perceptive imaging

Insight Paper

# Impact of PET on Imaging Related Endpoints in Lymphoma Trials

## Background

The first widely accepted imaging and clinical response criteria for non-Hodgkin lymphoma (NHL), were the International Working Group (IWG)–NHL 1999 criteria *(Cheson, 1999)*. The IWG-NHL 1999 criteria were largely dependent on computed tomography (CT), single photon emission CT gallium scans, and qualitative bone marrow assessment—the common technologies at that time.

While these criteria were broadly adopted by clinicians and regulatory agencies and used in regulatory approval for several new therapeutic agents, they were subject to high inter-observer and intra-observer variations. Other limitations of the criteria included non-inclusion of the extranodal disease in response assessment and misinterpretations of residual tumor masses on CT.

To address the limitations of the IWG-NHL 1999 criteria and the subsequent increased use of [18F] 2-fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET), bone marrow immunohistochemistry (IHC) and flow cytometry in clinical trials, the IWG published new guidelines–IWG-NHL 2007 criteria *(Cheson, 2007)*. These criteria recommended integration of PET with CT for tumor response assessments because of better sensitivity and superiority of PET over CT. The major advantage of PET over CT or magnetic resonance imaging (MRI) alone is its ability to distinguish between viable tumor and necrosis or fibrosis in residual mass after treatment *(Cheson, 2018)*.

While determining PET positivity of lesions provided more accurate tumor response assessments than using CT alone, there was a potential for discordance in the interpretation of lesion positivity on PET, primarily due to a dichotomous PET assessment (i.e., positive versus negative) based on a subjective interpretation of what represents "background" for FDG uptake (i.e., mediastinal blood pool versus adjacent regions) and the subjective judgement of what represents significant uptake compared to the background. To provide further clarification on these limitations, the IWG-NHL 2007 criteria were further refined into the Lugano Classification guidelines, which recommend the Deauville 5-point scale for interpretation of PET *(Cheson, 2014; Barrington, 2014; Cheson, 2017)*.

## PET and the Lugano Classification

The goal of the Lugano Classification was to achieve more consistent and uniform therapeutic response assessments for subjects enrolled in clinical trials evaluating treatment for lymphoma. The most significant aspects of the Lugano Classification pertain to the folowing components:

- Replacement of the dichotomous evaluation of FDG uptake (positive versus negative) with a 5-point scale assessment for interim and end of treatment (EOT) analyses
- All FDG-avid disease for the applicable lymphomatous indications present in the individual subject is assessed
- Use of PET-CT in the assessment of FDG-avid lymphoma, while use of CT alone in the assessment of non-FDG-avid lymphomas

The incorporation of PET as the predominant imaging modality for measuring the distribution and extent of disease in FDG-avid lymphomas represents a major paradigm shift, as this approach moves away from a pure anatomic size-based response into a physiologic response assessment based on tumor metabolism. This approach allows for a more accurate early assessment of lymphoma treatment response.

In years after publication of the Lugano Classification, PET-CT has been used as a preferred modality for staging and assessing clinical response in FDG-avid lymphomas based on its ability to differentiate between viable tumor versus necrosis or fibrosis after therapy, and a great sensitivity for bone marrow evaluation. However, despite its huge benefits, PET assessments are associated with limitations such as false positive findings due to infection, inflammation, and sarcoidosis. False negative findings on PET can also result due to low resolution of the equipment and technique used.

The established criteria provide an explanation on many aspects of the use of PET in clinical trials, the guidelines on the use of PET for imaging endpoints of a clinical trial and frequency of PET imaging in clinical trials. However, some aspects like tumor response assessments when PET is not available for a timepoint and incorporating PET with bone marrow biopsy/aspirate data are not clearly defined. In this document, we further clarify the above topics related to the use of PET in clinical trials for lymphoma.

## Impact of PET on imaging endpoints in clinical trials

Omar et al *(Omar, 2016)* studied 50 subjects with pathologically proven lymphoma. Contrast enhanced (CE) CT and PET-CT were performed for all subjects for initial staging, during the course of chemotherapy and at the EOT. The tumor assessments on PET-CT and CE CT were agreeable in 75% of cases overall, in 61% cases during treatment, and in 41% cases at the EOT. PET-CT showed higher sensitivity and specificity over CE CT. The major strength of PET-CT over CE CT was its higher ability for detection of extra nodal sites of lymphoma and confirming residual nodal mass lesions at follow-up to be metabolically inactive disease.

