Journal of Nuclear Medicine, published on July 14, 2022 as doi:10.2967/jnumed.122.264124

APPLICATION OF THE LUGANO CLASSIFICATION FOR INITIAL EVALUATION, STAGING, AND RESPONSE ASSESSMENT OF HODGKIN AND NON-HODGKIN LYMPHOMA: THE PROLOG CONSENSUS INITIATIVE (PART 2- TECHNICAL).

Short running title

Lugano classification: PRoLoG technical

Authors list

Fabien Ricard* (1), Sally Barrington* (2), Ron Korn (3), Guenther Brueggenwerth (4), Judith Trotman (5), Bruce Cheson (6), Gilles Salles (7), Larry Schwartz (8), Greg Goldmacher (9), Rudresh Jarecha (10), Jayant Narang (11), Florence Broussais (12), Paul Galette (13), Min Liu (14), Surabhi Bajpai (15), Eric Perlman (16), Julie Gillis (17), Ira Smalberg (18), Pierre Terve (19), Gudrun Zahlmann (20), Annette Schmid (11)

*contributed equally to this work

Institution/Affiliation

(1) Relay Therapeutics, Cambridge, Massachusetts
(2) King's College London and Guy's and St Thomas' PET Centre, School of Biomedical Engineering and Imaging Sciences, King's College London, King's Health Partners, London, UK
(3) Adjunct Faculty TGEN/City of Hope and Imaging Endpoints Core Lab, Scottsdale, AZ, USA
(4) Bayer AG, Berlin, Germany
(5) Concord Repatriation General Hospital, University of Sydney, Concord, Australia
(6) Lymphoma Research Foundation, New York, NY, USA
(7) Department of Medicine, Memorial Sloan Kettering Cancer Center and Weil Cornell Medicine, New York, NY, US
(8) Department of Radiology, Columbia University College of Physicians and Surgeons and Radiologist-in-Chief, The New York Presbyterian Hospital, Columbia, NY, USA

(9) Merck & Co, Inc. Kenilworth, NJ, USA

(10) Deciphera Pharmaceuticals LLC, Waltham, MA, USA

(11) Takeda Pharmaceutical Company Ltd, Cambridge, MA, USA

(12) Lymphoma Study Association Research Center LYSARC, Pierre Benite, France Calyx international, Billerica, MA, USA

(13) GSK, Collegeville, PA, USA

(14) Autolus Therapeutics, London, UK

(15) Calyx international, Billerica, MA, USA

(16) Perlman Advisory Group LLC, Boynton Beach, FL, USA

(17) Merigold LLC, Branford, CT, USA

(18) Associate Professor of Radiology, Saint John's Cancer Institute and Tower imaging medical group, Sherman Oaks, CA, USA

(19) KEOSYS Medical Imaging, Saint Herblain, France

(20) QIBA/RSNA, Radiological Society of North America, Oak Brook, IL, USA

*contributed equally to this work

First and Corresponding Author

Fabien Ricard MD MS

399 Binney St,

Cambridge, MA 02139

fabienricardmd@gmail.com

1 (917) 862-2616

Financial support

no

Word count manuscript (main text)

4989 words

ABSTRACT

Rationale: To provide consensus recommendations from a consortium of academic and industry experts in the field of lymphoma and imaging for the consistent application of imaging assessment with the Lugano classification.

Methods: Consensus was obtained through a series of meetings from July 2019 until October 2021 sponsored by the PINTaD (Pharma Imaging Network for Therapeutics and Diagnostics) as part of the ProLoG (PINTaD RespOnse criteria in Lymphoma wOrking Group) consensus initiative.

Results: Consensus recommendations encompass all technical imaging aspects of the Lugano classification. Some technical considerations for PET-CT and diagnostic CT are clarified with regards to required imaging series and scan visits, as well as acquisition and reconstruction of PET images and influence of lesion size and background activity. Recommendations are given on the role of imaging and clinical reviewers as well as on training and monitoring. Finally, an example template of imaging case report form is provided to support efficient collection of data with Lugano Classification.

Conclusion: Consensus recommendations are made to comprehensively address technical and imaging areas of inconsistency and ambiguity in the classification encountered by end users. Such guidance should be used to support standardized acquisition and evaluation with the Lugano 2014.

