Response Assessment Criteria and Their Applications in Lymphoma: Part 1

Mateen C. Moghbel[1], Lale Kostakoglu[2], Katherine Zukotynski[3], Delphine L. Chen[4], Helen Nadel[5], Ryan Niederkohr[6], and Erik Mittra[1]

[1]Stanford University Medical Center, Stanford, California; [2]Mount Sinai Medical Center, New York, New York; [3]McMaster University, Hamilton, Ontario, Canada; [4]Washington University, St. Louis, Missouri; [5]University of British Columbia, Vancouver, British Columbia, Canada; and [6]Kaiser Permanente, Santa Clara, California

The effectiveness of cancer therapy, both in individual patients and across populations, requires a systematic and reproducible method for evaluating response to treatment. Early efforts to meet this need resulted in the creation of numerous guidelines for quantifying posttherapy changes in disease extent, both anatomically and metabolically. Over the past few years, criteria for disease response classification have been developed for specific cancer histologies. To date, the spectrum of disease broadly referred to as lymphoma is perhaps the most common for which disease response classification is used. This review article provides an overview of the existing response assessment criteria for lymphoma and highlights their respective methodologies and validities. Concerns over the technical complexity and arbitrary thresholds of many of these criteria, which have impeded the long-standing endeavor of standardizing response assessment, are also discussed.

Key Words: lymphoma; PET; CT; RECIST; PERCIST

Lymphoma comprises a heterogeneous collection of lymphoproliferative malignancies with varying clinical behaviors and response profiles. These disorders are commonly categorized as either Hodgkin lymphoma (HL) or non-Hodgkin lymphoma (NHL), with the latter group constituting most cases. HL tends to be less aggressive and carries a relatively high 5-y survival rate of 85.3% (1). In 2015, this subtype of lymphoma was diagnosed in an estimated 9,050 patients and caused 1,150 deaths in the United States

(2). By comparison, NHL includes dozens of distinct conditions with varying etiologies and prognoses. Together, these conditions accounted for approximately 71,850 new cases and 19,790 deaths in the United States in 2015 (2), with a 5-y survival rate of 69.3% (1). The guidelines of the World Health Organization (WHO) subdivide NHL according to cell lineage into mature B-cell neoplasms and mature T-cell and NK-cell neoplasms (3). Diffuse large B-cell lymphoma, which falls into the first classification, represents approximately 40% of all cases of NHL, making it the most common form of the disease (4).

The nodular enlargements characteristic of lymphoma were noted in the medical literature as early as 1661 (5), but the constellation of "lymph node and spleen enlargement, cachexia and fatal termination" was first described by Thomas Hodgkin in 1832 (6). The development of modern treatments occurred over a century later, when the discovery of marked lymphoid and myeloid suppression in soldiers exposed to mustard gas during the Second World War led Louis S. Goodman and Alfred Gilman to test the effects of a related compound—nitrogen mustard—on patients with lymphoma and other hematologic diseases (7).

Even these early chemotherapeutic agents required an objective means of evaluating their in vivo effectiveness in human subjects. Initially, standardized methods for the manual measurement of tumor size before and after therapy were proposed for this purpose. But with the advent of anatomic medical imaging techniques, most notably CT, an array of novel guidelines for response assessment was developed. More recently, functional information from PET has been integrated to complement the anatomic information of CT. Currently, numerous criteria that rely on CT and PET individually, as well as a handful of criteria that combine these imaging modalities, have been reported for assessing treatment response in both solid tumors and hematologic malignancies (Supplemental Table 1; supplemental materials are available at http://jnm.snmjournals.org). Although progress has recently been made toward the standardization of response assessment, the clinical and research communities remain somewhat fragmented in their use of these various criteria. This review article outlines the

available criteria and highlights their differences in an attempt to facilitate a more uniform approach to response assessment.

HISTORICAL REVIEW OF RESPONSE ASSESSMENT IN SOLID TUMORS

From the development of the first chemotherapeutic agents in the 1940s to the advent of modern imaging techniques in the 1970s, objective and systematic assessment of treatment response depended largely on physical examination [8]. However, palpation as a method of assessing response was imprecise, as demonstrated by a 1976 study by Moertel and Hanley in which 16 oncologists palpated and measured 12 simulated tumor masses using "variable clinical methods" [9]. The authors found that criteria defining response as 25% and 50% reductions in the perpendicular diameters of these palpated tumors resulted in false-positive interpretations in 19%–25% and 6.8%–7.8% of cases, respectively.

