

 Open Access Full Text Article

ORIGINAL RESEARCH

# Lugano 2014 criteria for assessing FDG-PET/CT in lymphoma: an operational approach for clinical trials

Ronald L Van Heertum<sup>1</sup>Robert Scarimbolo<sup>1</sup>John G Wolodzko<sup>1</sup>Barbara Klencke<sup>2</sup>Richard Messmann<sup>3</sup>Feza Tunc<sup>4</sup>Levi Sokol<sup>4</sup>Rajan Agarwal<sup>5</sup>James A Strafaci<sup>1</sup>Michael O'Neal<sup>1</sup>

<sup>1</sup>Medical Affairs, Bioclinica, Inc., Princeton, NJ, <sup>2</sup>Medical Affairs, Sierra Oncology, Brisbane, CA, <sup>3</sup>Medical Affairs, Apexian Pharmaceuticals, Indianapolis, IN, <sup>4</sup>Radiology, University Radiology at RVJ University Hospital, New Brunswick, NJ, <sup>5</sup>Radiology, Abington Hospital, Abington, PA, USA

**Abstract:** An operationalized workflow paradigm is presented and validated with pilot subject data. This approach is reproducible with a high concordance rate between individual readers ( $\kappa$  0.73 [confidence interval 0.59–0.87;  $P < 0.0001$ ]) using a 5-point scale to assess [ $^{18}\text{F}$ ] labeled fluorodeoxyglucose metabolic activity in lymphomatous lesions. These results suggest an operationally practical 5-point scale workflow paradigm for potential use in larger clinical trials evaluating lymphoma therapeutics.

**Keywords:** lymphoma, Lugano criteria, molecular imaging, oncology trials

## Introduction

Lymphoma, typically categorized as either non-Hodgkin's lymphoma (NHL) or Hodgkin lymphoma (HL), is the most common hematological malignancy in the US and accounts for approximately 5% of all newly diagnosed cancers (according to the National Cancer Institute [NCI], 2016). In 2016 in the US, approximately 72,580 new cases of NHL and 8,500 new cases of HL were diagnosed.<sup>1,2</sup>

While classified under the general heading of lymphoma, NHL and HL, as well as subtypes within each histologic category, may differ in surface protein expression, histologic appearance, cell of origin, clinical evolution, response to treatment, and other features. These differences are yielding important insights into the natural biology of lymphoma, as well as potential markers for diagnostic and therapeutic development. A compelling example is the successful development of CD20-targeted therapy for management of a broad variety of lymphomatous and hematologic diseases.<sup>3</sup>

Similarly, there are a number of newer diagnostic imaging approaches available, one of which includes [ $^{18}\text{F}$ ] fluorodeoxyglucose (FDG) positron emission tomography (PET) imaging, to better distinguish the lymphoma subtypes. This approach visually assesses the metabolic activity of lymphoma by three-dimensionally measuring the uptake distribution post-administration of FDG.<sup>4</sup> In addition to identifying the presence and distribution of disease, FDG-PET imaging, particularly when combined with high quality computed tomography (CT) imaging (PET/CT), has also been shown to be a very effective tool for assessing response to treatment.<sup>5</sup>

Different sub-types of lymphomas exhibit varying degrees of FDG-avidity that correlate with the aggressiveness of the individual lymphoma.<sup>6</sup> Previous investigations have established that NHL exhibits a varying FDG-avidity range from 40% to 100%, depending on the lymphoma subtype, while HL exhibits a much narrower FDG-avidity range of 97% to 100% (Table 1).<sup>7</sup> Although the use of FDG-PET/CT for the assessment

Correspondence: Ronald L Van Heertum  
BioClinica, Inc., 100 Overlook Center,  
Princeton, NJ 08540, USA  
Tel +1 609 936 2868  
Fax +1 609 514 4964  
Email ronald.vanheertum@bioclinica.com

**Table I** Modified Lugano 5-point scale (5PS)

Score	Description
1	No uptake
2	Uptake $\leq$ mediastinum
3	Uptake $>$ mediastinum but $\leq$ liver
4	Uptake moderately increased above liver at any site
5	Markedly increased uptake above liver at any site
NE	Not evaluable
X	Any areas of uptake not likely to be related to lymphoma

**Note:** Data from Cheson et al.<sup>10</sup>

of lymphoma, particularly at the end of treatment (EOT), is supported by published literature, creating a standardized clinically practical methodology for the assessment of lymphoma by FDG-PET/CT continues to be a challenge.<sup>8,9</sup>

The purpose of this paper is to define a proposed operational workflow to improve the efficiency and reproducibility of evaluating FDG-avid lymphomas using PET/CT, following the most current published criteria for assessing treatment response in clinical trials.<sup>10</sup> The workflow methodology for evaluating FDG non-avid lymphomas using CT criteria will not be included in this manuscript.

