

Specialized second-opinion radiology review of PET/CT examinations for patients with diffuse large B-cell lymphoma impacts patient care and management

Peter Sawan, MD, FRCR^{a,b}, Karim Rebeiz, MD^b, Heiko Schoder, MD^b, Connie Batlevi, MD^c, Alison Moskowitz, MD^c, Gary A. Ulaner, MD^b, Mark Dunphy, DO^b, Lorenzo Mannelli, MD, PhD^{a,b,*}

Abstract

To identify discrepancies in fludeoxyglucose positron emission tomography/computed tomography (FDG PET/CT) reports generated by general radiologists and subspecialized oncological radiologists for patients with diffuse large B-cell lymphoma (DLBCL), and to assess if such discrepancies impact patient management.

Two radiologists retrospectively reviewed 72 PET/CT scans of patients with DLBCL referred to our institutions between 2009 and 2011, and recorded the discrepancies between the outside and second-opinion reports regarding multiple preset criteria using kappa statistic (K), including the disease stage. A multidisciplinary staging that considered all patient clinical data, pathology, and follow up radiological scans, was considered as standard of reference. A hemato-oncologist, blinded to the reports' origin, subjectively graded the quality and structure of these reports for each patient to determine if clinical stage and disease activity could be derived accurately from these reports.

Agreement was not, or slightly, achieved between the reports regarding the binary and multilevel criteria ($K < 0-0.2$ and weighted $K = 0.082$, respectively). Second-opinion reviews of PET/CT scans were concordant with the multidisciplinary staging in 78% of cases with an almost perfect agreement ($K = 0.860$). A change in staging was demonstrated in 36% of cases. In addition, 68% of second-opinion reports were assigned the highest grades on quality (grades 4 and 5) by the hemato-oncologist, compared with 15% of outside reports, with no noted agreement (weighted $K = -0.007$).

Second-opinion review of PET/CT scans by sub-specialized oncological radiologists increases accuracy of initial staging, posttreatment evaluation and also the clinical relevance of the radiology reports.

Abbreviations: DLBCL = diffuse large B cell lymphoma, FDG PET/CT = fludeoxyglucose positron emission tomography/computed tomography, PACS = picture archiving and communication system, SUVmax = maximum standardized uptake value.

Keywords: diffuse large B-cell lymphoma, FDG PET/CT, non-Hodgkin lymphoma, patient management, radiology report, second opinion

1. Introduction

Recent advances in chemotherapy, immunotherapy and bone marrow transplant for the treatment of patients with lymphoma have placed renewed emphasis on accurate staging and characterization of disease activity and response.^[1,2] Imaging

with fluorodeoxyglucose positron emission tomography-computed tomography (FDG PET/CT) is now an integral part of the staging and management of most patients with lymphoma.^[3,4] In daily practice, these PET/CT scans are interpreted by radiologists or nuclear medicine physicians with a variety of backgrounds and subspecialty training. As in other malignancies, the criteria for staging and response assessment for lymphoma have changed over the past decades,^[5,6] and for physicians interpreting these scans it is imperative to be familiar with such criteria to generate a clinically meaningful radiology report. Accurate reporting also requires familiarity with the various imaging patterns in which a disease can present on cross-sectional imaging for the interpreting radiologist. Sometimes, patients present to a secondary or tertiary care center with imaging studies and radiology reports generated in the community or at other institutions. In our institution, it is customary that such imaging studies are stored in the picture archiving and communication system (PACS) and presented to staff radiologists or nuclear medicine physicians for official interpretation prior to any management decision. This same standard applies to pathology specimens and outside pathology reports studies. Indeed, review and re-interpretation of pathology specimens in dedicated hemato-pathologists has been shown to cause a change in diagnosis in 17.8% and 16.4% of patients in years 2001 and 2006, respectively, in lymphoma patients.^[7] Similarly, several studies have suggested that interpretation of

Editor: Michael Masoomi.

This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748.

The authors declare no conflicts of interest.

^a Department of Radiology, ^b Department of Molecular Imaging and Therapy,

^c Department of Medicine, Memorial Sloan-Kettering Cancer Center, York Avenue, New York, NY.

* Correspondence: Lorenzo Mannelli, Department of Radiology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY (e-mail: mannelliorenzo@yahoo.it).

Copyright © 2017 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the Creative Commons Attribution-NoDerivatives License 4.0, which allows for redistribution, commercial and noncommercial, as long as it is passed along unchanged and in whole, with credit to the author.

Medicine (2017) 96:51(e9411)

Received: 14 August 2017 / Received in final form: 3 November 2017 /

Accepted: 28 November 2017

<http://dx.doi.org/10.1097/MD.00000000000009411>

radiologic imaging studies by expert radiologists is more accurate than interpretation by radiologists without subspecialization,^{18-10]} and that such re-interpretation of imaging studies may have significant impact on patient stage and management. In the current analysis, we chose to study lymphoma patients as this disease has the largest number of submitted outside PET/CTs to our department. We focused more specifically on patients with reliably FDG avid lymphoma histological subtype, diffuse large B-cell lymphoma (DLBCL), and investigated if review of outside FDG PET/CT studies by subspecialized oncoradiologists would alter patient stage and have implication for treatment response evaluation. We particularly focused on the accurate description of clinical stages and disease activity, using modern classification schemes.^[6,11]

2. Methods

2.1. Case selection

This study was approved by the institutional review board with a waiver of informed consent. The analysis was compliant with the Health Insurance Portability and Accountability Act.