With the goal of achieving "the standardization of reporting results of cancer treatment," WHO held a series of meetings between 1977 and 1979 that culminated in the publication of a handbook outlining response assessment criteria, which were subsequently widely publicized and rapidly adopted [10,11]. The criteria called for bidimensional tumor measurements to be obtained before and after therapy and the product of these bidimensional measurements to be calculated and summed across several sites of disease to form a single parameter by which to assess response. The changes in these parameters over time classified patients into 1 of 4 response groups: complete response, partial response, no response, and progressive disease (Supplemental Table 2).

Although these guidelines made strides toward standardization of response assessment, they did not explicitly specify critical factors, including the number of masses to be measured and the minimum measureable size of a tumor [12]. As a result of these ambiguities, as well as the introduction of imaging modalities such as CT, the WHO criteria eventually became the subject of reinterpretation by various research organizations and clinical groups, thus undermining the standardization it was designed to promote.

To address the gradual divergence of response assessment, institutions such as the National Cancer Institute and the European Organization for Research and Treatment of Cancer (EORTC) began revisiting the WHO criteria throughout the 1990s with the goal of developing new guidelines that would restandardize the practice of evaluating response to therapy. In 1999, the EORTC released its own recommendations for preimaging patient preparation, image acquisition and analysis, tumor sampling, and tumor response classification [13]. These were among the first guidelines to use a functional imaging modality, namely PET, as a means of assessing treatment response (Supplemental Table 3). The PET radiotracer ^{18}F-FDG was used to measure metabolic activity and tumor aggressiveness. Moreover, ^{18}F-FDG was shown to delineate the metabolically active tumor borders, providing insight into individual tumor biology. These metabolic classifications of treatment response laid the groundwork for similar ^{18}F-FDG–based criteria in the years that followed.

The incorporation of PET imaging helped to address the issue of residual masses detected after therapy, which frequently comprise inflammatory, necrotic, and fibrotic tissue rather than residual disease [14–16]. This phenomenon proved especially problematic for lymphoma, for which the response assessment criteria relied solely on anatomic imaging. Approximately 40% of NHL patients and 20% of HL patients continue to exhibit residual mediastinal or abdominal masses on CT after therapy [17,18]. In studies that restaged such patients via laparotomy, between 80% and 95% of residual masses were shown to be nonmalignant on pathology [17,19]. Moreover, the presence of residual masses on imaging was found not to be associated with time to relapse or survival [18]. Therefore, by shedding light on the metabolic activity and thereby viability of these masses, PET overcame a significant limitation of CT-based response assessment for lymphoma [20].

In 2000, shortly after the EORTC devised its PET-based criteria, a collaboration between the National Cancer Institute and EORTC provided a new set of CT-based guidelines called Response Evaluation Criteria in Solid Tumors (RECIST) [21]. Unlike earlier anatomic criteria [11,22], RECIST assessed tumor response on the basis of unidimensional measurements made on CT along the tumor's longest axis, rendering the process more reproducible and applicable to the clinical setting. RECIST also defined the parameters that had been the source of disagreement between groups implementing the WHO criteria: the maximum number of lesions to be measured was set at 10, with a maximum of 5 per organ, and the minimum size of a lesion to be measured was set at 1 cm. Finally, RECIST redefined the response categories that were established in the WHO criteria (Table 1). These reformulated classifications were conservative relative to the WHO criteria, placing fewer patients in the progressive disease category [21,23,24].

However, RECIST was not without shortcomings. RECIST was widely reported to be less suitable for particular cancers, such as mesothelioma and pediatric tumors [23,25,26]. Furthermore, the arbitrary number of tumor foci to be measured according to the criteria and the relatively narrow definition of progressive disease were points of contention [27]. It was also suggested that the routine clinical implementation of RECIST would significantly increase the workload of radiologists [28].

To address these limitations, the RECIST Working Group set out to amend the criteria, publishing "RECIST 1.1" in 2009 [29]. There were a handful of significant changes both to simplify and clarify the criteria and to allow for application in additional cancers and modalities. First, the maximum number of measured tumors was reduced to five, with a maximum of two per organ. This amendment was based on data showing that such a reduction did not result in a significant loss of information [30]. Second, the definition of progressive disease was changed to require a minimum absolute increase of 5 mm in the sum of the tumor diameters, thereby preventing changes in individual small lesions from leading to unnecessary classifications of progression. Third, specific guidelines were established for

TABLE 1
Tumor Response Classifications of RECIST (2000)

Classification	Criteria
Complete response	Disappearance of all tumors for ≥4 wk
Partial response	Shrinkage of tumor by at least 30% for >4 wk
Stable disease	<30% reduction and 20% increase in tumor size
Progressive disease	Increase in tumor size by ≥20%

the assessment of lymph node involvement, defining nodes spanning at least 15 mm on their short axis as assessable target lesions and nodes shrinking to less than 10 mm on their short axis as normal. Finally, the criteria paved the way for the incorporation of information from functional imaging modalities such as PET.