Keywords: Lugano classification, technical recommendations, consensus, standardization

INTRODUCTION

In 2014, the Lugano classification (*1*) together with an imaging-focused companion report (*2*) (referred together as Lugano 2014) provided a standardized approach to classifying response based upon Fluorodeoxyglucose (FDG) positron emission tomography/computed tomography (PET-CT) in FDG-avid lymphomas. The Lugano 2014 was an update to the Revised Response Criteria for Malignant Lymphoma published in 2007 (*3*) (referred as Cheson 2007).

The Lugano 2014 has since been used by regulatory agencies for recent drug approval and widely adopted both by the pharmaceutical industry and clinicians for evaluation of Hodgkin (HL) and Non-Hodgkin Lymphoma (NHL). Currently, hundreds of actively recruiting and ongoing investigational trials are using the Lugano classification (*4*).

The PRoLoG committee (PINTaD RespOnse criteria in Lymphoma wOrking Group), sponsored by the PINTaD (*5*) (Pharma Imaging Network for Therapeutics and Diagnostics), is a cross-functional group of volunteers from the industry and academy who engaged in discussions to provide expert end-users consensus recommendations for the consistent application of the Lugano classification.

This manuscript, focusing on the technical imaging recommendations, is not intended to replace the classification. It may also be applied to some extent to the newer lymphoma response assessment criteria (e.g. Lymphoma Response to Immunomodulatory Therapy Criteria 2016 (*6*) and Response Evaluation Criteria in Lymphoma 2017 (*7*)). While these recommendations are primarily given for clinical trial end-users, it may be valuable information for healthcare providers as well.

METHODS

Task forces (TF) were created to evaluate technical imaging and clinical considerations of the Lugano classification that could affect its uniformity in evaluating lymphoma response.

The TF members included representatives from academic/scientific organizations (n=3), pharmaceutical industry (n=9), clinical research organizations (n=13) and other clinical trial specialists (n=4), as well as independent research leaders (BC, SFB, JT, GS, LS). A steering committee (FR, RK, AS, GB) oversaw the activities of each TF. All meetings were held virtually, from July 2019 to October 2021, recorded and transcribed into minutes that were approved by the TF members. In instances where there was lack of evidence-based data, or consensus, a call for future research on that topic was suggested. Additional recommendations from the TF, primarily for clinical imaging considerations will be available elsewhere ("Application of the Lugano Classification for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Prolog Consensus Initiative (Part 1- clinical)").

Any individual involved in the implementation of the Lugano classification is considered an "end-user". Any physician responsible for assessing response in lymphoma is considered a "reviewer".

TECHNICAL CONSIDERATIONS FOR IMAGE ACQUISITION, RECONSTRUCTION AND EVALUATION

Required Images and Viewing Stations

The assessment of the Lugano classification is informed by both anatomic imaging (diagnostic CT preferred; however, can be interchangeable with magnetic resonance imaging [MRI]; ultrasound should not be used due to the operator-dependency of the method), and metabolic (18F-FDG PET-CT) imaging, for FDG-avid lymphomas.

The following images should be provided to the reviewers:

- 18F-FDG PET-CT
    - PET attenuation-corrected images (AC)
    - PET non attenuation-corrected images (NAC)
    - Low Dose CT for attenuation correction (CTAC) and for localization purposes
    - Reconstructed images: AC MIP (maximum intensity projection) and PET-CT fusion images, unless the viewing software enables creation from AC images. Care should be taken that no patient identifiers are embedded on reconstructed images.
- Diagnostic CT
    - CT images with anatomic coverage to include all areas of known or suspected disease with appropriate acquisition settings for kVp, mAs, slice thickness of ≤ 5mm, intravenous (IV) contrast, and patient positioning and breathing instructions (e.g. deep inspiration breath-hold)
    - Standard soft tissue and lung reconstruction images

Viewing Stations for Image review and interpretation should provide adequate functionality to allow multiplanar display (i.e. axial, coronal and sagittal views) of PET, diagnostic CT and fused PET-CT images for image interpretation and lesion cross-referencing purposes. PET images should be scaled to a set SUV range and color table.

- PET software should allow creation of MIP images (of special importance for providing visual scoring assessments of distant lesions to mediastinum and liver reference tissues).
- Reading software should allow for vendor neutral evaluation of PET images including semi-quantitative uptake measurements, of CT images including size measurements and may ideally allow for volumetric assessments (which are interesting exploratory measurements, but not included in the Lugano classification).
- The Quantitative Imaging Biomarkers Alliance (QIBA (*8*)) has provided guidance on system's technical performance standards (*9,10*) when the aim is to use 18F-FDG PET as a quantitative imaging biomarker.