## Background

In 1999, the NCI Lymphoma International Working Group (IWG) first published imaging and clinical response guidelines for NHL, commonly known as Cheson 1999 criteria.<sup>11</sup> These guidelines formed a standardized approach for assessing the presence of NHL, and measuring response to therapeutic intervention, by evaluating imaging and clinical data. The imaging aspects of the Cheson 1999 criteria were primarily based on CT technology which was widely incorporated in clinical trials at that time. The Cheson 1999 criteria guidelines were then updated in 2007, when the IWG published revised response criteria for malignant lymphoma.<sup>12</sup> The revised Cheson 1999 criteria, more commonly known as the Cheson 2007 criteria, were developed to address limitations of the Cheson 1999 criteria and to incorporate bone marrow (BM) immunohistochemistry, flow cytometry, and the increased use of FDG-PET imaging as a recognized and effective modality for visualizing the presence and distribution of lymphoma. As a result of incorporating FDG-PET, the response designation of “complete response/unconfirmed” permitted in the Cheson 1999 criteria was eliminated for FDG-avid histologies which converted from FDG-“positive” to FDG-“negative” following treatment.

In the Cheson 2007 criteria, these lesions which changed from FDG-positive to FDG-negative following treatment, regardless of residual size on CT, were designated as a complete response. While these changes represented marked improvements to the Cheson 1999 criteria, the Cheson 2007 guidelines did present some challenges to the interpretation of lymphoma progression or response to treatment. Specifically, there was significant potential for ambiguity in the interpretation of lesion positivity due to a dichotomous (ie, positive vs negative) PET response criteria which was based on a subjective interpretation of what represented FDG “background” (ie, blood pool vs adjacent regions) and the degree of significantly discernible uptake, compared to the background.

Despite the challenges, the Cheson 2007 criteria remained the standard for evaluating HL and NHL until 2014 when the most recent revised criteria were published.<sup>10</sup> The evolution of this revised criteria was based on the need to define tumor FDG-avidity with greater objectivity and reproducibility. These new criteria were the direct outgrowth of integrating the previously defined Deauville criteria with the input of investigators at follow-up International Workshop Conferences in 2011 and 2013.<sup>13</sup>

## Essence of the Lugano 2014 guidelines

The goal of the Lugano 2014 guidelines was to revise the Cheson 2007 criteria, in order to reduce ambiguity and achieve more consistent therapeutic response assessments for patients enrolled in clinical trials evaluating treatment for lymphoma. The most significant aspects of the new guidelines pertain to three major components:

- the predominant use of FDG-PET/CT in the assessment of FDG-avid lymphoma, while CT remains the designated standard for assessment of non-FDG-avid lymphomas;
- the replacement of the dichotomous evaluation of FDG uptake (positive vs negative) with a 5-point scale (5PS) assessment for interim and EOT analyses;
- the premise that all FDG-avid disease (for applicable lymphomatous indications) present in the individual patient is included in each time point (TP) analysis.

Other updates in the Lugano 2014 criteria include:

- the discontinuation of routine BM biopsies in HL and FDG-avid NHL;
- the modification of the Ann Arbor staging method;
- the recommendation to reduce the total number of routine follow-up surveillance scan procedures.

The three major components, along with the other modifications to the prior guidance, are intended to help achieve

a more uniform and consistent assessment necessary for multiple TP assessments in clinical trials.

The incorporation of FDG-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 should allow for a more accurate early assessment of lymphoma tumor treatment response; however, there is still some controversy in the literature regarding the use of interim PET assessment of treatment response in the clinical trial setting.<sup>14,15</sup> While the new criteria should reduce the subjective variability which has existed in regards to determination of FDG-PET lesion positivity, this observation has yet to be documented in any multi-center lymphoma clinical trials.

## Challenges to developing an operational reader workflow paradigm for clinical trials

The Lugano 2014 criteria, as defined in the article by Cheson et al, were used as the basis for our approach.<sup>10</sup> Our goal was to develop a reproducible and time-efficient operational paradigm, utilizing the Lugano 2014 criteria, which could be routinely employed in clinical drug trials assessing FDG-avid lymphoma. The Lugano 2014 criteria was studied by the authors and modifications were incorporated into the workflow in order to operationalize this approach, as described in the following section.

## Recommendation for revisions to the staging criteria

At present, FDG-PET/CT is generally accepted as the preferred procedure for the clinical staging of FDG-avid lymphomas.<sup>16,17</sup> In recognition of the wide-spread utilization of FDG-PET/CT and the supporting literature, the IWG recommends that this modality be routinely employed in clinical trials assessing subjects with FDG-avid lymphomatous disease. In addition to the FDG-PET imaging, a contrast-enhanced CT scan should be included at baseline for accurate measurement of lesion size, separation of bowel from lymphadenopathy, and differentiation of vascular structures from lymph nodes. Another significant modification to the staging criteria is the integration of a 5PS to achieve a more accurate assessment of the degree of FDG avidity at baseline and during follow-up, as it relates to the evaluation of treatment response and progression of disease (Table 1).