In the same year, 2009, Wahl et al. published a paper outlining "PET Response Criteria in Solid Tumors" (PERCIST) (12). These criteria followed several earlier guidelines for response assessment that used PET, namely those proposed by the EORTC in 1999 (13), Hicks et al. in 2001 (31), and Juweid et al. in 2005 (32,33). PERCIST uses similar criteria to those developed for RECIST, but incorporation of the metabolic information to anatomic information sets it apart. The authors stated that CT alone possesses "poor predictive ability" because the residual masses that are detected by this modality often reflect scarring that is mistaken for active tumor. As a PET-based criterion for response assessment, PERCIST was "designed to facilitate trials of drug development, but, if sufficiently robust, could be applied to individual patients" (12).

In their report outlining PERCIST, Wahl et al. specified a host of parameters that would facilitate the standardization of PET-based response assessment once the criteria were widely adopted. Among these suggestions was a proposed maximum of 5 tumor foci of the highest ^{18}F-FDG avidity, with up to 2 foci per organ, to be measured for comparison before and after therapy. It was also recommended that patients undergo ^{18}F-FDG PET scans at least 10 d after an early cycle of chemotherapy to maximize the prognostic value of the scan and minimize the effect of ^{18}F-FDG–avid inflammation caused by chemotherapy and radiation. Moreover, the authors called for the SUVs derived from a PET scan to be corrected for lean body mass (SUL) and compared with reference uptake in the liver or, if necessary, background blood pool (12). Finally, PERCIST retained the same 4 response classifications that were established in RECIST but amended their respective specifications (Table 2). Although not yet fully validated, the PERCIST criteria are increasingly used in clinical trials for assessing therapy response in cancer (34). Such data will potentially help support their more widespread clinical application.

MODERN RESPONSE ASSESSMENT CRITERIA IN LYMPHOMA

Although the guidelines included in the WHO criteria, RECIST, and PERCIST are generalizable to a wide array of cancers, several specialized criteria have also been proposed specifically for the spectrum of hematologic malignancies. As early as the late 1980s, as guidelines began to be developed for response assessment in chronic lymphocytic leukemia (35), HL (36), and acute myelogenous leukemia (37), there were calls for similar efforts toward standardization in NHL (38). However, in the decade that followed, various organizations simply adapted existing criteria to create their own guidelines for response assessment in NHL, thereby hindering the ability to compare data across different groups.

At meetings sponsored by the National Cancer Institute in February and May 1998, an international working group that comprised both American and European experts reached a consensus on response assessment criteria specifically for NHL (39). The resulting International Working Criteria (IWC) defined anatomic parameters, obtained by clinical or radiologic examination, that could be used to group patients into the traditional classifications of complete response, partial response, stable disease, and progressive disease, as well a new classification of "unconfirmed complete response" (Table 3). To support these anatomically based criteria, the IWC defined the upper limit for the size of a normal lymph node as 1 cm along its short axis on the basis of several prior studies (40–42). In the years after its publication, the IWC were also adopted for HL (43).

In 2005, Juweid et al. integrated the originally CT-based IWC with ^{18}F-FDG PET to create the IWC+PET criteria, which were initially designed and validated for NHL (32) but were subsequently validated for HL as well (44). Citing the prevalence of posttherapy residual masses and the unique ability of PET to accurately predict tumor viability in these masses, the investigators sought to establish a standardized approach that would join the anatomic information of CT with the functional information of PET. The IWC+PET criteria retained the classifications of the original IWC criteria but amended the guidelines to incorporate PET findings (Supplemental Table 4). Juweid et al. found that IWC+PET was a better predictor of progression-free survival than IWC in NHL.