18F-FDG PET-CT and diagnostic CT Scan Visits

PET-CT should provide sufficient anatomical coverage to accurately assess whole body tumor burden. It should include common areas of disease involvement including the neck, chest, abdomen and pelvis (including groin) as a minimum for all patients. Coverage should be adjusted to include additional areas of known or suspected disease (e.g. extremities). Inclusion of the brain is dependent on the lymphoma disease status and imaging center standard protocol. It is highly recommended that 18F-FDG PET emission scanning commences in the pelvis/thigh region, and extend to the upper body, to avoid reconstruction artifacts due to high bladder uptake. The same

PET-CT scanner and scanning direction should be used on follow-up timepoints and consistent patient positioning and breathing instructions should be ensured across all imaging visits. Time from injection of 18F-FDG to acquisition of PET images should be kept rigorously constant across successive scans in a patient to allow for comparability of metabolic images (ideally +/- 5 minutes, up to +/- 10 minutes, compared to time used at baseline) and acquisition should always be timed to close to 60 minutes post injection (55-75 min is acceptable) (*9*-*12*). Factors affecting SUV calculation (e.g. injection time, but also administered activity, weight) that are entered manually onto the scanner, should be carefully checked, and documented for quality control purposes.

Whenever possible, 18F-FDG PET-CT and diagnostic CT scans, if both are required at the same timepoint, should be acquired on the same scanner during the scheduled imaging visit for patient convenience. CTAC scans should be obtained without IV or positive oral bowel contrast. The use of intravenous contrast for diagnostic CT should be performed after the PET-CTAC acquisition in order to avoid over-attenuation of the PET images from the CT contrast medium.

A CT should be considered of diagnostic quality (so-called diagnostic CT) if it has adequate resolution to detect and accurately measure lesions and spleen size and should contain intravenous contrast, unless contra-indicated, ideally in the portal venous phase for clinical trials. Oral contrast is recommended per site standard of care, especially in patients with known or suspected hollow viscus involvement or mesenteric lymphadenopathy. Technical acquisition parameters, use of intravenous contrast unless medically contraindicated, breath-hold techniques and arm positioning should be prespecified in study documents and kept as consistent as possible for a given subject across time points, and as much as possible for the trial. The CT portion of a PET-CT can be used for lesions and spleen measurements if it is considered of acceptable diagnostic quality.

For situations in which a patient is diagnosed at a center different from the treating institution, it is of utmost importance that the baseline scan (images and image acquisition fields) be made available in DICOM format to enable comparison to subsequent imaging. Ideally, all scans for a same patient should be conducted at the same scanner and same institution throughout the trial.

Further recommendations are provided in Supplemental Table 1.

PET Acquisition and Image Reconstruction

Phantom-based quantitative calibration validation is strongly recommended prior to starting a clinical trial, and is even critical in trials where main endpoints require SUV/activity-concentration-based quantitative measurements; although for trials with no quantitative measurements the imaging facilities manufacturer and institutional recommended regular quality control that is used for clinical care, may be sufficient.

Semi-quantitative SUV read-outs can be of interest in trials using the Lugano classification (*2*) and it is highly recommended that the comprehensive QIBA FDG profile (*9,10*) be implemented at each site as a guideline for standardization of the 18F-FDG PET workflow. Other guidance exists such as the European Association of Nuclear Medicine procedure guidelines for tumor imaging with FDG PET/CT (*12*). Key PET reconstruction parameters should be agreed between scanning sites and the study sponsor to harmonize image quality and quantification.

Change in SUVmax (ΔSUV) and Metabolic Tumor Volume (MTV) may be promising tools for response evaluation and prognosis in lymphoma (*13,14*) including for clinical trials, further emphasizing the need for standardization of PET acquisition (*15*). A change in SUV measurement (e.g. ΔSUVmax of less than or equal to 66% in 18F-FDG PET-CT after 2 cycles of

chemotherapy for diffuse large B-cell lymphoma as a correlate to an unfavorable outcome (*16-19*)) has been suggested for response and prognosis evaluation at interim PET, as well as for assessment in PET-guided therapy (*20*) and this promising measurement is ongoing further validation (*21-23*).