A retrospective analysis of 10 lymphoma studies using both PET and CT, with a total of 1,537 subjects and 17,394 timepoints demonstrated significant impact of PET on the overall tumor assessment and clinical trial endpoints *(Sharma, 2019 suppl)*. Five (5) clinical trials with a total of 1,078 subjects with 15,480 timepoints used IWG-NHL 2007 criteria, while the other 5 trials with a total of 459 subjects and 1,914 timepoints used the Lugano Classification. The number of subjects and timepoints with PET and the impact of PET on the overall tumor assessment was calculated (Table 1).

| | TOTAL # | # WITH PET | IMPACT ON ASSESSMENT |
|---|---|---|---|
| Subjects | 1,537 | 1,159 (75.4%) | 462 (39.9%) |
| Timepoints | 17,394 | 4,688 (26.9%) | 956 (20.4%) |

Table 1: Impact of PET on Overall Tumor Assessment

PET was available for 1,159 out of 1,537 subjects (75.4%) and 4,688 out of 17,394 timepoints (26.9%) across all 10 studies. Out of 1,159 subjects with PET, in 462 subjects (39.9%) PET had an impact on the overall tumor assessment and study endpoints due to metabolic uptake not possible to detect on CT imaging. Out of 4,688 timepoints with PET, for 956 (20.4%) timepoints there was an impact of PET on the overall tumor assessment due to information associated with metabolic uptake.

The impact of PET on the subjects and endpoints across different studies was between 7% to 59%. This high variability is possibly due to difference in indication, criteria used, line of therapy and blinded independent review design. The most common impact of PET was on the complete response and disease progression related study endpoints.

## PET for evaluation of FDG-avid versus non-FDG-avid lymphomas in clinical trials

PET-CT has been shown to be a very promising and effective tool for disease staging and assessing response to treatment, especially for FDG-avid lymphomas (e.g., Hodgkin lymphoma (HL), diffuse large B-cell lymphoma (DLBCL)). CT-only evaluation is preferred for staging and response assessment of low or variably FDG-avid histologies (e.g. marginal zone lymphoma, small lymphocytic lymphoma).

A positive PET of bone or bone marrow is adequate to designate advanced stage in diffuse DLBCL. However, bone marrow biopsy can be considered in DLBCL when PET does not show an evidence of bone marrow involvement, particularly if identification of the discordant histology is relevant for patient management or if the results would alter treatment *(Cheson, 2015)*.

## PET-based 5 point scoring assessments for FDG-avid lymphomas

The Lugano Classification recommends the use of PET-based Deauville 5-point scoring for clinical trials, which is based on comparison of tumor lesion FDG uptake on PET with the mediastinal blood pool as well as the liver. The standardized uptake values (SUV) serve as marker of metabolic activity. This scoring represents a significant improvement over assessment based only on visual comparison. Response assessment is categorized according to the 5-point scale, which includes the following scores *(Johnson, 2015)*:

**Score 1:** No FDG uptake > background

**Score 2:** FDG uptake ≤ mediastinum

**Score 3:** FDG uptake > mediastinum but ≤ liver

**Score 4:** FDG uptake moderately > liver

**Score 5:** FDG uptake markedly > liver and/or new lesions.

Though the assessment appears more complex with potential for higher variability, Heertum et al *(Heertum, 2017)* showed that the assessment of FDG metabolic activity in lymphomatous lesions using Deauville 5-point scale showed reproducible results between readers with a high concordance rate (kappa 0.73 [confidence interval 0.59–0.87; P=,0.0001]). The results show that 5-point scale is not just more accurate but also operationally practical for use in larger lymphoma clinical trials.