Two publications in 2007, one authored by Cheson et al. (45) and the other by Juweid et al. (33), amended the existing IWC+PET criteria and made recommendations for their clinical use in both HL and NHL as the International Harmonization Project. To avoid false-positive results on PET as a result of therapy-induced inflammation, which can persist for as long as 2 wk after chemotherapy and 3 mo after radiation therapy, both reports recommended that PET acquisition occur at least 3 wk, and preferably 6–8 wk, after chemotherapy and 8–12 wk after radiation therapy. Cheson et al. also addressed the possibility of false-positive PET findings due to "rebound thymic hyperplasia, infection, inflammation, sarcoidosis, and brown fat," as well as "[spatial] resolution…technique, and variability of ^{18}F-FDG avidity among histologic subtypes" (45). For evaluating the tumor viability of residual masses larger than 2 cm in their greatest transverse diameters, mediastinal blood-pool activity was recommended as a reference. On the other hand, for residual masses smaller than 2 cm, background activity was the recommended reference. Residual hepatic and splenic lesions larger than 1.5 cm detected on CT were deemed positive if their metabolic activity was higher than that of the liver and spleen. These amendments permitted the elimination of the unconfirmed complete response category of tumor response, returning the classification scheme to the classic tetrad of complete response, partial response, stable disease, and progressive disease (Table 4).

An international workshop that first met in Deauville, France, in 2009 conceived of novel criteria for both HL and NHL that signaled a significant change on multiple fronts (46–49). In contrast to the predominantly quantitative guidelines proposed previously, the Deauville 5-point scoring system (D5PS) assessed

TABLE 2
Tumor Response Classifications of PERCIST (2009)

Classification	Criteria
Complete response	Decline of metabolic activity of measureable lesions to below mean liver activity and on a par with background blood-pool activity
Partial response	Reduction of peak SUL of ≥30% and 0.8 SUL unit in measurable lesions
Stable disease	<30% reduction and 30% increase in peak SUL in measurable lesions
Progressive disease	Increase in peak SUL by ≥30% and 0.8 SUL unit in measurable lesions

TABLE 3
Tumor Response Classifications of International Workshop Criteria (1999)

Classification	Criteria
Complete response	Total disappearance of all signs and symptoms of disease on clinical and radiographic evaluations
	Regression of nodal masses to normal size in greatest transverse diameter (≤1.5 cm in nodes >1.5 cm before therapy, ≤1 cm in nodes 1.1–1.5 cm before therapy)
	Regression of spleen to normal size so as not to be palpable on physical examination
	Clear infiltrate on bone marrow aspiration and biopsy in sites that were previously involved
Unconfirmed complete response	Fulfillment of complete-response requirements, except:
	Presence of residual lymph nodes >1.5 cm that have regressed by >75% in sum of product of their greatest diameters (SPD)
	Bone marrow aspiration indeterminate for infiltration
Partial response	Regression of 6 largest nodal masses by ≥50% in SPD
	No increase in size of other nodes, liver, or spleen
	Regression of splenic and hepatic nodules by ≥50% in SPD
	No new sites of disease
Stable disease	Results not meeting the criteria for partial response or progressive disease
Progressive disease	≥50% increase in SPD of lymph node mass
	Appearance of any new lesions

treatment response qualitatively—specifically, in the form of a 5-point scale that graded the intensity of ^{18}F-FDG uptake relative to the reference activity of the mediastinal blood pool and liver (Table 5) (50). The technical simplicity of this classification system facilitated its widespread clinical adoption. Moreover, the D5PS became a standard-bearer for the rising trend of interim response assessment, which enabled improved determinations of prognosis and earlier treatment modifications during the course of therapy.

The D5PS have since been modified by a comprehensive set of recommendations developed at the 11th International Conference on Malignant Lymphomas in 2011 and presented at the Fourth International Workshop on PET in Lymphoma, held in Menton, France, in 2012, and at the 12th International Conference on Malignant Lymphomas, convened in Lugano, Switzerland, in 2013 (51,52). The consensus revision of both the staging criteria and the 2007 IWG response criteria led to the development of the Lugano classification, in which separate sets of response criteria were proposed for PET and CT imaging, although the former is generally preferred for ^{18}F-FDG–avid lymphomas. The PET-based criteria built on the 5-point categoric scale established by D5PS by adding considerations for new or recurring involvement of lymph nodes and bone marrow as well as organomegaly (Table 6) (53). Stand-alone CT-based guidelines were also included, despite the known limitations of anatomic response assessment in ^{18}F-FDG–avid lymphoma, for use when PET/CT imaging is unavailable or when lymphomas have low or variable ^{18}F-FDG avidity.