PET/CT scans were performed and reported in a variety of outside institutions (including private practice clinics, community hospitals, academic hospitals, and less frequently cancer centers).

Imaging studies and the accompanying radiology report were submitted to our institutional PACS. As shown in Figure 1, we conducted a retrospective data base search using the following

inclusion criteria: biopsy-proven DLBCL; outside FDG PET/CT between the years 2009 and 2011 with at least 1 year of radiological and clinical follow-up (recent enough to ensure use of modern PET/CT technology but distant enough to ensure sufficient follow-up); available outside PET/CT report; ability to properly display outside studies and fuse raw axial CT and PET data using our reader software (GE healthcare volume viewer AW server 3.2), and the ability to extract the maximum standardized uptake values (SUVmax) from these imaging data. From the initial sample of 169 patients and scans collected in these 3 years, 72 were found eligible for this current analysis.

2.2. Data review and analysis

In this study, the reports from outside institutions will be referred to as “outside” reports whereas the reports from our institution will be referred to “second-opinion” reports.

Two sub-specialized oncoradiologists (PS and KR) with 6 years of experience and 2 fellowships training in nuclear medicine and oncological imaging, reviewed retrospectively the 72 scans, the original reports and the reports generated by staff radiologists/nuclear medicine physicians at our institution, and evaluated if these reports addressed critical clinical questions, including: sites of disease with exact description of location and size, the intensity of FDG uptake, appropriate classification of disease extent to enable assigning a clinical stage according to Ann Arbor classification, and appropriate measurement and description of residual FDG uptake during or after therapy to enable

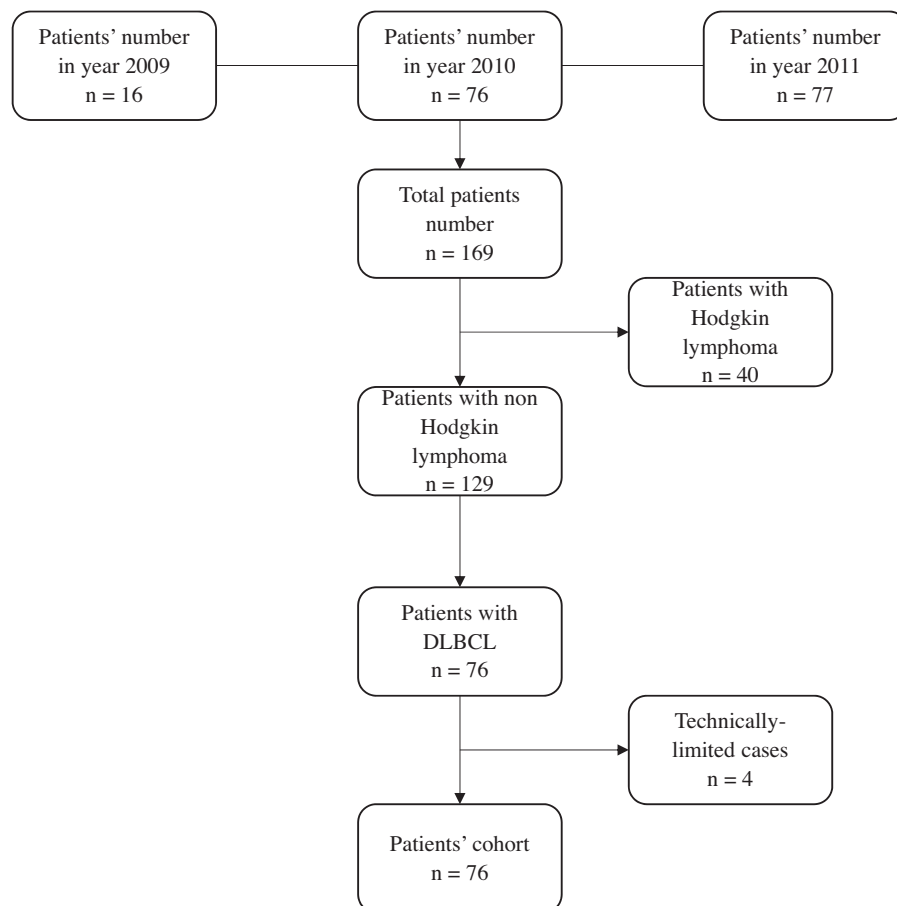


Figure 1. Patients' flow chart.

characterization of response using current recommendations .^[6] The 2 radiologists also recorded whether radiology reports contained SUV number standardized reference regions in mediastinum and liver (needed for quality assurance and response assessment), and evaluated the level of confidence with which the outside report distinguished between malignant and benign etiology, elaborated on differential diagnoses, and further recommended management.