In interpreting the 5-point scale, a score of 1 or 2 is considered negative for lymphoma, while a score of 4 or 5 is considered positive. A score of 3 at interim suggests good prognosis for a low risk disease and is therefore usually considered to be negative for lymphoma and no further treatment is necessary (e.g., Follicular Lymphoma (FL), Mantle Cell Lymphoma (MCL), etc.). A score of 3 for a high-risk disease that is curable and aggressive (e.g., HL and DLBCL) is typically considered positive for lymphoma. Other factors to consider when assessing score 3 are lymphoma FDG avidity and treatment under evaluation. A score of 3 in many patients indicates a good prognosis with standard treatment, especially at interim. However, in trials involving PET where de-escalation is investigated, it may be preferable to consider a score of 3 as an inadequate response, to avoid undertreatment *(Barrington, 2014; Cheson, 2014; Biggi, 2013)*.

Since score 3 can be considered as PET positive or PET negative, it can have an impact on the overall study endpoints. For example, for a clinical trial with progression-free survival (PFS) as the primary endpoint, a PET-positive score 3 may result in lower PFS compared to a PET-negative score 3. On the other hand, considering score 3 as PET-negative might result in higher relapse rates thereby affecting the PFS endpoint. To assess the importance of score 3 in response assessments, we evaluated the PET 5-point scoring for 583 subjects with a total of 1,492 PET timepoints in 3 lymphoma clinical trials. Of the 366 subjects, 35 (~9.56%) showed a score of 3 (Table 2).

| TOTAL NUMBER OF SUBJECTS | SUBJECTS WITH 5-POINT ASSESSMENT | ALL TIMEPOINTS WITH A 5-POINT ASSESSMENT | ALL SUBJECTS WITH A 5-POINT SCORE OF 3 | ALL TIMEPOINTS WITH A 5-POINT SCORE OF 3 | % SUBJECTS WITH A SCORE OF 3 |
|---|---|---|---|---|---|
| 583 | 366 | 1,492 | 35 | 55 | 9.56 |

Table 2: Score 3 in Lymphoma Clinical Trials

## Frequency of PET acquisition in clinical trials for FDG-avid lymphomas

Frequency of PET acquisition should depend on the lymphoma subtype and scope of the trial. Baseline/staging and EOT PET scans are recommended for all FDG-avid lymphomas.

Eichenauer et al *(Eichenauer, 2018)* suggest that all subjects with FDG-avid lymphoma should undergo interim staging using PET 2-3 cycles after the initiation of treatment to exclude false assessment of disease progression during treatment and to stratify treatment.

Interim PET-CT in advanced HL may prove valuable in predicting disease outcome. Several prospective, risk-adapted trials have demonstrated that the use of PET-CT can improve decision making and consequently the outcome in high-risk *(interim PET–positive)* subjects and can limit the amount of treatment in lower-risk (interim PET–negative) subjects (Cheson, 2017).

The below points can be considered when determining the frequency of PET acquisition:

- For studies with overall response rate (ORR) as a primary endpoint, PET is mandatory at screening, optional at interim and again required 6-8 weeks after completion of treatment. In addition, during treatment PET may also be acquired to confirm a complete response assessed on CT alone.
- For studies with PFS as a primary endpoint, PET is mandatory at screening, can be considered at interim and 6-8 weeks after completion of treatment.
- CT scan can suffice for PFS surveillance. Surveillance PET scans are not recommended due to high false positive rates leading to unnecessary investigations, radiation exposure, biopsies, expense, and subject anxiety. Follow-up scans may be done in case of equivocal EOT, high risk of relapse or clinical suspicion.

## Assessment in case of missing PET

While CT scans are acquired regularly during a study, PET frequency is generally variable. As a result, PET may not be available for correlation of disease burden assessed on CT for all timepoints, which may cause incorrect comparison of disease assessments between timepoints with and without PET. To maintain uniformity of assessments across timepoints, for the timepoints at which PET is not available but CT is, the overall tumor assessment from the previous timepoint or baseline with PET can be carried forward provided the disease burden on CT remains stable or improves. A few examples are shown in Table 3 and Table 4. In Table 3, PET is not available at Timepoint 3. However, disease burden on CT has remained stable from Timepoint 2 and has not worsened. Therefore, at Timepoint 3, PET 5-point scoring can be considered to have

remained stable and an overall tumor assessment of complete metabolic response (CMR) is possible. On the other hand, in Table 4, while PET is not available at Timepoint 3, CT at Timepoint 3 shows worsening disease compared to Timepoint 2. In this scenario, the previous overall tumor response may not be carried forward and will be progressive metabolic disease (PMD).