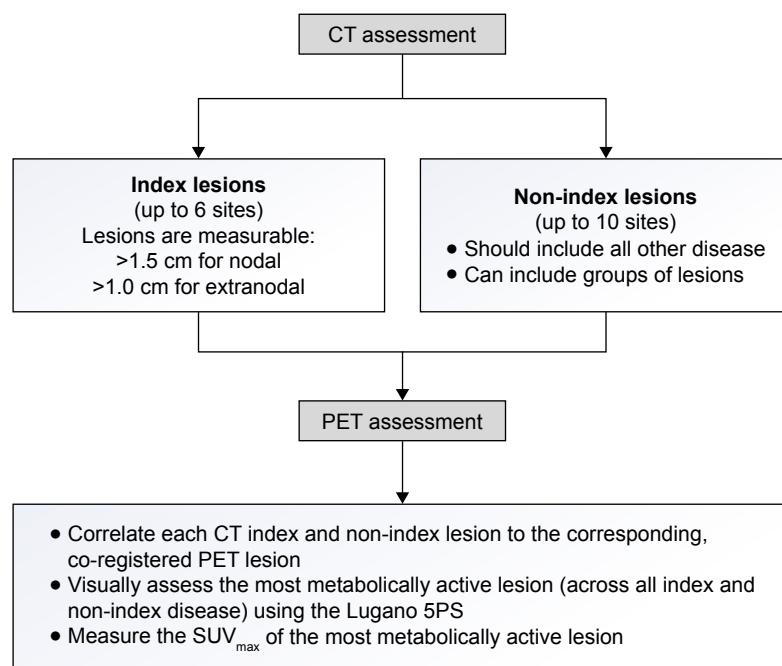
The 5PS ranges from a score of 1 (where no uptake is discernible in the lesion) to a score of 5 (where the uptake in the lesion is markedly increased compared to the uptake in the liver parenchyma). A single 5PS score, which represents the most FDG-avid (ie, metabolically intense) area of disease (across all index and non-index lesions), is assigned for each TP. The designation of “X” in the Lugano 2014 5PS has been removed, since under the proposed operational workflow, the readers are trained to provide a comment in their assessment if there are new areas of observed uptake that are unlikely to be lymphoma.

The assessment of BM according to the Lugano 2014 criteria is also very different from the Cheson 2007 criteria, since BM biopsy (BMB) is no longer required for all patients with FDG-avid lymphomas. Disease involvement in BM can now, in most cases, be solely evaluated using FDG-PET imaging; however, confirmation by BMB is still recommended in certain cases. Some of these include patients with certain FDG-avid lymphoma subtypes, cases of negative focal FDG BM activity with additional discordant clinical data, and cases of persistent focal FDG BM activity.

## FDG-PET assessment at baseline and on-study TPs

At the baseline imaging assessment, whenever possible all sites of lymphomatous disease are selected (as described in Figure 1) and should represent the patient’s overall FDG-avid tumor burden. The most effective implementation of the methodology for response assessment in clinical trials described in this manuscript, requires consistency of FDG-PET/CT image acquisition performed at multiple sites. Consensus guidelines for FDG-PET/CT image acquisition have been published by experts in the field, and should be incorporated into any clinical trial paradigm.<sup>18</sup> Correlation of the FDG-avid sites on CT imaging should be performed to confirm lesion size and morphology, and differentiate sites of disease from bowel and vascular structures. Finally, a 5PS score (as previously described) is assigned to the baseline TP to represent disease avidity on the FDG-PET imaging. The baseline assessment workflow is summarized in Figure 1. When assessing post-baseline imaging TPs, the same method used at baseline is employed.

Although the Lugano 2014 criteria represent a major advance in the assessment of FDG-avid lymphomas, there are a number of specific modifications to consider when optimizing the criteria for use in multi-center clinical trials. Our proposed approach is summarized in the following sections.

**Figure 1** Proposed baseline assessment workflow.

**Abbreviations:** CT, computed tomography; PET, positron emission tomography; 5PS, 5-point scale;  $SUV_{max}$ , maximum standardized uptake value.

## Proposed operational approach and recommendations for including Lugano 2014 criteria in FDG-avid lymphoma clinical trials

### Baseline assessment workflow

At baseline, a maximum of six sites with most metabolically active FDG-avid disease (classified as index [or target] lesions) should be selected which, when possible, include the largest lesions most representative of the patient's overall tumor burden. When possible, index lesions should be chosen from disparate regions of the body and include mediastinal and retroperitoneal areas of disease. These lesions must meet the minimum size requirement of being  $>15$  mm in longest diameter (LD<sub>i</sub>) for nodal disease, or  $>10$  mm in LD<sub>i</sub> for extranodal lesions. The LD<sub>i</sub> and shortest diameter should be recorded for each index lesion. All other disease, consisting of up to ten individual or grouped sites, should be selected at baseline as non-index (or non-target) disease. These can include nodal or extranodal lesions or groups of lesions which are not measurable (or measurable beyond the six sites chosen to be followed as index lesions). In addition, non-index disease can include markedly diffuse FDG uptake in the liver or spleen and marked focal FDG uptake in the BM. All selected sites of disease should be followed throughout the course of treatment. Although the PET criteria in Lugano 2014 for FDG-avid lymphomas do not specifically

require the designation of index and non-index disease and size measurements on CT, this approach allows investigators to follow all disease in a logical manner, where the PET findings can be easily correlated with CT imaging and clinical observations. This methodology ensures that all sites of disease, which are reflective of the patient's overall tumor burden, are accounted for in each TP assessment.

Our approach specifies that each designated index and non-index CT lesion should be correlated to the corresponding and co-registered PET lesion. A visual assessment using the 5PS should then be performed on the most metabolically active lesion out of all index and non-index disease. In addition, a quantitative standardized uptake value (SUV) measurement, which represents the maximum SUV ( $SUV_{max}$ ), should also be documented for this lesion. The  $SUV_{max}$  will be used to calculate the change in uptake compared to post-baseline TPs.

