ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

Annals of Oncology 28: 1436–1447, 2017
doi:10.1093/annonc/mdx097
Published online 3 April 2017

REVIEW

International Working Group consensus response evaluation criteria in lymphoma (RECIL 2017)

A. Younes[1*], P. Hilden[2], B. Coiffier[3], A. Hagenbeek[4], G. Salles[3], W. Wilson[5], J. F. Seymour[6], K. Kelly[7], J. Gribben[8], M. Pfreunschuh[9], F. Morschhauser[10], H. Schoder[11], A. D. Zelenetz[1], J. Rademaker[11], R. Advani[12], N. Valente[13], C. Fortpied[14], T. E. Witzig[15], L. H. Sehn[16], A. Engert[17], R. I. Fisher[18], P.-L. Zinzani[19], M. Federico[20], M. Hutchings[21], C. Bollard[22], M. Trneny[23], Y. A. Elsayed[24], K. Tobinai[25], J. S. Abramson[26], N. Fowler[27], A. Goy[28], M. Smith[29], S. Ansell[15], J. Kuruvilla[30], M. Dreyling[31], C. Thieblemont[32], R. F. Little[33], I. Aurer[34], M. H. J. Van Oers[35], K. Takeshita[36], A. Gopal[37], S. Rule[38], S. de Vos[39], I. Kloos[40], M. S. Kaminski[41], M. Meignan[42], L. H. Schwartz[43], J. P. Leonard[44], S. J. Schuster[45] & V. E. Seshan[2]

[1]Lymphoma Service; [2]Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, USA; [3]Hematology, Université Lyon-1, Lyon-Sud Charles Mérieux, Lyon, France; [4]Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; [5]Lymphoid Malignancies Branch, National Cancer Institute, Bethesda, USA; [6]Peter MacCallum Cancer Centre and University of Melbourne, Australia; [7]Pediatrics Department, Roswell-Park Cancer Institute, Buffalo, USA; [8]Department of Haemato-Oncology, Barts Cancer Institute, London, UK; [9]Department of Internal Medicine, Universität des Saarlandes, Homburg, Germany; [10]Department of Hematology, Université de Lille 2, Lille, France; [11]Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York; [12]Department of Oncology, Stanford University, Stanford; [13]Genentech, San Francisco, USA; [14]Statistics, EORTC, Brussels, Belgium; [15]Mayo Clinic, Rochester, USA; [16]British Columbia Cancer Agency, Vancouver, Canada; [17]Department of Internal Medicine, University Hospital of Cologne, Cologne, Germany; [18]Fox Chase Cancer Center, Philadelphia, USA; [19]Department of Hematology, University of Bologna, Bologna; [20]Department of Diagnostic Medicine, University of Modena, Modena, Italy; [21]Department of Hematology, University of Copenhagen, Denmark; [22]Children's National Health System, Washington, USA; [23]Lymphoma and Stem Cell Transplantation Program, Charles University, Prague, Czech Republic; [24]Janssen Research & Development, Spring House, USA; [25]Department of Hematology, National Cancer Center Hospital, Tokyo, Japan; [26] Massachusetts General Hospital, Center for Lymphoma, Boston; [27]U.T. M.D.Anderson Cancer Center, Houston; [28]John Theurer Cancer Center, Hackensack University Medical Center, Hackensack; [29]Cleveland Clinic, Cleveland, USA; [30]Division of Medical Oncology and Hematology, Princess Margaret Cancer Centre, Toronto, Canada; [31]Medicine Clinic III, Ludwig Maximilian University, Munich, Germany; [32]Haemato-Oncology, Hôpital Saint Louis, Paris, France; [33]Divisions of Cancer Treatment and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, USA; [34]Department of Hematology, University Hospital Centre Zagreb, Zagreb, Croatia; [35]Academic Medical Center, Amsterdam, The Netherlands; [36]Celgene, Summit; [37]Fred Hutchinson Cancer Research Center, Seattle, USA; [38]Haematology Department, Plymouth University, UK; [39]Oncology, UCLA, Los Angeles, USA; [40]Servier, Neuilly sur Seine, France; [41]University of Michigan Comprehensive Cancer Center, Ann Arbor, USA; [42]Nuclear Medicine, Hôpitaux Universitaires Henri Mondor, Créteil, France; [43]Columbia University College of Physicians and Surgeons and New York Presbyterian Hospital, New York; [44]Weill Cornell Medicine and and New York Presbyterian Hospital, New York; [45]University of Pennsylvania School of Medicine, Philadelphia, USA