All the PET/CT studies were from centers where the radiologists are involved in a wide range of diagnostic imaging studies. None of the radiologists was a dedicated oncological imaging radiologist. We defined a radiologist as a subspecialized oncoradiologist, if he had at least 1 year of oncological imaging fellowship training.

A multidisciplinary staging considering all patients clinical, pathological, and imaging data, and clinical and imaging follow-up was considered as the standard of reference.

In addition, a single hemato-oncologist with 7 years' experience in managing and treating lymphomas compared the text of the outside PET/CT report to the report generated by our institutional specialized oncologic radiologist or nuclear medicine physician without any information regarding the origin of the report. This physician recorded the clinical stage based on the information that could be derived from the report text and graded the report clarity on a scale from 1 to 5, with "grade 1" being poor, "grade 2" being fair, "grade 3" being average, "grade 4" being good and "grade 5" being excellent. The grade was a subjective measure of the report structure, addressing how easily and accurately the hemato-oncologist was able to extract the information needed for proper staging, restaging and subsequent management.

2.3. Statistical analysis

Kappa statistic (K) was used to assess the agreement on PET/CT features from outside and second-opinion reports. Weighted kappa with squared weights was used for features with multiple levels, including grade, mediastinum (yes, no, and unknown) and staging. We then compared the likelihood of the criteria present in the report, the staging, and the grade between the outside reports and second-opinion reports using McNemar test for 2-level features and Bowker's test of symmetry for multiple-level features, respectively. Agreement between radiology staging in each report and the multidisciplinary standard of reference staging was also assessed using weighted kappa. Kappa statistic is interpreted as follows: < 0, no agreement; 0.00–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement. A test with P < .05 was considered statistically significant. All statistical analyses were performed in software packages SAS 9.4 (SAS Institute Inc., Cary, NC).

3. Results

Of the 72 PET/CT scans analyzed between the years 2009 and 2011, 35 were obtained for initial-staging and 37 were obtained during follow-up and for assessment of response to treatment.

3.1. Head to head comparison of the second-opinion and outside readings

Differences between the second-opinion and outside readings regarding the studied criteria are shown in Table 1. The second-opinion reviews showed no agreement or slight agreement with

Table 1
The second-opinion and outside readings regarding the studied criteria: present or absent.

	Initial reading	Second-opinion reading	kappa (95% CI)	McNemar test P
Anatomical site of disease	No	No 0 (0%)	Yes 17 (24%)	-0.054 (-0.177, 0)
	Yes	2 (3%)	51 (73%)	
Measuring SUVmax	No	No 3 (4%)	Yes 12 (18%)	-0.008 (-0.214, 0.256)
	Yes	11 (16%)	42 (62%)	
Deauville/Lugano terminology applied	No	No 6 (8%)	Yes 43 (60%)	0.024 (-0.107, 0.109)
	yes	2 (3%)	21 (29%)	
DD for avid lesions	No	No 0 (0%)	Yes 16 (25%)	0
	Yes	0 (0%)	47 (75%)	
Lesion with highest SUVmax noted	No	No 3 (5%)	Yes 6 (10%)	-0.016 (-0.203, 0.217)
	Yes	19 (31%)	34 (55%)	
Size measurement	No	No 17 (27%)	Yes 12 (19%)	0.185 (-0.066, 0.402)
	Yes	14 (22%)	21 (33%)	
Confidence in differentiating benign versus malignant lesions	No	No 0 (0%)	Yes 18 (25%)	-0.053 (-0.168, 0)
	Yes	2 (3%)	51 (72%)	
Structured report	No	No 2 (3%)	Yes 38 (53%)	0.045 (0, 0.160)
	Yes	0 (0%)	32 (44%)	
Further imaging recommended	No	No 48 (67%)	Yes 5 (7%)	0.078 (-0.113, 0.326)
	Yes	16 (22%)	3 (4%)	

CI = confidence interval, DD = differential diagnosis, SUV = standardized uptake value. kappa statistic is interpreted as follows: ≤0 no agreement; 0.00–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement.

Table 2**Comparison of quality grading (top) and radiological staging (bottom) of second-opinion and outside reports.**

		Second-opinion reading					Weighted <i>kappa</i> (95% CI)	Bowker's test of symmetry <i>P</i>
Outside staging	Grade	1	2	3	4	5	-0.007 (-0.123, 0.132)	<0.001
	1	0 (0%)	1 (1%)	1 (1%)	4 (6%)	2 (3%)		
	2	2 (3%)	3 (4%)	2 (3%)	6 (8%)	10 (14%)		
	3	1 (1%)	3 (4%)	7 (10%)	7 (10%)	9 (12%)		
	4	0 (0%)	1 (1%)	1 (1%)	5 (7%)	2 (3%)		
	5	0 (0%)	0 (0%)	1 (1%)	4 (6%)	0 (0%)		
Staging	NED	I	II	III	IV	0.674 (0.469, 0.824)	0.308	
	NED	17 (24%)	5 (7%)	2 (3%)	2 (3%)			2 (3%)
	I	1 (1%)	4 (6%)	1 (1%)	1 (1%)			1 (1%)
	II	1 (1%)	3 (4%)	5 (7%)	0 (0%)			0 (0%)
	III	0 (0%)	2 (3%)	1 (1%)	8 (11%)			2 (3%)
	IV	1 (1%)	0 (0%)	1 (1%)	0 (0%)			12 (17%)