METHODOLOGIC COMPARISON OF EXISTING RESPONSE ASSESSMENT CRITERIA

The various therapy response criteria discussed in this review apply varying approaches to the use of imaging modalities. The RECIST and 1999 IWC criteria primarily use CT; EORTC and PERCIST rely on PET; and the 2007 IWC and Lugano classifications make use of both modalities, with the former using the International Harmonization Project criteria and the latter the D5PS criteria for PET interpretation. The assorted definitions of the response classifications across these criteria are shown in Table 7, which presents a simplified and standardized scheme comprising 4 groupings: complete response, partial response, stable disease, and progressive disease. Although there are identifiable trends across criteria, even those using the same modality demonstrate considerable variability in their thresholds for each response

TABLE 4
Tumor Response Classifications of International Harmonization Project (2007)

Classification	Criteria
Complete response	PET-negative nodes of any size or PET-positive nodes that have regressed to normal size
Partial response	Reduction in SPD of 6 largest masses by ≥50% (and no growth in other nodes); PET-positive at previously uninvolved site or PET-negative
Stable disease	PET-positive at sites of previous involvement but not at new sites; no change in size of nodes on CT
Progressive disease	Appearance of new lesions >1.5 cm or 50% growth in SPD of existing nodes; PET-positive nodes

TABLE 5
Tumor Response Classifications of D5PS (2009)

Classification	Criteria
1	No uptake > background activity
2	Uptake ≤ mediastinal blood-pool activity
3	Uptake between mediastinal blood pool and liver activity
4	Uptake moderately > liver activity
5	Uptake markedly > liver activity

classification. For example, progressive disease is defined as a tumor size increase of at least 20% by RECIST, at least 25% by the WHO criteria, and at least 50% by IWC.

In recent years, the relative simplicity of the D5PS and the associated Lugano classification have distinguished them from their quantitative predecessors, whose technical demands and complexity often precluded their widespread clinical use. However, questions about the reproducibility of the simplified qualitative criteria remain. The literature includes several comparisons between the D5PS and other guidelines using functional imaging for response assessment. A 2010 study of diffuse large B-cell lymphoma patients by Horning et al. compared interobserver agreement in the D5PS and the International Harmonization Project–based Eastern Cooperative Oncology Group criteria,

reporting κ values of 0.502 and 0.445, respectively (54). However, this study was limited to a small study population. Another study on diffuse large B-cell lymphoma, by Itti et al., found lower interobserver agreement with D5PS (κ = 0.66) than with a semiquantitative counterpart based on SUV_{max} (κ = 0.83) (55). In larger standardized studies of HL that used D5PS, Barrington et al., Furth et al., and Gallamini et al. reported κ-values of 0.79–0.85, 0.748, and 0.69–0.84, respectively, proving its superiority (56–58). The implications of these findings on the reproducibility and clinical applicability of the Lugano classification have yet to be determined in prospective studies with large datasets.

FUTURE TRENDS

The recent advent and adoption of the D5PS and Lugano classifications have marked a step toward standardization of interpretation and brought a relatively more objective system. However, this qualitative system of response assessment should also be tested against quantitative criteria to determine their relative effectiveness in patient management. In an earlier study, the tradeoff between simplicity and reproducibility was studied by Lin et al., who compared the prognostic ability of qualitative and quantitative PET analysis in patients with diffuse large B-cell lymphoma (59). Visual analysis was able to predict event-free survival with an accuracy of 65.2%, whereas SUV-based analysis did so with an accuracy of 76.1%. A reduction of 65.7% in the SUV_{max} of an interim PET scan was found to be the optimal cutoff value in differentiating between favorable and unfavorable responses to therapy. These earlier results

TABLE 6
Tumor Response Classifications of Lugano Criteria (2014)

Classification	PET/CT-based criteria	CT-based criteria
Complete response	1, 2, or 3 points on D5PS No new lesions No bone marrow involvement	>1.5 cm along longest transverse diameter Regression of enlarged organs to normal size No new lesions No bone marrow involvement
Partial response	4 or 5 points on D5PS Reduced uptake compared with baseline No new lesions Residual bone marrow uptake that is reduced from baseline	>50% reduction from baseline in sum of product of perpendicular diameters of up to 6 nodes >50% reduction from baseline in size of enlarged spleen No new lesions
Stable disease	4 or 5 points on D5PS Unchanged uptake compared with baseline No new lesions Unchanged bone marrow involvement	<50% reduction from baseline in sum of product of perpendicular diameters of up to 6 nodes No new lesions
Progressive disease	4 or 5 on D5PS Increased uptake compared with baseline New or recurrent involvement in nodes and bone marrow demonstrated by ^{18}F-FDG avidity	>50% increase in product of perpendicular diameters of node Increase in nodal diameter (by 0.5 cm if node is ≤2 cm, 1.0 cm if >2 cm) New or recurrent splenomegaly New or recurrent involvement of nodes and bone marrow