### On-study assessment workflow

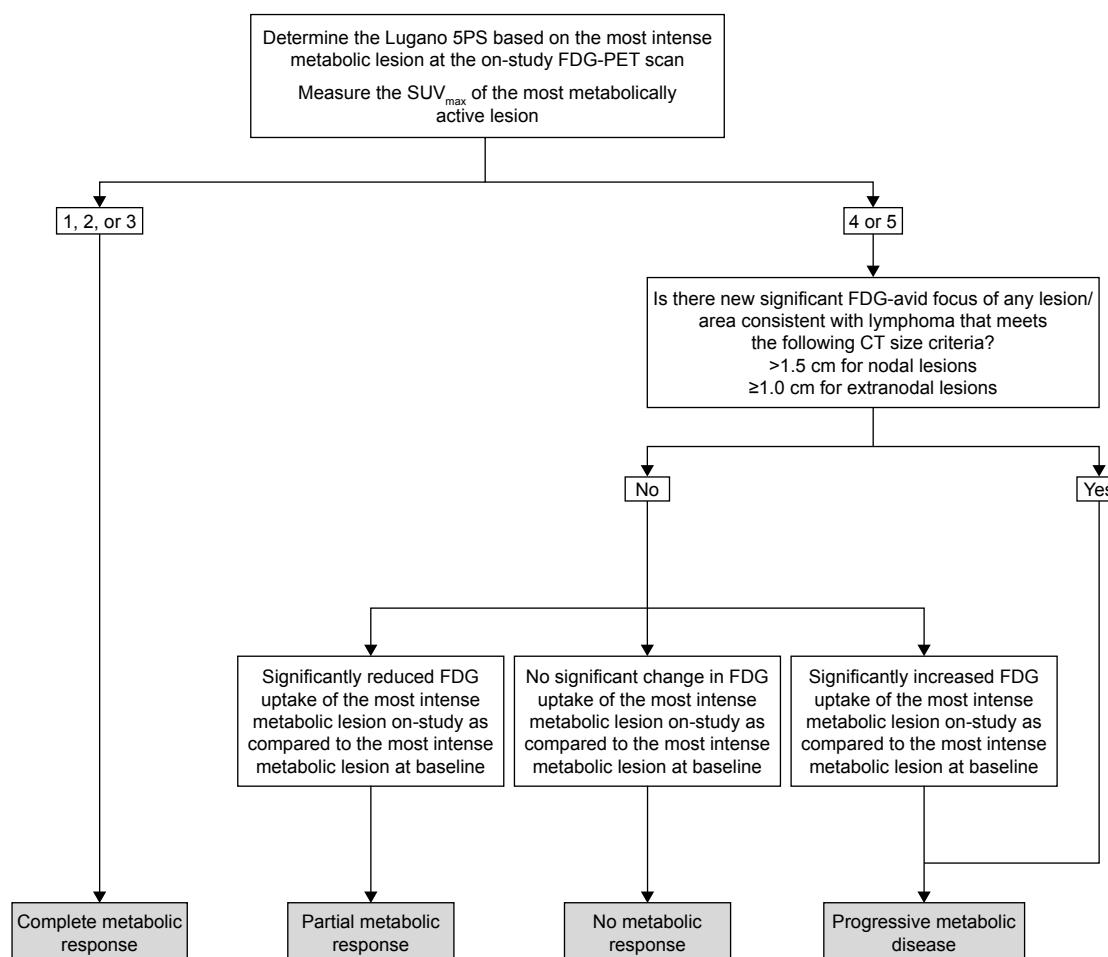
At each on-study TP, the index and non-index lesions identified at baseline are assessed on the PET/CT exam. The most metabolically active lesion is again assessed using the 5PS approach and the  $SUV_{max}$  of that lesion is determined. Of note, it is possible that the most metabolically active lesion identified on-study, when a subject is undergoing treatment, may be different from the most metabolically active lesion which had been identified at the baseline TP. The on-study  $SUV_{max}$  measurement is then utilized to perform the on-study

response assessment. The proposed on-study operational approach is illustrated in Figure 2.

The  $SUV_{max}$  of the most metabolically active lesion at each on-study TP is compared to the most metabolically active lesion at baseline in order to quantify changes in FDG uptake and obtain a percent change in  $SUV_{max}$ . An on-study Lugano 5PS score of 1, 2 or 3 is considered complete metabolic response at both interim and EOT, whereas a score of 4, 5 represents a different response outcome depending on the measured change in FDG uptake (Table 2) and the type of TP being evaluated (ie, an interim or EOT TP). At interim, a score of 4 or 5 with a decrease of  $>25\%$  in  $SUV_{max}$  is considered to be a significant decline in FDG uptake, representative of a partial metabolic response. A score of 4 or 5 with an increase of  $>50\%$  is considered a significant increase in FDG uptake, representative of progressive metabolic disease (PMD), and a change metric between  $\leq 25\%$  decrease and  $\leq 50\%$  increase in FDG uptake is considered to be no significant change in

FDG uptake, representative of no metabolic response. At EOT, a score of 4 or 5 is representative of treatment failure (TF) regardless of any significant change in  $SUV_{max}$ . These threshold metrics may be modified in the context of different clinical trial requirements.