CI=confidence interval, NED=no evidence of disease.

outside reports for most of the binary criteria (Table 1). For example, there was no agreement for accurate anatomic description of anatomical sites of disease ($K=-0.054$). In addition, the likelihood of presenting each criterion was different between the outside and second-opinion reading ($P<.02$), except for size and SUVmax measurements ($P=.695$ and $.835$ respectively).

3.2. Comparison of quality grading and radiological staging of second-opinion and outside reports

Table 2 shows the discordance in quality of reports generated by second-opinion readers versus outside readers and the related

discordance in clinical stage that could be assigned based on these reports. In general, second-opinion reports were judged to be of higher quality by the hemato-oncologist who assigned higher grades (4 and 5) in 49/72 patients (68%), as compared with the outside reports for which only 11/72 (15%) received a grade of 4 or 5 (weighted $K=-0.007$ (-0.123, 0.132).

Discordance between the second-opinion and outside reports regarding disease staging was noted in 26/72 (36%) of patients (weighted $K=0.674$, 95% confidence interval, CI: 0.469, 0.824). Of these 26 patients, 4 patients (15%) were upstaged from “no evidence of disease” to “stages 3 and 4”; and 3 patients (11%) were downstaged to no evidence of disease. Figure 2 represents a

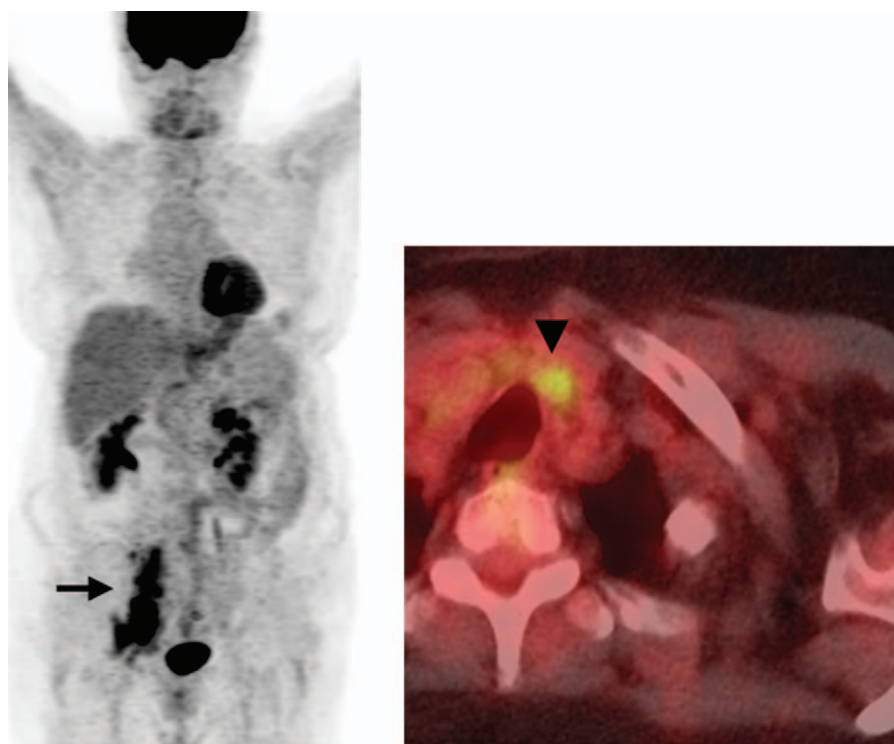


Figure 2. A 71-year-old female patient with DLBCL undergoing a PET/CT for initial staging. The outside report indicated hypermetabolic right iliac chain adenopathy (black arrow on MIP image) and a cervical central compartment adenopathy (black arrowhead on axial fused PET/CT image) consistent with stage 3 disease. Second-opinion review at our institution reported infradiaphragmatic adenopathy consistent with stage 1 disease. The cervical hypermetabolic focus corresponds to a benign left thyroid nodule.