TABLE 7
Comparison of Simplified Classifications of Various Response Criteria

Criteria	Classification			
	CR	PR	SD	PD
WHO	100% reduction in tumor size	≥50% reduction in tumor size	≥25% and ≤50% reduction in tumor size	≥25% increase in tumor size
EORTC	Reduction of ^{18}F-FDG uptake to background levels	≥15% reduction in ^{18}F-FDG uptake	≤15% reduction and ≤25% increase in ^{18}F-FDG uptake	≥25% increase in ^{18}F-FDG uptake
RECIST	100% reduction in tumor size	≥30% reduction in tumor size	≤30% reduction and ≤20% increase in tumor size	≥20% increase in tumor size
PERCIST	Reduction of ^{18}F-FDG uptake to level of background blood pool	≥30% reduction in peak SUL	≤30% reduction or increase in peak SUL	≥20% increase in peak SUL
IWC	Reduction of nodes to normal size	≥50% reduction in size of 6 largest nodes	≤50% reduction or increase in size of nodes	≥50% increase in size of nodes
IWC+PET	CR by IWC plus negative PET scan	CR/PR by IWC plus positive PET scan	SD by IWC plus positive PET scan	PD by IWC plus positive PET scan
International Harmonization Project	PET− nodes or PET+ nodes of normal size	≥50% reduction in size of 6 largest nodes	New PET+ nodes	New PET+ ≥1.5 cm or ≥50% increase in size of existing nodes
D5PS*	^{18}F-FDG uptake at background level	^{18}F-FDG uptake ≤ mediastinal blood pool/liver activity	Mediastinal blood-pool activity ≤ ^{18}F-FDG uptake ≤ liver activity	^{18}F-FDG uptake > liver activity
Lugano				
PET/CT	Normalized ^{18}F-FDG uptake (1–3 on D5PS)	Reduced ^{18}F-FDG uptake (4–5 on D5PS)	Unchanged ^{18}F-FDG uptake (4–5 on D5PS)	Increased ^{18}F-FDG uptake (4–5 on D5PS) + new lesions
CT	Reduction of nodes/ organs to normal size	≥50% reduction in size of up to 6 nodes/spleen	<50% reduction in size of up to 6 nodes	≥50% increase in size of node + new lesions

*D5PS was converted from 5-point scale to 4 categories of CR, PR, SD, and PD (1 = CR, 2 = PR, 3 = SD, 4/5 = PD).
CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease.

suggest that, if optimized for clinical use, standardized PET criteria using a quantitative method may be more adept at assessing tumor response. However, various study biases and the suboptimal technical methodology inherent in the retrospective design of prior studies make it difficult to arrive at a firm conclusion. Moreover, multiple factors, including variability in instrumentation, scanner calibration, and human biology, complicate obtaining reliable PET measurements across medical centers in future studies (*34*). Thus, the most reproducible and accurate method for PET quantification remains to be determined. As techniques for automated segmentation and quantification continue to improve, these advancements will likely be more readily implemented in the clinical setting and facilitate the use of quantitative response assessment.

CONCLUSION

Over the past 6 decades, the techniques used for evaluating the efficacy of cancer therapies have steadily increased both in precision and in intricacy, moving from crude manual measurement toward more complex structural and functional data acquisition, with many more advanced techniques such as heterogeneity measures, parametric mapping, and kinetic acquisitions on the way. The integration of CT and PET in particular greatly enhanced the ability to assess disease progression, adjust therapeutic regimens, and form an accurate prognosis. However, the vast array of interpretative guidelines that were introduced, each with its own protocols and thresholds, created stifling methodologic variability and sparked calls for "harmonization" that have rung out since the early days of response assessment and continue to echo to this day.

With respect to lymphoma in particular, recent PET/CT-based criteria have made significant strides toward standardization, and their simplified qualitative guidelines have remedied the technical complexity and time intensity that impeded the clinical application of prior quantitative criteria. However, their suitability for certain scenarios, such as in patients with lymphomas of low or variable ^{18}F-FDG avidity or in those receiving immunochemotherapy or biologic therapy, remains to be determined. Moreover, as technologic

advances ease the use of quantitative criteria, continued efforts to maintain harmonization in response assessment will likely be necessary to avoid renewed fragmentation.

REFERENCES

1. Howlader N, Noone A, Krapcho M, et al. *SEER Cancer Statistics Review, 1975–2011.* Bethesda, MD: National Cancer Institute; 2014.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin.* 2015;65:5–29.

3. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood.* 2009;114:937–951.

4. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* 2000;403:503–511.