Correspondence to: Dr Anas Younes, Lymphoma Service, Memorial Sloan Kettering Cancer center, 1275 York Avenue, New York, New York, USA. Tel: +001-212-639-7715; E-mail: 10021.younesa@mskcc.org

In recent years, the number of approved and investigational agents that can be safely administered for the treatment of lymphoma patients for a prolonged period of time has substantially increased. Many of these novel agents are evaluated in early-phase clinical trials in patients with a wide range of malignancies, including solid tumors and lymphoma. Furthermore, with the advances in genome sequencing, new "basket" clinical trial designs have emerged that select patients based on the presence of specific genetic alterations across different types of solid tumors and lymphoma. The standard response criteria currently in use for lymphoma are the Lugano Criteria which are based on [^{18}F]2-fluoro-2-deoxy-D-glucose positron emission tomography or bidimensional tumor measurements on computerized tomography scans. These differ from the RECIST criteria used in solid tumors, which use unidimensional measurements. The RECIL group hypothesized that single-dimension measurement could be used to assess response to therapy in lymphoma patients, producing results similar to the standard criteria. We tested this hypothesis by analyzing 47 828 imaging measurements from 2983 individual adult and pediatric lymphoma patients enrolled on 10 multicenter clinical trials and developed new lymphoma response criteria (RECIL 2017). We demonstrate that assessment of tumor burden in lymphoma clinical trials can use the sum of longest diameters of a maximum of three target lesions. Furthermore, we introduced a new provisional category of a minor response. We also clarified response assessment in patients receiving novel immune therapy and targeted agents that generate unique imaging situations.

Key words: response criteria, FDG-PET, targeted therapy, immunotherapy, waterfall plots, lymphoma

Introduction

The National Cancer Institute-sponsored international consensus response criteria for lymphoma guidelines were published in 1999 and were subsequently revised in 2007 to incorporate assessment of tumor metabolism by [^{18}F]2-fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) [1–3]. More recently, these criteria were further refined in the so-called Lugano Classification, to incorporate a scoring system to enhance the reproducibility of the interpretation and reporting of FDG-PET results, in addition to updating the recommended procedures for staging evaluation [4]. These guidelines have facilitated the conduct and enhanced the analysis of clinical trials across different institutions and geographic regions and provided the basis for an objective comparison of the response assessments from different treatment regimens. These guidelines are also frequently used by regulatory agencies in their evaluation and approval processes of new lymphoma drugs. However, these lymphoma response criteria were mainly based on expert opinion and were not supported by large-scale data analysis. As the number of novel antilymphoma targeted drugs that have entered clinical trials has substantially increased; new clinical situations have emerged that were not envisaged in the original criteria developed in the era of cytotoxic chemotherapy [5]. Some of these new agents have unique mechanisms of action and have demonstrated excellent safety profiles allowing extended dosing administration until disease progression. However, the current definition of disease progression and partial remission depends on historically arbitrary criteria that may not adequately reflect an individual patients' clinical benefit and often do not support clinical decisions to continue or to stop therapy [6, 7]. Furthermore, many phase I/II clinical trials of novel agents include patients with both solid tumors and lymphoma, yet response assessment of these two disease categories is based on different criteria, resulting in different interpretations [8]. Moreover, with recent advances in genome sequencing studies and the identification of driver genetic alterations across tumor types, novel clinical "basket" trial designs have emerged for the treatment of patients with different tumors that harbor specific genetic defects across different types of solid tumors and lymphoma. To facilitate the evaluation of lymphomas in the era of precision medicine oncology trials, it is important to align lymphoma response criteria with the response evaluation criteria in solid tumors (RECIST) [9].