In frontline therapy lymphoma trials, where a finite number of drug treatment cycles is frequently part of the study design, the recommendation is to use both interim and EOT assessment methods. Conversely, in relapsed and/or refractory lymphoma trials, it is recommended to only use interim assessment methods, as continuation of therapy can be based on a wide range of factors. In this paper, the focus is on the use of an interim assessment method; however, when accounting for an EOT analysis, the main difference is that the term TF is incorporated in the Lugano guidelines as a descriptor for patients who have demonstrated persistent FDG lesion uptake. This term can be confusing when used along with the term PMD. In our proposed workflow, both TF and PMD are



**Figure 2** On-study PET response workflow.

**Abbreviations:** PET, positron emission tomography; 5PS, 5-point scale; FDG, [ $^{18}\text{F}$ ] fluorodeoxyglucose;  $SUV_{max}$ , maximum standardized uptake value; CT, computed tomography.

**Table 2** Determination of Lugano PET-based on-study response

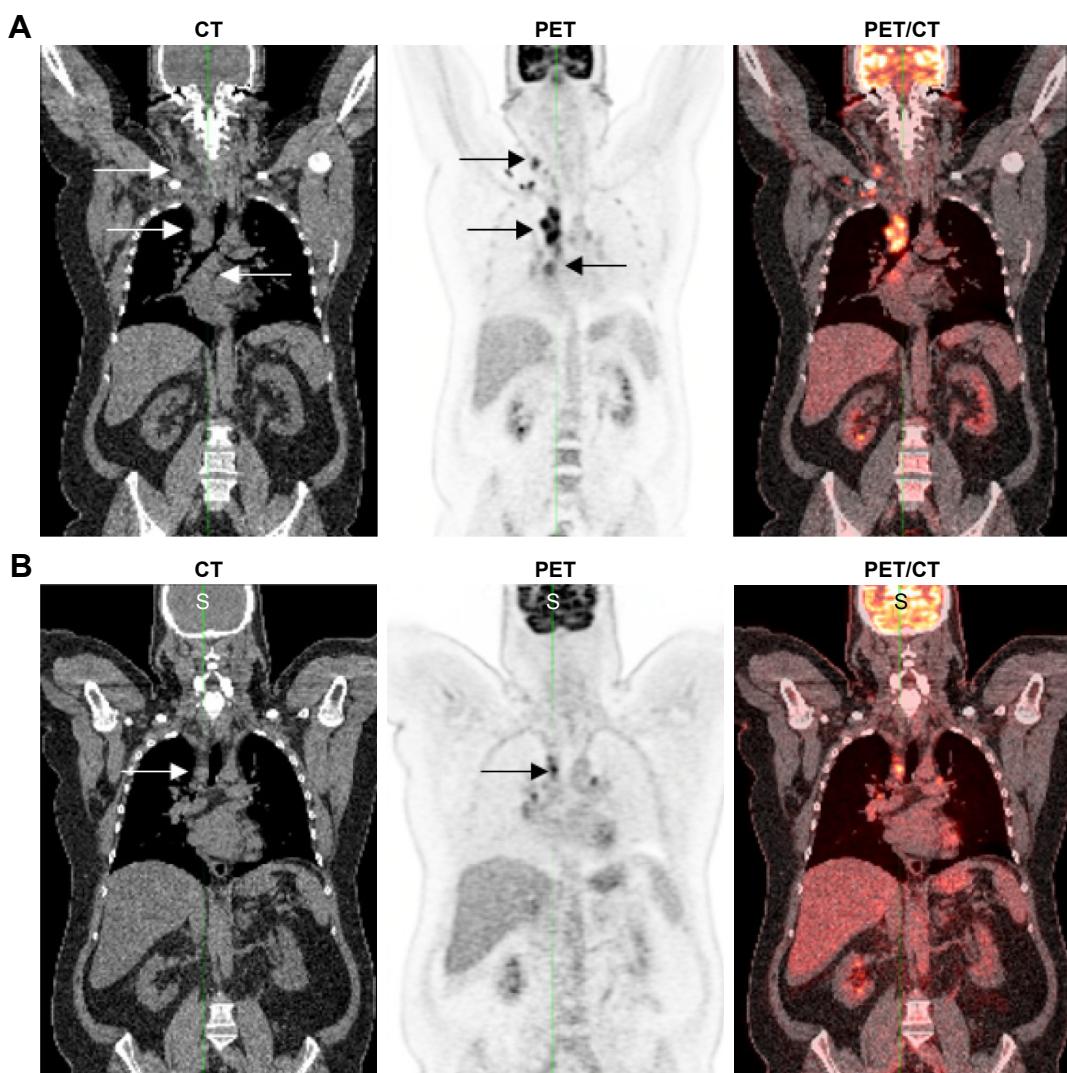
Lesion 5PS	Change from screening	New lesions	Lugano response
1, 2, or 3	Decrease	No	CMR
4 or 5	Decrease	No	PMR
4 or 5	No change	No	NMR
4 or 5	Increase	Yes/no	PMD
Any	Any	Yes	PMD

**Abbreviations:** PET, positron emission tomography; 5PS, 5-point scale; CMR, complete metabolic response; PMR, partial metabolic response; NMR, no metabolic response; PMD, progressive metabolic disease.

classified as one category labeled as PMD. It is important to note that, depending on the type of therapy being evaluated in a particular oncology trial, different change metric thresholds and response categories may be used.