5. Hoster HA, Dratman MB. Hodgkin's disease, 1832–1947. *Cancer Res.* 1948;8:1–48.

6. Hodgkin T. On some morbid appearances of the absorbent glands and spleen. *Med Chir Trans.* 1832;17:68–114.

7. Goodman LS, Wintrobe MM, Dameshek W, Goodman MJ, Gilman A, McLennan MT. Nitrogen mustard therapy: use of methyl-bis (beta-chloroethyl) amine hydrochloride and tris (beta-chloroethyl) amine hydrochloride for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *J Am Med Assoc.* 1946;132:126–132.

8. Gehan EA, Schneiderman MA. Historical and methodological developments in clinical trials at the National Cancer Institute. *Stat Med.* 1990;9:871–880.

9. Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer.* 1976;38:388–394.

10. *WHO Handbook for Reporting Results of Cancer Treatment.* Geneva, Switzerland: World Health Organization; 1979.

11. Miller A, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer.* 1981;47:207–214.

12. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50(suppl):122S–150S.

13. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer.* 1999;35:1773–1782.

14. Canellos G. Residual mass in lymphoma may not be residual disease. *J Clin Oncol.* 1988;6:931–933.

15. Reske S. PET and restaging of malignant lymphoma including residual masses and relapse. *Eur J Nucl Med Mol Imaging.* 2003;30(suppl):S89–S96.

16. Lewis E, Bernardino ME, Salvador PG, Cabanillas FF, Barnes PA, Thomas JL. Post-therapy CT-detected mass in lymphoma patients: is it viable tissue? *J Comput Assist Tomogr.* 1982;6:792–795.

17. Surbone A, Longo DL, DeVita V, et al. Residual abdominal masses in aggressive non-Hodgkin's lymphoma after combination chemotherapy: significance and management. *J Clin Oncol.* 1988;6:1832–1837.

18. Orlandi E, Lazzarino M, Brusamolino E, et al. Residual mediastinal widening following therapy in Hodgkin's disease. *Hematol Oncol.* 1990;8:125–131.

19. Fuks JZ, Aisner J, Wiernik PH. Restaging laparotomy in the management of the non-Hodgkin lymphomas. *Med Pediatr Oncol.* 1982;10:429–438.

20. Mikhaeel N, Timothy A, Hain S, O'Doherty M. 18-FDG-PET for the assessment of residual masses on CT following treatment of lymphomas. *Ann Oncol.* 2000;11(suppl 1):147–150.

21. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst.* 2000;92:205–216.

22. Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. *Invest New Drugs.* 1992;10:239–253.

23. Therasse P, Eisenhauer E, Verweij J. RECIST revisited: a review of validation studies on tumour assessment. *Eur J Cancer.* 2006;42:1031–1039.

24. Verweij J, Therasse P, Eisenhauer E. Cancer clinical trial outcomes: any progress in tumour-size assessment? *Eur J Cancer.* 2009;45:225–227.

25. McHugh K, Kao S. Response evaluation criteria in solid tumours (RECIST): problems and need for modifications in paediatric oncology? *Br J Radiol.* 2003;76:433–436.

26. Barnacle A, McHugh K. Limitations with the response evaluation criteria in solid tumors (RECIST) guidance in disseminated pediatric malignancy. *Pediatr Blood Cancer.* 2006;46:127–134.

27. Eisenhauer E. Response evaluation: beyond RECIST. *Ann Oncol.* 2007;18:ix29–ix32.

28. Padhani A, Ollivier L. The RECIST criteria: implications for diagnostic radiologists. *Br J Radiol.* 2001;74:983–986.

29. Eisenhauer E, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45:228–247.

30. Bogaerts J, Ford R, Sargent D, et al. Individual patient data analysis to assess modifications to the RECIST criteria. *Eur J Cancer.* 2009;45:248–260.

31. Hicks RJ, Kalff V, MacManus MP, et al. The utility of 18F-FDG PET for suspected recurrent non–small cell lung cancer after potentially curative therapy: impact on management and prognostic stratification. *J Nucl Med.* 2001;42:1605–1613.

32. Juweid ME, Wiseman GA, Vose JM, et al. Response assessment of aggressive non-Hodgkin's lymphoma by Integrated International Workshop criteria and fluorine-18–fluorodeoxyglucose positron emission tomography. *J Clin Oncol.* 2005;23:4652–4661.

33. Juweid ME, Stroobants S, Hoekstra OS, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *J Clin Oncol.* 2007;25:571–578.

34. Bai B, Bading J, Conti PS. Tumor quantification in clinical positron emission tomography. *Theranostics.* 2013;3:787–801.