Acquisition and reconstruction methods should be kept consistent throughout the trial and between patient visits. PET 3D mode acquisition with Time of Flight (ToF) is preferred when available. In the interest of harmonization of the image acquisition across sites, newer reconstruction methods that may not be widely available (i.e. Point spread Function [PSF] corrections, regularized reconstructions, AI-based acquisition and reconstruction algorithms, etc.) and whose effect on the 5-PS is not yet known, should be used cautiously to assess study outcomes for PET-guided therapy decisions until its impact on the 5-PS is better understood.

However, the TF acknowledges that phantom harmonization programs that align scanner performances across institutions may help to mitigate such differences between such newer reconstruction methods, especially for semi-quantitative assessments (e.g., SUV, MTV, etc). Although prospective harmonization of PET scanners in a multi-institutional clinical trial setting is desirable, it may not always be entirely practical or feasible due to variety of reasons (including the use of different reconstruction algorithms, such as e.g., Bayesian penalized likelihood [BPL] and PSF) compared to older methods (e.g., traditional ordered subset expectation maximization [OSEM]).

Technical Influence of Lesion Size and Background Activity

The influence of lesion size and activity concentration on partial volume is difficult to correct for in smaller lesions. This is particularly relevant when using the 5-PS to assess small residual lesions in lymphoma response assessment. In phantom studies using different sized spheres filled with identical concentrations of ^{18}Fluorine to mimic tumor sizes, smaller lesions (< 2 cm) appeared to have less 18F-FDG activity than larger lesions ($\geq$2 cm) (*24-26*). This is due to the inability of PET scanners to fully recover all the counts (i.e. partial volume effects) from smaller compared to larger spheres (or lesions) (*24*).

Although newer scanners may have advanced reconstruction algorithms to account for the loss of signal (point-response function and/or regularized-reconstructions), there have been no well-controlled studies addressing this issue or its influence on the application of the 5-PS.

Therefore, a uniform recommendation by the TF on how to integrate lesion size information into Lugano evaluation is not possible at this time, and further investigation is encouraged.

Signal to noise ratio (SNR) plays an important role in lesion detection. Image reconstruction and post processing of images with available reconstruction algorithms and filtering helps to control for and remove noise which should be optimized for individual scanners based upon either phantom testing and/or according to the suggested recommendations of manufacturers specifications. However, the conspicuity of lesions is not only dependent upon lesion signal but also on the uptake or signal in surrounding tissue and organs. Therefore, the reader should be aware of this phenomenon when performing scan interpretation.

SUV measurements

Some semi-quantitative measurements are routinely recorded (e.g. most hypermetabolic lesion, reference regions) and such measurement may be used to confirm visual assessment, e.g. to assign a score of 5 on the 5-PS (*2*,*13*).

Standardized uptake values (SUV) that are captured (e.g. most hypermetabolic lesion, reference regions) usually represent the maximum values (SUVmax), in alignment with the Lugano classification. However, other types of measurements (e.g. lesion SUVpeak, reference region SUVmean) are frequently recorded in clinical trials (*13*).

SUVmax represents the uptake in the single voxel exhibiting the highest tracer uptake in the region of interest. It is easily available on read stations, has good interreader reproducibility and is relatively unaffected by partial volume effects. However, SUVmax is influenced by noise.

SUVpeak is the average of the SUV in the $1cm^3$ of voxels with the highest activity in a volume of interest. SUVpeak (corrected for lean body mass) is used in PERCIST (*27*) that was proposed in 2009 to better standardize PET response criteria in solid tumors, in order to combine good interreader reproducibility, reduce the influence of partial volume with SUVmax and improve count rate stability.

SUVmean represents the mean tracer uptake in the region of interest. Usually, the most metabolically active portion within the area of interest should be used within the region of interest in which SUVmean is calculated. Measurement of the mean is dependent on the size of the region or volume of interest which should be standardized.

Further work is warranted in this field to identify the optimal measure for lymphomas. Besides, metabolic assessments (e.g. metabolic tumor volumes) and other radiomic features may become more important in the future.

Terminology for image evaluation and reporting

Lugano 2014 considers both metabolic and anatomic assessments when evaluating FDG avid lymphoma. With regards to response assessed on diagnostic CT, both “radiographic” and “anatomic” terminology have been used. The TF recommends naming it “anatomic”.

