)	
	24. (#18 or #19 or #20 or #23)	
	25. (#8 and #17 and #24)	