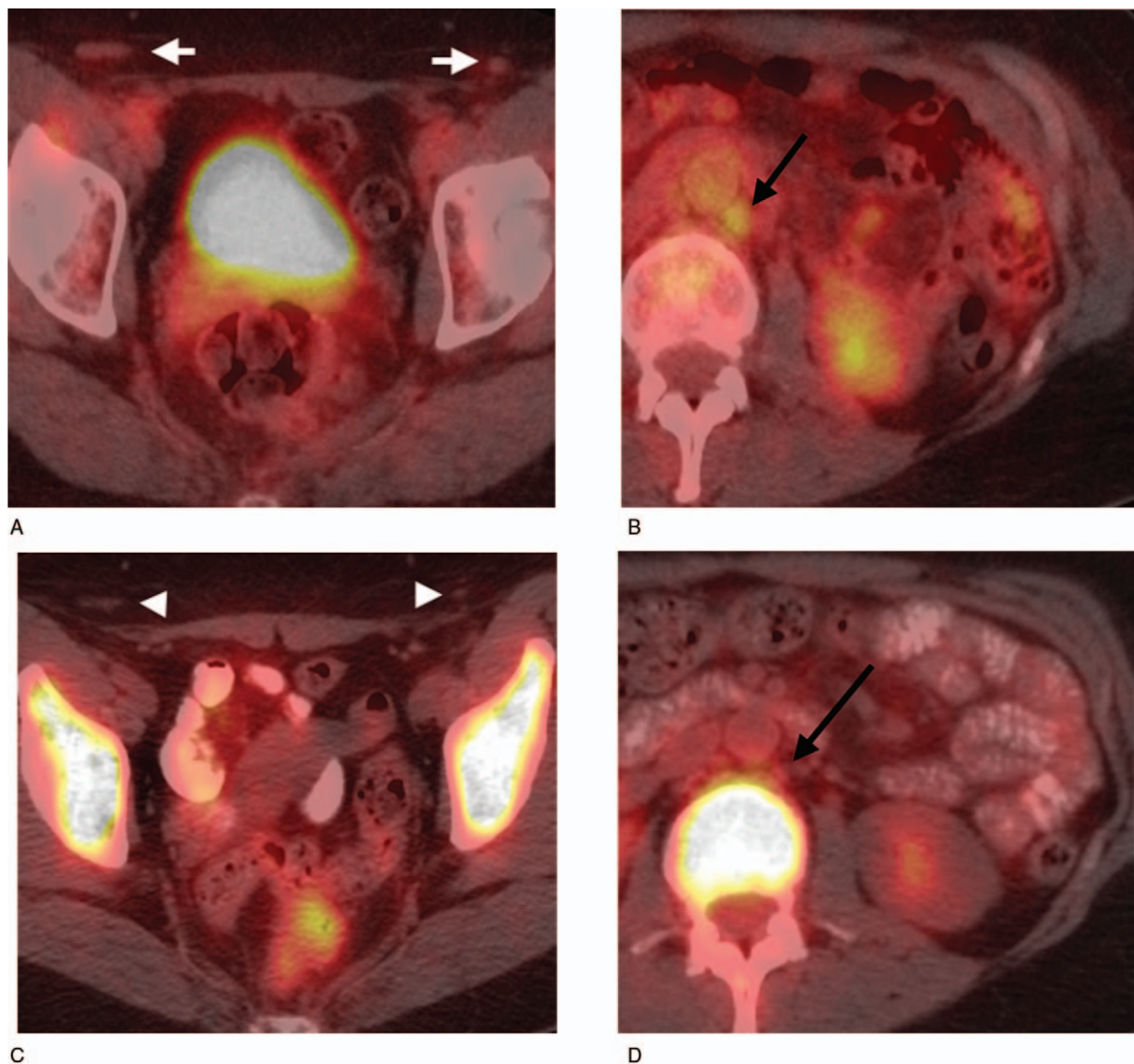


Figure 3. A 58-year-old female with stage I DLBCL in the left cervical lymph nodes on chemotherapy (no access on the initial PET/CT). Interim PET/CT was performed at an outside institution after 4 cycles of chemotherapy. The outside report indicated resolution of left cervical adenopathy and suspicious inguinal (white arrowheads) (A) and left para-aortic (black arrow) (B) adenopathy. Second-opinion review at our institution reported these nodes as “probably reactive” and the staging as “no evidence of disease.” This staging was concordant with the multidisciplinary staging, and the follow-up PET/CT shows resolution of the inguinal nodes (white arrows) (C) and left paraaortic nodes (black arrow) (D).

good example, in which second-opinion review demonstrated the exact and correct anatomical location of a hypermetabolic lesion wrongly described on outside PET/CT report, therefore changing initial disease stage. Figure 3 illustrates the case of a stage I DLBCL patient with interim outside PET/CT describing reactive lymph nodes as sites of disease.

3.3. Comparison to the standard of reference

Compared with the multidisciplinary staging (Table 3), 56/72 of second opinion reports (78%), were in agreement with this ultimate standard of reference, with an almost perfect agreement ($K=0.860$). In contrast, only 40/72 (55%) of outside reports were found to be concordant with the standard of reference, with a fair agreement ($K=0.513$). As an example, Figures 4 and 5

represent a common pitfall and a challenge for every PET/CT reader to discriminate between reactive versus malignant bone marrow uptake.