The standard response criteria currently in use for lymphoma are the Lugano Criteria which are based on PET or bidimensional tumor measurements on computerized tomography (CT) for non-FDG avid lymphomas, or when PET imaging is not available [4]. These differ from the RECIST criteria used in solid tumors which use unidimensional measurements [9]. In a pilot study, a lymphoma-adapted RECIST was found to be simpler to use than the 2007 lymphoma Response Criteria, while yielding similar response rates [10, 11]. A second pilot study was conducted in 2013 by some of the authors of this manuscript using 175 cases from four major academic centers also supported our hypothesis and lead to the development of this projects (data not shown). With this background, a group of leading international lymphoma experts from academic centers and pharmaceutical companies, radiologists, and statisticians established a collaboration to harmonize the lymphoma response criteria with RECIST, and to evaluate the effect of using bidimensional or unidimensional measurements on the assessment of best response for each subject, the proportion of subjects in each response category, and progression-free survival (PFS). The new response evaluation criteria in lymphoma (RECIL) was introduced and approved at the International Workshop on non-Hodgkin lymphoma (iwNHL) in San Diego on September 25, 2016.

Methods

Data were collected using a predefined purpose-designed template, and the data were transcribed to the template from study-specific case report forms. Measurements were collected from previously measured lesions entered on study-specific case report forms from 10 prospective multicenter trials (supplementary Table S1, available at *Annals of Oncology* online). For each subject, the target lesions (nodal and/or extranodal) were coded by the number of lesions, from 1 to K, where K represents the number of lesions for a given patient. The imaging dates were coded as 0 for the initial scan (baseline) and as the number of days from the baseline date for each follow-up scan. The length of the maximal diameter and its perpendicular short diameter were provided for lesions at each scan. Lesion measurement data which did not include lengths in both diameters, or which did not include the date of measurement were excluded. Additionally, non-numeric measurement values (such as "enlarged," "improved," "still abnormal but better"), measurements which were not present at baseline, or measurements with 0 length in both dimensions at baseline were also removed. Lesions in which a positive measurement was recorded in the maximal diameter while a zero was recorded for its perpendicular were recoded assigning the minimum perpendicular value in a given study in lieu of the zero value (supplementary Figure S1, available at *Annals of Oncology* online). Lesions not measured at visits post baseline were considered to have 0 length for both axes. Both Johnson and Johnson data sources provided multiple reviewers for each lesion measurement, of which one was selected at random for each patient for analysis. Given the limited number of follow-up visits provided within the Children Oncology Group data, and the considerably greater proportion of lesion measurements which were missing or of poor quality, these data are excluded from analyses involving response. In the three large industry-sponsored trials, data from randomly selected patients were provided by the sponsor. Data from FDG-PET scan results were not provided, and therefore were not part of this analysis.

For a subject let X_{ti} denote the maximal diameter for lesion i at scan time t and Y_{ti} the corresponding perpendicular dimension. The uni-dimensional measurement is $X_{t1} + \cdots + X_{tk}$ and the bi-dimensional measurement is $X_{t1}Y_{t1} + \cdots + X_{tk}Y_{tk}$, where k is the number of lesions for a given patient. The bidimensional measurement is in square of the units of the unidimensional measurement. Note that if every lesion changes by the same factor r in both dimensions (that is X value at a follow-up visit is $X(1+r)$ and likewise for Y) then the change will be r for the unidimensional measurement and $2r+r^2$ for the bidimensional measurement. However a square-root transformation of the bidimensional measurement will make the changes in the two measurements equal. It is unrealistic to expect that a lesion that

has a change in the longest diameter will have an identical change in the perpendicular. For this reason, the goal of this analysis was to evaluate how comparable the changes obtained using either unidimensional or square-root transformed bidimensional measurement are, and to what extent these changes effect response designation and time to progression of disease (PD).