## An example of the baseline and on-study workflows

For example, if five lesions are being followed and only one is observed to have markedly increased uptake compared to the liver, then the overall 5PS score for that TP is 5, regardless of the uptake in the other four lesions. Within this 5PS approach, a score of 1, 2, and 3 is generally considered to be PET-negative and values of 4 and 5 would be classified as PET-positive; however, in certain situations (specifically, in response-adaptive trials assessing de-escalation therapy) a 5PS score of 3 may be viewed as PET-positive and therefore considered to be an inadequate treatment response (Figure 3). Note: the classification of a 5PS of 3 as PET-positive should be prospectively defined

**Figure 3** Lugano operational workflow example.

**Notes:** (A) A 50-year-old male with NHL. CT (left image) reveals diffuse adenopathy in the neck and mediastinum (white arrows). PET (middle image) reveals marked FDG uptake (5PS of 5) in neck and mediastinum (black arrows) confirmed on fused PET/CT (right image). (B) Interim treatment follow-up CT (left image) at 8 weeks reveals significant residual adenopathy (white arrow). Follow-up PET (middle image) still assessed as marked uptake above liver (5PS of 5), demonstrates a significant decrease in FDG uptake with residual activity in the mediastinum (black arrow) which is also confirmed on the PET/CT (right image). The overall findings are consistent with a significant partial metabolic response.

**Abbreviations:** NHL, non-Hodgkin's lymphoma; CT, computed tomography; PET, positron emission tomography; FDG, [<sup>18</sup>F] fluorodeoxyglucose; 5PS, 5-point scale.

prior to the commencement of reads, preferably during development of the trial protocol.

## Validation of the Lugano operational workflow

### Introduction

In order to validate the proposed Lugano workflow, a pilot study cohort consisting of 12 NHL patients with a total of 34 imaging TPs was evaluated. The objective of this validation was to determine if the proposed Lugano workflow is a feasible method to improve the reproducibility and efficiency amongst readers evaluating the radiographic response in lymphoma patients.

### Methods

The study cohort consisted of 12 well-documented, NHL patients, eight male, four female, ranging in age from 40 to 80 years, who were a subset of patients enrolled in an Institutional Review Board (IRB)-approved early-phase, commercially-sponsored, clinical trial.<sup>19</sup> All patients signed written informed consents acknowledging all aspects of the clinical trial including an independent review of their imaging data. The participating IRBs included Schulman Associates IRB, St John Hospital and Medical Center IRB, and New England IRB. All of the patients had documented FDG-avid disease. Three of the authors, all experienced independent radiology reviewers (RA, LS, FT), blindly assessed all FDG-PET/CT scans for each patient. The reviewers conducted their evaluations using a modified 5PS (Table 1) according to the modified Lugano<sup>10</sup> response assessment criteria for FDG-avid lymphomas and the proposed operational workflow described in this paper. Specifically, the reviewers were instructed to categorize all lymphomatous disease visualized on FDG-PET/CT into one of two groups. The six most dominant lesions (labeled as index 001–006) which were representative of the patient's overall tumor burden were identified on the baseline CT and PET images. These index lesions were measured for anatomic size on CT and evaluated for metabolic activity on the PET images. The remainder of the patient's tumor burden was assigned by location to a maximum of ten single (or grouped) lesion sites (labeled as non-index 200–209). Baseline lesions were considered positive when they met the criteria for a 5PS of 4 or 5 (ie, uptake moderately or markedly > liver).

At on-study TPs, CT index and non-index lesions were assessed for continued presence of disease, however, no size measurements or change metrics were required. PET index and non-index lesions were again assessed using the on-study

workflow (Figure 2). The change metric at each on-study TP was calculated by the following formula:

$$\% \Delta \text{SUV}_{\max} = \left( \frac{\text{SUV}_{\max_{TP_x}} - \text{SUV}_{\max_{\text{Screening}}}}{\text{SUV}_{\max_{\text{Screening}}}} \right) \times 100$$

Note:  $TP_x$  = Time point X

New lesions (labeled as 300–302) were required to meet the minimum anatomic size criteria (>15 mm) on CT and were also required to be FDG-avid on PET (ie, 5PS of 4 or 5).

### Statistical methods

The overall concordance rate between the three reviewers was determined. A 5PS of either 1, 2 or 3 (PET-negative) was considered concordant across readers, as was a score of 4 or 5 (PET-positive). Reader discordance was observed when the 5PS assessment recorded by readers differed between PET-positive and PET-negative at a given TP.

In order to evaluate inter-reader reliability, Fleiss' kappa, an extension of Scott's pi to more than two observers and nominal categories, was calculated for the Lugano TP 5PS assessments of three readers who independently reviewed multiple FDG-PET imaging series.<sup>20</sup>

### Results

All three reviewers assessed all patients at all individual TPs using the Lugano TP 5PS (Table 1). The three independent reviewers were in concordance in 97 out of 102 TPs assessed, which equates to an overall 95% concordance rate. The kappa statistic for 5PS agreement (Table 3) was 0.73 (confidence interval 0.59–0.87;  $P=<0.0001$ ) using the Fleiss kappa statistic methodology, indicative of an overall good to excellent correlation between the three readers. Furthermore, two out of the three reviewers were in concordance at all TPs.