Focusing specifically on the 32 reports for which discrepancies were noted between outside and second opinion reports, the second-opinion reviews showed a moderate agreement ($K=0.699$) with the standard of reference and 18 of these 32 reports (56%) were concordant with the results of the multidisciplinary staging (Table 4).

4. Discussion

Recent advances in imaging technology, image interpretation criteria, and clinical management of patients with lymphoma place increasing responsibility on radiologists to provide precise

Table 3**Agreement between outside and second-opinion staging as compared with the standard of reference multidisciplinary staging.**

Outside staging	Multidisciplinary staging					Weighted kappa (95% CI)
	NED	I	II	III	IV	
NED	11 (15%)	6 (8%)	4 (6%)	3 (4%)	4 (6%)	0.513 (0.318, 0.697)
I	1 (1%)	4 (6%)	0 (0%)	0 (0%)	3 (4%)	
II	1 (1%)	3 (4%)	5 (7%)	0 (0%)	0 (0%)	
III	0 (0%)	1 (1%)	0 (0%)	8 (11%)	4 (6%)	
IV	1 (1%)	1 (1%)	0 (0%)	0 (0%)	12 (17%)	

Second opinion staging	Multidisciplinary staging					Weighted kappa (95% CI)
	NED	I	II	III	IV	
NED	14 (19%)	3 (4%)	2 (3%)	0 (0%)	1 (1%)	0.860 (0.685, 0.925)
I	0 (0%)	11 (15%)	0 (0%)	3 (4%)	0 (0%)	
II	0 (0%)	1 (1%)	7 (10%)	1 (1%)	1 (1%)	
III	0 (0%)	0 (0%)	0 (0%)	7 (10%)	4 (6%)	
IV	0 (0%)	0 (0%)	0 (0%)	0 (0%)	17 (24%)	

CI= confidence interval, NED=no evidence of disease.



Figure 4. MIP image of a 65-year-old male with DLBCL on chemotherapy. Interim PET/CT was performed at an outside institution. The outside report indicated "no evidence of disease." A second-opinion review at our institution reported heterogenous uptake in the bone marrow (arrowheads) suspicious for residual lymphomatous infiltration, concordant with the multidisciplinary staging as "stage IV" and with bone marrow biopsy results.



Figure 5. MIP image of an 84-year-old female with DLBCL on chemotherapy. PET/CT was performed at an outside institution after completion of treatment. The initial report indicated diffuse bone marrow uptake suspicious for lymphomatous involvement (stage IV). Second-opinion review at our institution described the diffuse bone marrow uptake as reactive to endogenous and/or exogenous stimulation. Therefore, the PET/CT was reported as "no evidence of disease," concordant with the bone marrow biopsy and the multidisciplinary staging as "complete remission."

Table 4**Agreement between second-opinion staging and the multidisciplinary staging for reports where discordant results were noted between outside and second opinion reports.**

Second-opinion Staging	Multidisciplinary staging					Weighted kappa (95% CI)
	NED	I	II	III	IV	
NED	3 (9%)	3 (9%)	2 (6%)	0 (0%)	1 (3%)	0.699 (0.402, 0.865)
I	0 (0%)	7 (22%)	0 (0%)	2 (6%)	0 (0%)	
II	0 (0%)	1 (3%)	2 (6%)	0 (0%)	1 (3%)	
III	0 (0%)	0 (0%)	0 (0%)	1 (3%)	4 (12%)	
IV	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (16%)	

CI=confidence interval, NED=no evidence of disease.

disease evaluation to deliver the best patient care. The diagnostic benefits of FDG PET/CT in initial staging and response evaluation have been established.^[12] Residual FDG uptake on PET/CT scans obtained during and after first line therapy is also associated with patient prognosis.^[13] A meta-analysis reported a sensitivity and specificity of 72% (95% CI, 61% to 82%) and 100% (95% CI, 97% to 100%) for the detection of residual disease after completion of first-line therapy for aggressive non-Hodgkin lymphomas.^[14] Modern guidelines for the application and interpretation of FDG PET/CT in lymphoma were formulated at the 11th International Conference on Malignant Lymphomas^[5] and are now widely applied by oncologists in clinical practice. To ensure optimal patient management, it is imperative that oncologists and radiologists/nuclear medicine physicians “speak the same language” when interpreting and characterizing findings on diagnostic imaging studies, in particular with regard to FDG PET/CT. Based on general impression, discussions in interdisciplinary tumor boards and prior experience with second opinion review of pathology data generated at our center,^[7] we hypothesized that interpretation of FDG PET/CT by expert readers specialized in oncologic imaging would provide added benefit to the oncologist treating lymphoma patients. Indeed, our study demonstrates that interpretation of FDG PET/CT by expert readers and generation of standardized reports can improve the accuracy of disease characterization and the quality of reporting. Clarity of language, application of clear criteria, reporting of data points necessary for proper staging and response assessment (such as lesion size, SUVmax, and changes in these during therapy) are clearly appreciated by oncologists, and are in fact critical for proper patient management.