In order to evaluate the application of RECIST-like response criteria to lymphoma, a comparison of the rules for response and progression assessment was conducted. Under the current Lugano response criteria for non-FDG avid lymphoma, a partial response (PR) was defined as a 50% or greater reduction in the area based on bidimensional measurements [4, 10]. When the lesion has a similar change in both dimensions, a 30% reduction in each of the maximal and perpendicular axis results in nearly a 50% reduction in area, due to the fact that each diameter is then 70% of its original size with $0.7 * 0.7 = 0.49$, or 49% of the original area. Thus, we define the equivalent reduction in unidimensional measurement as 30% (which is the threshold used by RECIST) [12]. Likewise progression in lymphoma is defined as a 50% or greater in the area from nadir. Note that a 22.5% increase in each of the maximal and perpendicular axis is needed for a 50% increase in area ($1.225 * 1.225 = 1.50$, i.e. a 50% increase in area). Thus, we define the equivalent increase in unidimensional measurement as 22.5%, slightly higher than RECIST who uses a threshold of 20% increase in diameter. With these modified threshold, 30% for PR and 22.5% for progression, we computed: the best response for each subject, the proportion of subjects in each response category, the time to response and PFS.

Since depth of response is a post treatment time-varying covariate we used a landmark analysis to determine the effect depth of response had on PFS [13]. In the landmark analysis, a fixed time point (landmark time) post baseline was selected and the depth of response was defined as the best percent change observed before the landmark time. Patients who progressed or were lost to follow-up before the landmark time are excluded from the analysis. PFS was then determined from the landmark time and the effect of depth of response was assessed. We analyzed the association by treating depth of response as a continuous variable and estimating percentiles of PFS times using smoothing techniques [14]. This method estimates the PFS percentile at a given depth of response using a weighted Kaplan–Meier estimator using data from patients whose depth of response is close to the target level. The percentiles are presented as smooth function of depth of response (percent change in tumor). As this analysis is qualitative in nature, the results are presented descriptively [14]. The stronger the association between depth of response and PFS, the steeper we would expect the percentile curves to be and hence shallow curves are indicative of minimal association.

Results

Tumor characteristics

A total of 47 828 unique measurements from 2983 individual patients enrolled on 10 multicenter clinical trials representing different lymphoma histologies, age groups (pediatrics versus adult), line of therapy, and phase of study, were collected and included in this analysis (supplementary Table S1, available at *Annals of Oncology* online) [15–23]. The number of baseline target lesions that were measured in each study, number of visits for imaging test used to perform tumor measurement, and the median and range of target lesion measurements for each clinical trial at baseline, are shown in supplementary Table S2, available at *Annals of Oncology* online.

Comparison between unidimensional and bidimensional tumor measurements

Supplementary Figure S2A, available at *Annals of Oncology* online, shows the scatterplot of the change in unidimensional (longest diameter) measurement plotted against the change in the square-root of the bidimensional measurement plotted for all subjects over all trial tumor-measurement visits. These same data are also presented in a Bland–Altman plot [24] (supplementary Figure S2B, available at *Annals of Oncology* online) with the locally weighted (LOESS) smooth fit [25]. As shown in supplementary Figure S3, available at *Annals of Oncology* online, a strong correlation of these two measurements was observed at the individual trial level, regardless of the number of lesions measured at baseline for each patient, lymphoma histology, or line of therapy (newly diagnosed versus previously treated). For the best response category, we computed the maximum decrease in the longest diameter or area before progression. Additionally, any patient who did not achieve a complete response (CR)/PR or PD within 6 months was considered as having stable disease (SD). Overall 94.5% [95% confidence interval (CI) 93.4% to 95.5%) of patients remained in the same best response category by both unidimensional and bidimensional methods. This relative change is shown as a scatterplot (supplementary Figure S4A, available at *Annals of Oncology* online) and waterfall plot (supplementary Figure S4B, available at *Annals of Oncology* online) where the best response by unidimensional and bidimensional methods is plotted in red and blue, respectively, with purple representing where the two methods overlap.

For calculation of PFS, progression was defined as the time that the lesions first exhibited a $\geq$50% increase in area or $\geq$20% increase (we also computed $\geq$22.5% increase) in diameter from the nadir value before that time-point. As shown in supplementary Figure S5, available at *Annals of Oncology* online, the PFS curve for the $\geq$50% increase in area definition is quite similar to the $\geq$22.5% increase in diameter with the $\geq$20% increase in diameter curve coming in slightly below the other two. The association between PFS and criteria for progressive disease definition was similar in patients treated with the first-line regimens and those who were treated at disease relapse (supplementary Figure S5, available at *Annals of Oncology* online)

Relationship between the depth of response and PFS

The landmark analysis which assesses the association between depth of response before the landmark time and PFS is shown in supplementary Figure S6, available at *Annals of Oncology* online. Since the previously untreated patients have different prognosis than those who were previously treated, we grouped the trials into the treated and untreated categories for this analysis.