### Discussion

The incorporation of the Lugano 2014 criteria for the assessment of lymphoma patients' response to therapy represents an important paradigm shift. In particular, the use of FDG-PET/CT imaging as the dominant imaging technique for the evaluation of FDG-avid lymphomas, allows for the pathophysiologic assessment of tumor metabolic activity in the ongoing evaluation of lymphoma patients.

There are a number of issues that must be considered when implementing a workflow paradigm for the Lugano criteria. These include the proper selection of representative disease, lesion size thresholds for index and new lesions, and

**Table 3** Reviewer results – Lugano time point 5PS assessment

Patient		Lugano scores			
ID	TP	2	3	4	5
1	B	0	0	0	3
	I	0	0	0	3
	2	0	0	0	3
2	B	0	0	0	3
	I	0	0	0	3
3	B	0	0	0	3
	I	0	0	3	0
	2	I	I	I	0
4	B	0	I	2	0
	I	0	0	2	I
	2	0	3	0	0
5	B	0	3	0	0
	2	0	0	3	0
6	B	0	0	3	0
	I	0	0	3	0
	2	0	0	3	0
7	B	0	0	0	3
	I	0	0	0	3
	2	0	0	0	3
8	B	0	0	0	3
	I	0	0	3	0
	2	0	3	0	0
9	B	0	0	0	3
	I	0	0	0	3
	2	0	2	I	0
10	B	0	0	2	I
	I	0	2	I	0
	2	0	2	I	0
11	B	0	0	0	3
	I	0	0	0	3
	2	0	0	0	3
12	B	0	0	I	2
	I	0	0	3	0
	2	0	0	3	0

**Abbreviations:** 5PS, 5-point scale; TP, time point; B, baseline.

ensuring that proper reader training is provided prior to trial commencement.

### Lesion selection

The optimum approach is to have each reader select all FDG-avid lesions present at the baseline TP. In situations where a reader chooses to select and follow fewer lesions from the total number of FDG-avid lesions, there may be a higher risk of discordance between the assessment by a reader who chooses to select and follow a larger representative number of lesions. This is particularly true at on-study TPs where there may be a difference in the 5PS assessment between individual readers. In this situation, the difference between reader assessments may be spurious due to the lack of complete lesion selection by one of the readers. For example, if one reader selects one index lesion and no non-index

lesions while another reader selects four index lesions and three non-index lesions, there is a high probability that there may be a discrepancy in the PET 5PS response assessment between the two readers.

### Lesion size threshold

It is also essential to establish a minimum size for the selection of index lesions at baseline and the identification of new lesions at post-baseline TPs in order to ensure optimum reproducibility between individual readers. At baseline, a minimum size threshold for selecting index lesions will facilitate a higher level of reader harmonization with respect to selection of representative disease. At on-study TPs, a minimum size threshold for new lesions is also necessary for optimal reader concordance. For example, if one reader selects a new lesion which does not meet a minimum size criteria, the response assessment may result in PMD designation, whereas another reader who did not believe the same lesion to be significant, would end up with a different response assessment.

### Reader training

Comprehensive reader training for an individual lymphoma clinical trial using the Lugano 2014 criteria, should be based on the presumption that the readers already have in-depth experience and knowledge of the radiographic assessment of lymphomatous disease and the Lugano 2014 criteria. The specific training material for an individual clinical trial should include ample clinical case examples and test case examples, to ensure that each reader adequately understands the overall workflow guidelines and any individual study-specific rules. Ideally, a discussion of lessons learned and known pitfalls should also be included in the training session(s). The overall goal of the training process is to achieve optimal reader harmonization and concordance which is reproducible on both an inter- and intra-reader level. For example, in this pilot validation study, the 5PS discordance noted between readers occurred most frequently between a score of 3 ( $>$  mediastinum,  $\leq$  liver) and 4 (moderately  $>$  liver). This observation emphasizes the importance of training readers to be aware of the subtle differences in discerning FDG uptake in lesions compared to the background activity in the liver.

The approach taken in this validation study was to categorize all FDG-avid lesions as either index or non-index. This approach allows for the correlation of the FDG-PET assessments with the separate lesion size measurements on CT imaging, which is particularly helpful in clinical

trials which require independent CT and PET assessments. In studies that do not require a separate CT analysis with a classification of index and non-index disease, it is possible to assess PET scans in a different manner, which is beyond the scope of this paper. The workflow presented here is an operational scheme that can be utilized for single-center or multi-center clinical trials which include both CT and FDG-PET imaging requirements. The proposed approach has inherent flexibility, which allows further modification to address the specific goals and/or issues of an individual clinical trial or pharmacologic therapy.

Regardless of the approach taken, monitoring total tumor burden response at the cellular level is essential, and the process implemented to apply the Lugano 2014 criteria should result in an assessment method that more accurately aligns with a patient's clinical outcome. This is in contrast to the Cheson 2007 guidelines, where a limited number of representative lesions are selected and assessed using a simple dichotomous (ie, positive vs negative) scale.