Our results showed major discordance between second-opinion and outside reports for majority of the studied criteria (Table 1). For example, the lack of agreement for accurate anatomic description of anatomical sites of disease makes it difficult for clinicians to judge the extent of disease at staging, and also to compare findings at baseline with findings upon follow-up. In addition, outside report often lacked confidence in differentiating between benign and malignant FDG uptake on scans, and often did not elaborate on potential differential diagnoses for FDG uptake in a particular location, such as lung, skeletal muscle, or bone marrow. This uncertainty would have led to further (often unnecessary) radiologic imaging studies to work-up “equivocal” sites of FDG uptake, leading to potential delay in care, and additional patient exposure and cost. Such additional studies were recommended rarely by second opinion readers.

For follow-up scans, residual FDG uptake is characterized using modern scoring systems,^[1,15] which rely on comparison with reference regions in the same scan and reporting of SUV numbers. We found only a slight agreement in the ability to derive

a particular score from the 5-point scales when comparing outside and second opinion reports ($K=0.024$) because most of the outside reports failed to provide the lesion with the highest SUVmax and the SUVmax at sites of residual disease ($K=-0.016$) or failed to include the comparison with FDG uptake in reference regions in mediastinum and liver ($K=0.082$).

Interestingly, outside and second-opinion reports were also discordant regarding disease stage in 26 of the 72 patients (36%) (weighted $K=0.674$), with sometimes considerable potential impact on patient management (eg, as shown in Figs. 4 and 5).

Again, second-opinion interpretations were found to be more accurate and clinically relevant. Oncologists appreciated the fact that second-opinion reports were more structured, with 97% of reports receiving a grade of 4 or 5 as compared with outside reports in only 44% ($P<.001$). Some of the limitations noted in outside reports could be remedied easily by local radiologists and nuclear medicine physician reading PET scans if they were familiar with patterns of disease spread and differential diagnoses, and applied contemporaneous criteria to their interpretation of scans.

In addition, the use of structured and standardized reporting is gaining favor because of its reproducibility and easy read by the referring physicians. Schwartz et al^[16] showed that structured radiology reporting might improve patient care by increasing clarity and thoroughness in the communication of imaging findings to referring physicians. When such reports are developed with input from interdisciplinary clinical teams, they receive significantly higher mean ratings for clarity and content by referring physicians when compared with conventional (unstructured, free flowing language) radiology reports. Therefore, we have been using structured reporting for several years in our institution.

Previous studies have looked at the value of second opinion radiology reports in other diseases and have generally found that these reports are highly valued. For example, Lakhman et al^[17] demonstrated that second-opinion review of gynecological oncologic MRI affected management in about 20% of patients. Unnecessary surgeries were prevented in about 7% of cases. When compared with histopathology and 6-month follow-up as reference, the second-opinion MRI reviews were correct in 83% of cases with clinically relevant discrepancies. Similarly, Gollub et al^[18] found major discordance between outside and second-opinion radiology reports in 24 of 143 body CT scans, with a subsequent change in management for 3.5% of patients. Zan et al^[19] evaluated 7,465 neuroradiology studies and showed clinically important discrepancy in interpretation 7.7% of cases (in 2/3 of cases related to lesion detection, and in 1/3 of cases related to patient management). The second-opinion review was more accurate in 84% of studies when compared with the definitive diagnosis. Finally, Dudley et al^[20] compared the initial

outside and second-opinion CT and MRI reports for 396 patients who were referred to surgical oncologists for initial staging or follow-up and concluded that reports disagreed in 41% of cases; the second-opinion interpretations were more often correct (in 94% of discrepant cases). With regard to PET/CT in general, Tahari et al^[21] showed that submitted outside PET/CT exams often lacked technical information needed for appropriate scan interpretation. In another analysis,^[22] it was shown that subspecialist review of submitted outside PET/CT examinations resulted in discrepant interpretation in 13% of cases in characterizing FDG uptake as benign or malignant, and the subspecialist review was accurate in 89% of cases.

Our analysis has certain limitations, mainly its retrospective nature. The subspecialized oncoradiologists had access to the original outside report when generating the second-opinion report. Second, the number of studies reviewed was relatively small. Even though we recruited all lymphoma studies submitted over the course of 3 years, the necessity for reasonably long follow-up and ability to review and measure scans with our institutional hardware and software tools limited the initially larger number of potential scans to only 72. Third, we used the agreement in the multidisciplinary meeting and follow-up as standard of reference as biopsy confirmation for each individual lesion is not possible. The hemato-oncologist grading report quality was not aware of the origin of each individual report. However, as we have used structured reporting for some time it is likely that reports generated by second-opinion readers could thus be possibly recognized. Nevertheless, this would not necessarily lead to higher rating of inhouse reports as the oncologist was asked to address a number of clinically relevant criteria in the reports. Another limitation is that the radiology reports were evaluated by a single oncologist from the same institution of the second opinion radiologist; however, the oncologist was blinded to the source of the reports.