We chose the landmark time such that most of the patients in the study have had the first follow-up scan by that time which in our case was 6 months for the previously untreated patients and 3 months for the previously treated ones. The top panels show the percentiles when progressions were determined using unidimensional measure and the bottom panels when using bidimensional measure. The percentiles were comparable across a broad range of depth of response. PFS was comparable between the analysis conducted using either the unidimensional or the bidimensional response measure.

Analysis of unidimensional measurement using the short axis and long axis

In all the preceding analyses, the unidimensional measure for a patient was defined as the sum of the long axis of the target lesions in a patient (up to the number of target lesions to be counted- see supplementary Table S2, available at *Annals of Oncology* online). The area was defined as the product of the long axis and the length of the axis perpendicular to it, also called the short axis. We repeated the analysis by defining the unidimensional measure as the sum of the short axes of the target lesions in a patient. The percent change using the short-axis unidimensional measure is highly correlated with the area measure in the pooled data set and individual trials (supplementary Table S3, available at *Annals of Oncology* online) with the least strong correlation observed being 0.893 within the PRIMA trial. The waterfall plot for the best response using the short-axis uni-dimensional measure is similar to that of the long-axis plot (supplementary Figure S7, available at *Annals of Oncology* online). The high correlation of the short- and long-axis unidimensional measures to the bidimensional area indicate that the changes in the tumor appear to be consistently occurring proportionally in both measured axes, and hence, either method of unidimensional measurement adequately captures this change.

Number of target lesions required for response assessment

We studied whether six target lesions should be included in response evaluation as recommended by the Lugano Criteria, or a smaller number could be used without loss of precision (as recommended by RECIST 1.1). To do so, we repeated the analysis with 3, 4, 5, or 6 of the largest lesions at baseline among all the recorded lesions in a given subject. The best overall response based on the N largest lesions is shown in supplementary Tables S4 and S5, available at *Annals of Oncology* online. These data show that using even as few as three target lesions allowed for 96.9% (95% CI 96.0% to 97.6%) and 97.4% (96.6% to 98.0%) of the patients to be assigned to the same response category as using six lesions, by unidimensional and bidimensional criteria, respectively. Additionally, in a comparison of best overall response category by either unidimensional or bidimensional method, the best overall response category remained identical for nearly 95% of all patients regardless of whether all lesions, the largest 6, or the largest 3 in terms of baseline diameter were used. Furthermore, the use of three target lesions produced similar depth of response as six target lesions (supplementary Figure S8, available at *Annals of Oncology* online). The percent change in lesion measurements over time was similar when up to 3, 4, 5, or 6 target lesions were used (data not shown). Combined with the concordance of uni- and bidimensional response categories, we can conclude that a valid response designation can be achieved with a limited number of lesions.

Consensus statement on staging and response evaluation

Staging and assessment of baseline tumor burden using unidimensional measurement of target lesions

One of the most important factors that determine response to therapy is related to the effect of treatment on the aggregate dimensions of all target lesions. The Lugano lymphoma response criteria currently estimate the tumor burden by using the sum of the products of the longest perpendicular diameters (SPD), which is calculated by multiplying the two longest perpendicular diameters for each target lesion. In contrast, RECIST 1.1 estimates tumor burden using sum of diameters of target lesions (longest diameter for non-nodal lesions and short axis for nodal lesions) [9, 12]. Finally, the lymphoma Lugano Criteria for non-FDG avid lymphoma calculate the baseline SPD for a maximum of six target lesions and follow them over time to determine tumor response. In contrast, RECIST 1.1 uses up to five target lesions. SPD measurements are performed by multiplying the longest perpendicular diameters for each of the target lesions. But as lymph node shape and dimensions change with therapy, these diameters can vary between observers, creating inconsistency among investigator-reported and central review responses. The use of one dimension, as required by RECIST, is easier to determine, and may enhance the reproducibility of response assessment. In fact, a recent study demonstrated that nodal tumor burden in follicular lymphoma strongly correlated with the longest diameter of the largest diseased lymph node [26].