Another potential challenge with the implementation of the Lugano 2014 criteria is in immunotherapy oncology clinical trials, where the observation of increased metabolic activity may be mistakenly interpreted as PMD. In these trials, the proper assessment of possible transient metabolic flare, which may be visualized on FDG-PET/CT scans at interim TPs during treatment, needs to be considered. One possible solution to this challenge is to raise the threshold for assessing PMD to allow for greater changes in immune-related FDG metabolic activity. An alternative approach for controlling against sudden increases in FDG uptake which may be due to immune-related metabolic activity, is to delay the confirmation of PMD until a subsequent imaging TP (approximately 6–12 weeks after the initial observation of PMD) is submitted and the initial assessment of PMD can be either confirmed or not confirmed by the reader. Additional recommendations for handling immune-related response assessments are discussed in a recent publication.<sup>21</sup>

## Conclusion

The workflow paradigm presented in this paper represents an operationalized method which can be utilized in single- and multi-center lymphoma clinical trials employing the Lugano 2014 PET criteria. The pilot validation data presented in this paper, confirm that the proposed workflow is a useful and reproducible methodology to achieve consistent imaging assessment results with a high level of concordance across readers. The proposed paradigm is a work-in-progress which will require validation in a larger multi-center clinical trial,

to further solidify the operational workflow as an assessment standard for clinical trials investigating therapies for PET-avid lymphomas.

## Disclosure

RLVH, RS, JGW, JAS and MON are full time employees at Bioclinica Inc; RA, FT and LS are reader consultants for Bioclinica Inc; BK is a full time employee of Sierra Oncology and RM is a full time employee at Apexian Pharmaceuticals. The authors report no other conflicts of interest in this work.

## References

1. seer.cancer.gov [homepage on the Internet]. National Institute of Health. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Available from: <https://seer.cancer.gov/statfacts/html/nhl.html>. Accessed May 9, 2017.
2. seer.cancer.gov [homepage on the Internet]. National Institute of Health. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Available from: <https://seer.cancer.gov/statfacts/html/hodg.html>. Accessed May 9, 2017.
3. Cang S, Mukhi N, Wang K, Liu D. Novel CD20 monoclonal antibodies for lymphoma therapy. *Journal of Hematology & Oncology*. 2012;5:64.
4. rsna.org [homepage on the Internet]. Radiological Society of North America. FDG-PET/CT Technical Committee. QIBA Profile FDG-PET/CT as an Imaging Biomarker Measuring Response to Cancer Therapy, Quantitative Imaging Biomarkers Alliance, Version 1.13, Technically Confirmed Version. QIBA, November 18, 2016. Available from: RSNA. ORG/QIBA. Accessed May 9, 2017.
5. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG-PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42(2):328–354.
6. Barrington SF, Mikhalev 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(27):3048–3058.
7. 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.
8. Svoboda J, Andreadis C, Elstrom R, et al. Prognostic value of FDG-PET scan imaging in lymphoma patients undergoing autologous stem cell transplantation. *Bone Marrow Transplant*. 2006;38(3):211–216.
9. Barrington SF, Mikhalev NG. PET scans for staging and re-staging in diffuse large B-cell and follicular lymphoma. *Curr Hematol Malig Rep*. 2016;11(3):185–195.
10. Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for Initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol*. 2014;32(27):3059–3068.
11. Cheson BD, Horning SJ, Coiffier B, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. *J Clin Oncol*. 1999;17(4):1244.
12. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol*. 2007;25(5):579–586.
13. Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the first international workshop on interim-PET-scan in lymphoma. *Leuk Lymphoma*. 2009;50(8):1257–1260.
14. Moskowitz CH. Interim PET-CT in the management of diffuse large B-cell lymphoma. *Hematology Am Soc Hematol Educ Program*. 2012; 2012:397–401.

15. Couglan M, Elstrom R. The use of FDG-PET in diffuse large B cell lymphoma (DLBCL): predicting outcome following first line therapy. *Cancer Imaging*. 2014;14:34.
16. Ha, et al. Follow-up of Hodgkin Lymphoma. ACR Appropriateness Criteria. *American College of Radiology*. 1999. Available from: <https://acsearch.acr.org/docs/69388/Narrative/>. Accessed May 9, 2017.
17. Standard operating procedures for PET/CT: a practical approach for use in adult oncology. Available from: [http://www.pub.iaea.org/MTCD/Publications/PDF/Pub1616\\_web.pdf](http://www.pub.iaea.org/MTCD/Publications/PDF/Pub1616_web.pdf). Accessed May 9, 2017.
18. Graham MM, Wahl RL, Hoffman JM, et al. Summary of the UPICT protocol for FDG PET/CT imaging in oncology clinical trials. *J Nucl Med*. 2015;56(6):955–961.
19. Klencke B. Sierra Oncology 2016. Available from: <https://clinicaltrials.gov/ct2/show/NCT01733238>. Accessed May 9, 2017.
20. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas*. 1973;33(3):613–619.
21. Cheson BD, Ansell S, Schwartz L, et al. Refinement of the Lugano Classification lymphoma response criteria in the era of immunomodulatory therapy. *Blood*. 2016;128(21):2489–2496.

**Drug Design, Development and Therapy****Publish your work in this journal**

Drug Design, Development and Therapy is an international, peer-reviewed open-access journal that spans the spectrum of drug design and development through to clinical applications. Clinical outcomes, patient safety, and programs for the development and effective, safe, and sustained use of medicines are the features of the journal, which

Submit your manuscript here: <http://www.dovepress.com/drug-design-development-and-therapy-journal>

**Dovepress**

has also been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.