When evaluating response to therapy, it is recommended to assess and record the metabolic response, the anatomic response, the imaging response (metabolic response, anatomic response or combination of both when both available) and the overall response (used to determine endpoints, integrating clinical data when available). In order to differentiate anatomic and overall response, which nowadays are using the same terminology, it has been discussed with the clinical data interchange standards consortium (CDISC) to incorporate “anatomic” when recording the anatomic response. Thus, the anatomic response is now referred to complete anatomic response (CAR), partial anatomic response (PAR), stable anatomic disease (SAD) and progressive anatomic disease (PAD). Metabolic response remains defined as complete/partial metabolic response (CMR, PMR), no metabolic response (NMR, preferred term, since “stable disease” usually refers to radiographic stability) or stable metabolic disease (SMD), and progressive metabolic disease (PMD). The overall response remains defined as complete response, partial response, stable disease and progressive disease. Thus, it is clear what each component of the response is, and how it complementarily results in the overall response.

ROLE OF THE REVIEWERS: EXPERIENCE AND QUALIFICATIONS, TRAINING AND MONITORING

Imaging Reviewers Qualifications and Experience

Dependent upon the read requirements of a clinical protocol, the selection of imaging reviewer should meet certain qualifications, including the documentation of competency in diagnostic CT and/or PET-CT.

Reviewers should be Board Eligible (BE) or Board Certified (BC) Nuclear Medicine Physician (or the regional/national equivalent) with experience and/or certification in CT/MRI, or BE/BC Radiology physicians with experience and/or training in PET-CT imaging.

In addition, all reviewers should provide documented evidence of prior clinical experience with lymphomas and clinical trial participation in lymphoma studies on their CV or through attestations of participation. In cases where a reviewer may have no prior experience in clinical trial reads, a program of appropriate training about the application of the Lugano classification in the context of clinical trials and including test imaging cases is required.

Clinical Reviewers Qualifications and Experience:

Although Lugano does not specifically recommend separate imaging and clinical reviews, if a hematology-oncology review is requested, then the selection of clinical reviewers should meet prespecified qualifications including the credentials as a BE or BC physician in Hematology and/or Oncology (or the regional/national equivalent).

Having additional experience in clinical care of hematologic malignancies- either through clinical practice or in clinical trials, is required.

In addition, all reviewers should provide documented evidence of prior clinical trial participation in lymphoma studies on their CV or through attestations of participation. In cases where a reviewer may have no prior experience in clinical trial reads, a program of appropriate training about the application of the Lugano classification in the context of clinical trials and including test cases is required.

Close monitoring of on-trial performance is recommended for all reviewers regardless of training or experience.

Role of the Imaging Reviewer

The role of a blinded independent central reviewer (BICR) is to provide independent review of cases without bias or unblinding to treatment. It is recommended, when possible, that the reviewer remains the same throughout the reads of all timepoints for a patient. Where feasible, it is ideal to have the same reviewer provide assessment of both the 18F-FDG PET-CT and diagnostic CT throughout the entire study for an individual patient basis. If separate reads of diagnostic CT and 18F-FDG PET-CT occur, it is recommended that both readers meet for an "integration read" of anatomic and metabolic assessments that should be conducted to provide one patient level imaging timepoint assessment.

Whenever there are two BICRs evaluating scans from the same patient and modality, a third independent reviewer (aka adjudicator) should be assigned to review the scans in cases of time point assessment discrepancies to resolve any disagreements that would impact the overall time point responses.

During an adjudication event, the adjudicator should select which reader he/she most closely agrees with, rather than providing a third independent assessment, and a rationale for the selection should be provided. Alternative adjudication workflows exist, which are beyond the scope of this manuscript.

Reviewers Training and Monitoring

Training on Lugano classification (and any protocol specified modifications or clarifications), imaging case report form completion, familiarization of workstation usage and group review of clinical cases for formulating consensus on scan interpretation and time point responses are recommended activities that both imaging and clinical reviewers should complete prior to the start of on-study reads.

Borderline and challenging cases should be involved in the training; the number of cases to be trained should be dependent upon the study design and experience of reviewers with the response criteria (best practice is to consider 3 cases as a minimum and it should be more especially in case of less experienced readers or more complex studies) recognizing that statistics on such small sample of training may not be significant.

Monitoring (e.g. intra and inter-reader variability, adjudication rates) is recommended per the FDA guidance documents (*28*) and should be performed for all reviewers regardless of training or experience. Members from the PINTaD recently published additional information on reader variability and monitoring of performance (*29*,*30*). Reader monitoring should start early in the course of the trial to allow for retraining in good time when necessary. Group retraining is recommended based on monitoring results or as periodic follow-up group retraining/reviews to

ensure that all readers are provided with identical information so as not to introduce systematic discordance.