Recommendation. Assessment of tumor burden in lymphoma clinical trials can use the sum of longest diameters (SLD). In patients with disseminated disease, a maximum of three target lesions should be selected and used to estimate tumor response. Target lesions should be selected from those with the largest size that can be reproducibly measured and preferably representing multiple sites and/organs. In most cases, lymph nodes can be considered target lesions if the lymph node longest diameter measures $\geq$15 mm. Similar to RECIST 1.1, a lymph node measuring between 10 and 14 mm is considered abnormal but should not be selected as a target lesion [12]. Lymph nodes measuring $<$10 mm in diameter are considered normal [9]. In certain anatomical sites (inguinal, axillary, and portocaval), normal lymph nodes may exist in a narrow, elongated form, and such nodes should not be selected as target lesions if alternatives are available. Extranodal lesions are selected as target lesions if they have soft tissue component, based on their size, and the ease of reproducibility of repeated measurements, with a minimum measurement of the longest diameter of $\geq$ 15 mm. All other lesions should be identified as nontarget lesions and should be recorded at baseline, without the need to measure them. Nontarget lesions should be followed and reported as present, absent, or clear progression.

Table 1. RECIL 2017: Response categories based on assessment of target lesions

	% **Change in sum of diameters of target lesions from nadir**				
	CR	**PR**	**MR**[a]	**SD**	**PD**
% change from baseline	• Complete disappearance of all target lesions and all nodes with long axis <10mm. • ≥30% decrease in the sum of longest diameters of target lesions (PR) with normalization of FDG-PET	≥30% decrease in the sum of longest diameters of target lesions but not a CR	≥10% decrease in the sum of longest diameters of target lesions but not a PR (<30%)	<10% decrease or ≤ 20% increase in the sum of longest diameters of target lesions	• >20% increase in the sum of longest diameters of target lesions • For small lymph nodes measuring <15 mm post therapy, a minimum absolute increase of 5 mm and the long diameter should exceed 15 mm • Appearance of a new lesion
FDG-PET	Normalization of FDG-PET (Deauvile score 1-3)	Positive (Deauville score 4-5)	Any	Any	Any
Bone marrow involvement	Not involved	Any	Any	Any	Any
New lesions	No	No	No	No	Yes or No

CR, complete response; CT, computerized tomography; FDG-PET, [^{18}F]2-fluoro-2-deoxy-D-glucose; MR, minor response; PD, progression of disease; PR, partial response; SD, stable disease.
[a]A provisional category.

For clinical trials, timing of pretreatment baseline scans should be based on the clinical situations. For aggressive lymphoma, such as diffuse large B-cell lymphoma, scans within 4 weeks would be appropriate, but for indolent disease baseline scans may be within longer window that should be defined in the study. Whenever possible, the same imaging modality should be used at baseline and subsequent visits. At the present time, CT scan imaging, preferably with oral and intravenous contrasts remains the gold standard for determining tumor measurements before, during, and after completion of therapy. In certain situations where minimization of exposure to ionizing radiation is desirable, or where CT provides suboptimal assessment (such as primary bone lymphoma), standard magnetic resonance imaging can be used to determine baseline and subsequent tumor measurements. FDG-PET should be included in the initial staging work up all FDG-PET avid lymphomas [4, 27]. In certain cases, measurements may be performed on the CT-component of a combined PET/CT images, provided this is of adequate resolution.

In patients with newly diagnosed lymphoma, a bone marrow biopsy should be performed at baseline to determine the stage of disease. A baseline bone marrow biopsy is mandatory for previously untreated patients with indolent B-cell lymphoma, mantle cell lymphoma, and T-cell lymphoma. Patients with diffuse large B-cell lymphoma with a negative FDG-PET uptake in the bone marrow, does not rule out bone marrow involvement, especially discordant histology [28, 29]. However, a positive FDG-PET uptake in the bone marrow may obviate the need for a bone marrow biopsy [29]. Patients with Hodgkin lymphoma without FDG uptake in the bone marrow or presence of B-symptoms do not need a bone marrow biopsy at baseline, as bone marrow biopsy in this situation is extremely unlikely to modify stage [30]. Bone marrow biopsy may still be indicated to address special clinical scenarios, such as evaluation for stem cell collection, and to rule out possible myelodysplasia in patients with prolonged cytopenia.