| TIMEPOINT | CT ASSESSMENT | PET 5-POINT SCORING | OVERALL TUMOR ASSESSMENT |
|---|---|---|---|
| Baseline | – | 2 | – |
| Timepoint 2 – Week 12 | SD | 2 | CMR |
| Timepoint 3 – Week 24 | SD | Not available | CMR |

Table 3: Example of Overall Tumor Assessment when PET is Missing and CT Shows Stable Disease Burden

| TIMEPOINT | CT ASSESSMENT | PET 5-POINT SCORING | OVERALL TUMOR ASSESSMENT |
|---|---|---|---|
| Baseline | – | 2 | – |
| Timepoint 2 – Week 12 | SD | 2 | CMR |
| Timepoint 3 – Week 24 | PD | Not available | PMD |

Table 4: Example of Overall Tumor Assessment when PET is Missing and CT Shows Worsening Disease

### Correlation between bone marrow assessment on biopsy and PET

The Lugano Classification suggests that PET-CT alone is adequate for determination of bone marrow involvement, and bone marrow biopsy is not required, especially for HL and DLBCL. To compare the bone marrow assessments on biopsy versus PET, we retrospectively analyzed 140 subjects with FDG-avid lymphoma from 3 studies, all reviewed using the Lugano Classification. Out of 140 subjects, 23 (16%) had discordance in the baseline bone marrow assessment on biopsy and PET (Table 5). 9 subjects (6%) were assessed as positive for bone marrow involvement on biopsy but showed no focal bone marrow uptake on PET. This suggests that there could be a discordance between bone marrow biopsy and PET assessment. Therefore, in a clinical trial setting, if discordance for bone marrow is identified between PET and biopsy at baseline; at CR confirmation, the parameter that was positive at baseline should be negative. If biopsy was positive and no focal uptake on PET was identified at baseline, biopsy should be repeated at the time of CR confirmation.

| TOTAL N = 140 | BIOPSY POSITIVE | BIOPSY NEGATIVE |
|---|---|---|
| Focal Bone Marrow uptake on PET | 8 | 14 |
| No Focal Bone Marrow uptake on PET | 9 | 109 |

Table 5: Bone Marrow Assessment on Biopsy versus PET

## Conclusion

Per the Lugano Classification PET-CT is considered as the standard for disease staging and response assessments for all FDG-avid histologies. Use of PET-CT in imaging of FDG-avid histologies has further improved baseline staging and facilitated functional evaluation of disease behavior, metabolic response to therapy, and earlier detection of disease recurrence.

In a clinical trial setting, considerations on the impact of the frequency of PET acquisition and bone marrow biopsy on overall timepoint assessments can prove to be critical for ensuring uniform comparison between visits and study endpoints.

perceptive™ imaging

perceptive.com
contact us at: hello@perceptive.com


1. Cheson BD, Horning SJ, Coiffer B, et al. J Clin Oncol, 1999, 17(4), 1244.
2. Cheson BD, Pfistner B, Juweid ME, et al. J Clin Oncol, 2007, 25(5), 579-86.
3. Cheson BD. Semin Nucl Med, 2018, 48(1), 76-81.
4. Cheson BD, Fisher RI, Barrington SF, et al. J Clin Oncol, 2014, 32(27), 3059-3067.
5. Barrington SF, Mikhaeel NG, Kostakoglu L, et al. J Clin Oncol, 2014, 32(27), 3048-58.
6. Cheson BD. Chin Clin Oncol, 2015, 4(1), 5.
7. Omar NN, Alotafy LM, Abolela MS, et al. Egypt J Rad and Nucl Med, 2016, 47, 1639–1647.
8. Eichenauer DA, Aleman BMP, Andre M, et al. Ann Oncol, 2018, 29(4).
9. Biggi A, Gallamini A, Chauvie S, et al. J Nucl Med, 2013, 54(5), 683-90.
10. Sharma M, Bohnsack O, Clark D et al. ASCO J Clin Oncol, 2019 suppl.
11. Johnson, SA, Kumar A, Matasar MJ et al. Radiology, 2015, 276 (2), 323-38.
12. Heertum vRL, Scarimbolo R, Wolodzko JG et al. Drug Des Devel Theor, 2017 13 (11), 1719-1728.