Example of an imaging case report form can be found in Supplemental 1.

Summary table of recommendations can be found in Supplemental Table 1.

CONCLUSION

The PRoLoG initiative has created a platform to gather recommendations from an international group of recognized imaging and clinical expert end-users from academia and industry in the field of lymphoma response assessment to standardize application of the Lugano classification in clinical trials and beyond.

These recommendations are intended for clinical users, at local sites and central facilities, in academic and pharmaceutical clinical trials to enhance standardized acquisition and evaluation with the Lugano Classification, facilitating clinical trial conduct and regulatory review, ultimately leading to improved lymphoma patient outcome.

AKNOWLEDGEMENTS

The authors thank all the members from the PINTaD who participated to the PRoLoG initiative, especially Klaus Noever, Melissa Burkett, Anand Devasthanam, Nick Enus, Andres Forero Torres, Nick Galante, Sayali Karve, Katarina Ludajic, Michael ONeal, Ravikanth Mankala, Daniel Mollura, and Jason Vilardi. In addition, the authors thank Tina Nielsen and John Sunderland for their review of the manuscript, as well as the original authors of the classification and companion paper. Finally, the authors thank Deming Litner for her administrative support and the CDISC group for discussion on the terminology to be used for anatomic and metabolic responses.

Professor Barrington acknowledges support from the National Institute for Health and Care Research (NIHR) [RP-2-16-07-001] and by core funding from the Wellcome/EPSRC Centre for Medical Engineering at King's College London [WT203148/Z/16/Z] and the NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London and the NIHR Clinical Research Facility. The views expressed by Professor Barrington are not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

DISCLOSURE

This project was sponsored by PINTaD and the views expressed are those of the authors, not necessarily from their institution. No financial support was received for this project. Fabien Ricard, employed by Bayer at the time of first submitting this manuscript, is now an employee and has shares of Relay Tx. Paul Galette is employed by and shareholder of GSK. Greg Goldmacher is employed by Merck and has stocks in Merck, ImmunoGen and Aveo. Julie Gillis, formerly employed by Imaging Endpoints, is now at Merigold LLC. Pierre Terve is with Keosys medical imaging. Min Liu has shares at Autolus. Larry Schwartz reports BMS, Regeneron, Merck as member of independent review panel and data safety monitoring board for clinical trials. Jayant Narang and Rudresh Jarecha, formerly employed by Calyx, are with Takeda Pharmaceuticals and Deciphera Pharmaceuticals, respectively. Ron Korn is owner and CMO at Imaging Endpoints Core Lab, consultant for Virginia G Piper Cancer Center, ImaginAB Technologies, and shareholder of Verve Medical, Telelite Health, Globavir, Renibus. No other potential conflicts of interest relevant to this article exist.

KEY POINTS

Question: How can the Lugano classification be consistently applied among imaging end users?

Pertinent Findings: These consensus recommendations should be used as a companion to the Lugano Classification with regards to required imaging series and scan visits, acquisition and reconstruction of PET images. The role of imaging and clinical reviewers as well as on training and monitoring is clarified.

Implications for patient care: This guidance will enhance usage of the Lugano Classification, facilitating clinical trial conduct and regulatory review, ultimately leading to improved lymphoma patient outcome.

REFERENCES

1. 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;32:3059–3067.
2. 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;32:3048–3058.
3. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol*. 2007;25:579–586.
4. https://clinicaltrials.gov (accessed 23Apr2022).
5. https://www.pintad.net (accessed 23Apr2022).
6. Cheson BD, Ansell S, Schwartz L, et al. Refinement of the Lugano classification lymphoma response criteria in the era of immunomodulatory therapy. *Blood*. 2016;128:2489-2496.
7. Younes A, Hilden P, Coiffier B, et al. International working group consensus response evaluation criteria in lymphoma (RECIL 2017). *Ann Oncol.* 2017;28:1436–1447.
8. https://rsna.org/QIBA (accessed 23Apr2022).
9. Kinahan P, Perlman E, Sunderland J, et al. The QIBA Profile for FDG PET/CT as an imaging biomarker measuring response to cancer therapy. *Radiology.* 2020;294:647–657.
10. FDG PET-CT Technical Committee. 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: https://rsna.org/QIBA (accessed 23Apr2022).

http://qibawiki.rsna.org/images/1/1f/QIBA_FDG-PET_Profile_v113.pdf (accessed 23Apr2022).