Response assessment

Complete response. In the current Lugano Criteria, CR is defined as complete normalization of FDG-PET uptake (Deauville score of 1 to 3) or complete resolution of all target lesions for non-FDG avid lymphoma or when PET cannot be performed [27]. PET results are also used to discriminate CR from a prior criterion termed CR unconfirmed (CRu). End of treatment PET usually refers to predefined number of chemotherapy-based regimens that are typically administered for six to eight cycles. Several new agents have been reported to modulate tumor metabolism, glucose uptake, and inflammation in the tumor microenvironment, and therefore may potentially increase the false-positive or false-negative FDG-PET results [31].

Recommendation: CR is defined as a complete resolution of all target lesions by CT scans with complete normalization of FDG-PET uptake in all areas (Deauville score of 1–3), and bone marrow biopsy negativity (if it was positive or unknown at baseline). If pretreatment PET scan was negative, lymph nodes that measured ≥15 mm in the long axis should regress to < 10 mm. CR is also defined as achievement of a partial remission by CT scan criteria (reduction in sum of longest diameters by CT imaging by >30%) with normalization (Deauville score 1–3) of FDG-PET activity in FDG-avid lymphoma (Table 1). Because many novel

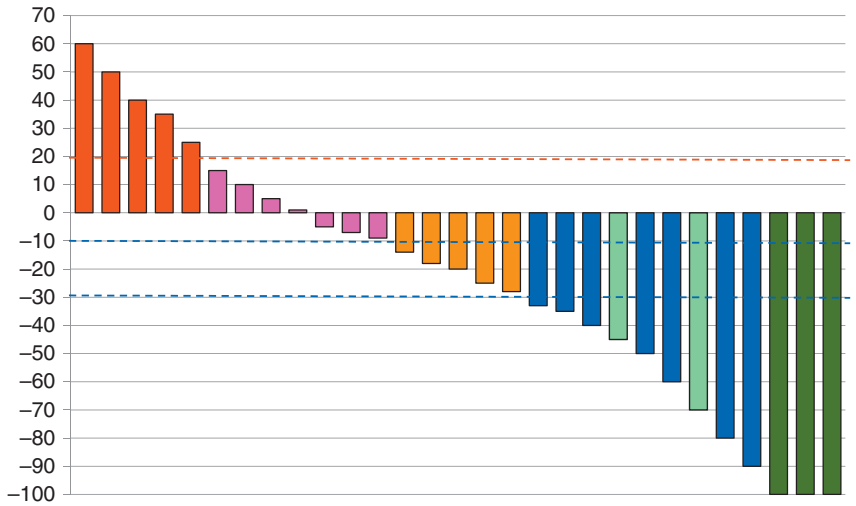

Figure 1. Treatment outcome by response category using a waterfall plot, Responses are color coded based on the cutoffs shown in Table 1. Red, progression of disease; pink, stable disease; orange, minor response; blue, partial response; dark green, complete response. Light green bars denote complete response based on integrating PET results. The horizontal dotted lines show the boundaries for partial response, minor response, stable disease, and progression of disease (Table 1).

Table 2. Calculating sum of diameters to include small responsive lymph nodes

Target lesions	Baseline measurement (long axis; cm)	Nadir actual measurement (cm) method 1	Nadir normalized measurement (cm) method 2
Lesion 1	1.6	0.9	0 (resolved)
Lesion 2	1.7	1.4	1.4
Lesion 3	2	1.8	1.8
Sum of diameters	5.3	4.1	3.2
% change from baseline	N/a	23	40
Response designation	N/A	Minor response	Partial response (or CR if PET is negative)

CR, complete response; PET, positron emission tomography.

targeted agents may alter glucose uptake and/or metabolism, normalizing of FDG-PET imaging alone is not sufficient by itself to determine CR status unless accompanied with a significant (>30%) decrease in the sum of diameters. Accordingly, a reduction in the sum of diameters by ≤30% with normalization of FDG-PET uptake should not be considered a CR unless documented by a negative tissue biopsy. PET-based CRs should be identified by specific designation on waterfall plots (Figure 1).