Finally, excluding all DLBCL patients whose submitted PET/CT scans lacked the original outside report may have led to selection bias.

In conclusion, second-opinion review of FDG PET/CT in lymphoma patients by subspecialized oncologic radiologists increases accuracy of staging and response evaluation and the clinical significance of the report. For ease of interpretation and comparison with follow-up scans, oncologists favor structured reports, with accurate description of disease sites and disease activity using contemporaneous interpretation criteria.

References

- [1] Johnson SA, Kumar A, Matasar MJ, et al. Imaging for staging and response assessment in lymphoma. *Radiology* 2015;276:323–38.
- [2] Shelly MJ, McDermott S, O'Connor OJ, et al. 18-Fluorodeoxyglucose positron emission tomography/computed tomography in the management of aggressive non-Hodgkin's B-cell lymphoma. *ISRN Hematol* 2012;2012:456706.
- [3] Tatsumi M, Cohade C, Nakamoto Y, et al. Direct comparison of FDG PET and CT findings in patients with lymphoma: initial experience. *Radiology* 2005;237:1038–45.
- [4] Terasawa T, Nihashi T, Hotta T, et al. 18F-FDG PET for posttherapy assessment of Hodgkin's disease and aggressive non-Hodgkin's lymphoma: a systematic review. *J Nucl Med* 2008;49:13–21.
- [5] Barrington SF, Mikhaeel NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *J Clin Oncol* 2014;32:3048–58.
- [6] Cheson BD. Staging and response assessment in lymphomas: the new Lugano classification. *Chin Clin Oncol* 2015;4:5.
- [7] Matasar MJ, Shi W, Silberstien J, et al. Expert second-opinion pathology review of lymphoma in the era of the World Health Organization classification. *Ann Oncol* 2012;23:159–66.
- [8] Eakins C, Ellis WD, Pruthi S, et al. Second opinion interpretations by specialty radiologists at a pediatric hospital: rate of disagreement and clinical implications. *AJR. AJR Am J Roentgenol* 2012;199:916–20.
- [9] Lu MT, Tellis WM, Avrin DE. Providing formal reports for outside imaging and the rate of repeat imaging. *AJR Am J Roentgenol* 2014;203:107–10.
- [10] Bell ME, Patel MD. The degree of abdominal imaging (AI) subspecialization of the reviewing radiologist significantly impacts the number of clinically relevant and incidental discrepancies identified during peer review of emergency after-hours body CT studies. *Abdom Imaging* 2014;39:1114–8.
- [11] Hutchings M, Barrington SF. PET/CT for therapy response assessment in lymphoma. *J Nucl Med* 2009;50(Suppl 1):21–30.
- [12] Barrington SF, Kirkwood AA, Franceschetto A, et al. PET-CT for staging and early response: results from the response-adapted therapy in Advanced Hodgkin Lymphoma study. *Blood* 2016;127:1531–8.
- [13] Schot B, van Imhoff G, Pruim J, et al. Predictive value of early 18F-fluorodeoxyglucose positron emission tomography in chemosensitive relapsed lymphoma. *Br J Haematol* 2003;123:282–7.
- [14] Zijlstra JM, Lindauer-van der Werf G, Hoekstra OS, et al. 18F-fluorodeoxyglucose positron emission tomography for post-treatment evaluation of malignant lymphoma: a systematic review. *Hematologica* 2006;91:522–9.
- [15] Itti E, Meignan M, Berriolo-Riedinger A, et al. An international confirmatory study of the prognostic value of early PET/CT in diffuse large B-cell lymphoma: comparison between Deauville criteria and DeltaSUVmax. *Eur J Nucl Med Mol Imaging* 2013;40:1312–20.
- [16] Schwartz LH, Panicek DM, Berk AR, et al. Improving communication of diagnostic radiology findings through structured reporting. *Radiology* 2011;260:174–81.
- [17] Lakhman Y, D'Anastasi M, Micco M, et al. Second-opinion interpretations of gynecologic oncologic MRI examinations by sub-specialized radiologists influence patient care. *Eur Radiol* 2016;26:2089–98.
- [18] Gollub MJ, Panicek DM, Bach AM, et al. Clinical importance of reinterpretation of body CT scans obtained elsewhere in patients referred for care at a tertiary cancer center. *Radiology* 1999;210:109–12.
- [19] Zan E, Yousem DM, Carone M, et al. Second-opinion consultations in neuroradiology. *Radiology* 2010;255:135–41.
- [20] Dudley RA, Hricak H, Scheidler J, et al. Shared patient analysis: a method to assess the clinical benefits of patient referrals. *Med Care* 2001;39:1182–7.
- [21] Tahari AK, Wahl RL. Quantitative FDG PET/CT in the community: experience from interpretation of outside oncologic PET/CT exams in referred cancer patients. *J Med Imaging Radiat Oncol* 2014;58:183–8.
- [22] Ulaner GA, Mannelli L, Dunphy M. Value of second-opinion review of outside institution PET-CT examinations. *Nucl Med Commun* 2017;38:306–11.