35. Cheson BD, Bennett JM, Rai KR, et al. Guidelines for clinical protocols for chronic lymphocytic leukemia: recommendations of the National Cancer Institute-sponsored working group. *Am J Hematol.* 1988;29:152–163.

36. Lister T, Crowther D, Sutcliffe S, et al. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *J Clin Oncol.* 1989;7:1630–1636.

37. Cheson BD, Cassileth PA, Head DR, et al. Report of the National Cancer Institute-sponsored workshop on definitions of diagnosis and response in acute myeloid leukemia. *J Clin Oncol.* 1990;8:813–819.

38. Dixon DO, McLaughlin P, Hagemeister F, et al. Reporting outcomes in Hodgkin's disease and lymphoma. *J Clin Oncol.* 1987;5:1670–1672.

39. Cheson BD, Horning SJ, Coiffier B, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. *J Clin Oncol.* 1999;17:1244.

40. Glazer G, Gross B, Quint L, Francis I, Bookstein F, Orringer M. Normal mediastinal lymph nodes: number and size according to American Thoracic Society mapping. *AJR.* 1985;144:261–265.

41. Kiyono K, Sone S, Sakai F, et al. The number and size of normal mediastinal lymph nodes: a postmortem study. *AJR.* 1988;150:771–776.

42. Dorfman RE, Alpern MB, Gross B, Sandler M. Upper abdominal lymph nodes: criteria for normal size determined with CT. *Radiology.* 1991;180:319–322.

43. Hampson F, Shaw A. Response assessment in lymphoma. *Clin Radiol.* 2008;63:125–135.

44. Brepoels L, Stroobants S, De Wever W, et al. Hodgkin lymphoma: response assessment by revised International Workshop Criteria. *Leuk Lymphoma.* 2007;48:1539–1547.

45. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol.* 2007;25:579–586.

46. 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;50:1257–1260.

47. Gallamini A, Fiore F, Sorasio R, Meignan M. Interim positron emission tomography scan in Hodgkin lymphoma: definitions, interpretation rules, and clinical validation. *Leuk Lymphoma.* 2009;50:1761–1764.

48. Meignan M, Gallamini A, Haioun C, Polliack A. Report on the Second International Workshop on Interim Positron Emission Tomography in Lymphoma held in Menton, France, 8-9 April 2010. *Leuk Lymphoma.* 2010;51:2171–2180.

49. Meignan M, Gallamini A, Itti E, Barrington S, Haioun C, Polliack A. Report on the Third International Workshop on Interim Positron Emission Tomography in

Lymphoma held in Menton, France, 26-27 September 2011 and Menton 2011 consensus. *Leuk Lymphoma.* 2012;53:1876–1881.

50. Le Roux P-Y, Gastinne T, Le Gouill S, et al. Prognostic value of interim FDG PET/CT in Hodgkin's lymphoma patients treated with interim response-adapted strategy: comparison of International Harmonization Project (IHP), Gallamini and London criteria. *Eur J Nucl Med Mol Imaging.* 2011;38:1064–1071.

51. 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.

52. Meignan M, Barrington S, Itti E, Gallamini A, Haioun C, Polliack A. Report on the 4th International Workshop on Positron Emission Tomography in Lymphoma held in Menton, France, 3-5 October 2012. *Leuk Lymphoma.* 2014;55:31–37.

53. 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;32:3059–3068.

54. Horning SJ, Juweid ME, Schöder H, et al. Interim positron emission tomography scans in diffuse large B-cell lymphoma: an independent expert nuclear medicine evaluation of the Eastern Cooperative Oncology Group E3404 study. *Blood.* 2010;115:775–777.

55. 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 D5PS and ΔSUVmax. *Eur J Nucl Med Mol Imaging.* 2013;40:1312–1320.

56. Barrington SF, Qian W, Somer EJ, et al. Concordance between four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging.* 2010;37:1824–1833.

57. Furth C, Amthauer H, Hautzel H, et al. Evaluation of interim PET response criteria in paediatric Hodgkin's lymphoma: results for dedicated assessment criteria in a blinded dual-centre read. *Ann Oncol.* 2011;22:1198–1203.

58. Gallamini A, Barrington SF, Biggi A, et al. The predictive role of interim positron emission tomography for Hodgkin lymphoma treatment outcome is confirmed using the interpretation criteria of the Deauville five-point scale. *Haematologica.* 2014;99:1107–1113.

59. Lin C, Itti E, Haioun C, et al. Early ^{18}F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. *J Nucl Med.* 2007;48:1626–1632.