11. FDG PET-CT UPICT V 2.0 Full document, publicly reviewed version (Dec 2014). Available from: https://rsna.org/QIBA (accessed 23Apr2022). http://qibawiki.rsna.org/images/7/71/UPICT_Oncologic_FDG-PETCTProtocol_Dec-2014a.pdf (accessed 23Apr2022).
12. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42:328–354.
13. Barrington SF, Kluge R. FDG PET for therapy monitoring in Hodgkin and non-Hodgkin lymphomas. *Eur J Nucl Med Mol Imaging*. 2017;44(suppl 1):97-110.
14. Schöder H and Moskowitz CH. Metabolic tumor volume in lymphoma: hype or hope? *J Clin Oncol*. 2016;34:3591–3594.
15. Kostakoglu L and Chauvie S. Metabolic tumor volume metrics in lymphoma. *Semin Nucl Med.* 2018;48:50-66.
16. Lin C, Itti E, Haioun C, et al. Early 18F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. *J Nucl Med*. 2007;48:1626–32.
17. 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 ΔSUVmax. *Eur J Nucl Med Mol Imaging*. 2013;40:1312–20.
18. Casasnovas RO, Meignan M, Berriolo-Riedinger A, et al. SUVmax reduction improves early prognosis value of interim positron emission tomography scans in diffuse large B-cell lymphoma. *Blood*. 2011;118:37–43.

19. Eertink J, Burggraaff C, Heymans M, et al. Optimal timing and criteria of interim PET in DLBCL: a comparative study of 1692 patients. *Blood Adv*. 2012;5:2375-2384.
20. Le Gouill S, Ghesquieres H, Oberic L, et al. Obinutuzumab vs rituximab for advanced DLBCL: a PET-guided and randomized phase 3 study by LYSA. *Blood*. 2021;137:2307-2320.
21. Rekowski J, Hüttmann A, Schmitz C, et al. Interim PET evaluation in diffuse large B-cell lymphoma using published recommendations: comparison of the Deauville 5-Point scale and the ΔSUV max method. *J Nucl Med.* 2021;62:37-42.
22. Kurch L, Andreas H, Georgi TM, et al. Interim positron emission tomography in diffuse large B-cell lymphoma. *J Nucl Med.* 2021;62:1068-1074.
23. Schöder H, Polley MYC, Knopp MV, et al. Prognostic value of interim FDG PET in diffuse large cell lymphoma: results from the CALGB 50303 Clinical Trial. *Blood*. 2020;135:2224-2234.
24. Raylman R, Kison P, Wahl R. Capabilities of two- and three-dimensional FDG-PET for detecting small lesions and lymph nodes in the upper torso: a dynamic phantom study. *Eur J Nucl Med*. 1999;26:39-45.
25. de Langen AJ, Vincent A, Velasquez LM, et al. Repeatability of 18F-FDG uptake measurements in tumors: a metaanalysis. *J Nucl Med*. 2012;53:701-708.
26. Itti E, Juweid ME, Haioun C, et al. Improvement of early 18F-FDG PET interpretation in diffuse large B-cell lymphoma: importance of the reference background. *J Nucl Med*. 2010;51:1857-1862.
27. Wahl RL, Jacene H, Kasamaon Y, et al. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50(suppl 1):122S-50S.

28. Clinical Trial Imaging Endpoint Process Standards: Guidance for Industry. U.S. Department of Health and Human Services. Food and Drug Administration Center for Drug Evaluation and Research (CDER). Center for Biologics Evaluation and Research (CBER). April 2018, Clinical/Medical. https://www.fda.gov/files/drugs/published/Clinical-Trial-Imaging-Endpoint-Process-Standards-Guidance-for-Industry.pdf (accessed 23Apr2022)
29. Schmid AM, Raunig DL, Miller CG, et al. Radiologists and clinical trials: Part 1 the truth about reader disagreements. *Ther Innov Regul Sci*. 2021;55:1111–1121.
30. Raunig DL, Schmid AM, Miller CG, et al: Radiologists and clinical trials: Part 2: practical statistical methods for understanding and monitoring independent reader performance. *Ther Innov Regul Sci.* 2021;55:1122-1138.

GRAPHICAL ABSTRACT