In cases where pretreatment baseline tumor burden is low, with only a few lesions measuring around 2 cm in longest diameter, treatment effect may shrink the long axis of a target lymph node to a normal values of <10 mm. However, even though the lymph node is now within normal size range, consistent with CR, the percentage of diameter reduction may be <30% (less than a PR). In these cases, a normalized diameter of "0, or resolved" (Table 2, method 2) should be used to calculate the sum of diameters, and therefore ensuring accurate response designation. Accordingly, a "normalized" calculation should be used when creating a waterfall plot.

Partial response. In the current Lugano lymphoma Response Criteria, PR is defined as a decrease in the SPD of target lesions by ≥ 50%, with no increase in the size of any lesion, and no appearance of new lesions. Typically, one or more lesions are also PET avid. A scenario is often encountered whereby the size of one or more lesions is increased by ≥ 50%, even though the overall SPD is decreased by more than 50% from baseline, This "mixed response" is designated as PD in the current Lugano Criteria [4]. In contrast, RECIST designation of a response is based on the overall changes in the sum of diameters, irrespective of a mixed response. Because many phase-I studies include patients with solid tumors and lymphoma, the discrepancy between the Lugano Criteria and RECIST creates regulatory concerns of how these responses should be reported.

Recommendation: Consistent with RECIST, PR is defined as a reduction of the sum of longest diameters of target lesions by ≥30%, but without meeting the definition of CR described above (Table 1; Figure 1). If one or more target lesions grew in size but the sum of the diameters remains ≤30% of the baseline measurement, and no new lesions appear, the response should be designated PR. This revised definition will eliminate the false interpretation of disease progression due to treatment-related inflammatory flares that were recently reported with some new

agents (see "Special Cases" section) and will eliminate early termination of potentially beneficial therapy in an otherwise noncurable clinical setting.

Minor response. Neither the Lugano Criteria nor RECIST include the designation of minor response (MR). In the Lugano Criteria, the designation of SD includes changes in the SPD, ranging between -49% and +50% (for RECIST, the range is −29% to +20% of the sum of diameters). Lumping such a wide range of changes in tumor measurements in one response category is clinically uninformative, as different changes may require different therapeutic interventions. Furthermore, with the use of modern imaging methods, precise changes in tumor measurements are frequently reported in waterfall plots. Many new agents have been shown to reduce tumor measurements relative to baseline, but not fulfilling the criteria for a PR. In patients who did not achieve a CR, the depth of response did not correlate with PFS (see supplementary Methods and Results available at *Annals of Oncology* online). This is not surprising, since the definition of PR was based on a historically arbitrary cutoff in tumor reduction. In fact, recent analysis of several phase II studies of single agents in relapsed lymphoma suggested that patients with PR had a similar PFS compared with those who had some tumor reduction that was below the cutoff of a PR [32–34].

Recommendations: A new provisional category of MR should be included in the response assessment. Using single diameter long-axis measurements, MR is defined as a reduction in the SLD of target lesions by ≥10% but <30%, without the appearance of any new lesions, and irrespective of PET scan results (Table 1; Figure 1). This cutoff was conservatively chosen, to eliminate a potential margin of measurement error, and therefore, it is considered a provisional category. A mixed response will be called a MR, as long as the sum of longest diameters is consistent with a MR.

Stable disease. In many cases of chronically administered new agents, including immune therapies, the best response to treatment may only be achieved after prolonged administration of therapy [35, 36]. Accordingly, an initial designation of SD should not be a basis for premature termination of therapy, especially if treatment is well tolerated, as the quality of response may improve with time to become MR, PR, or CR. In addition, the patient may derive benefit from a given therapy even if a response is not achieved.

Recommendation: SD is defined as changes in the SLD of targeted lesions ranging between reduction of <10% to an increase by ≤20% without the appearance of a new lesion, and irrespective of PET results (Table 1; Figure 1). Mixed responses will be called SD as long they fulfill the above criteria. This definition is used in lymphoma-specific clinical trials when an MR is included in the study aims. In clinical trials that include both lymphoma and solid tumors, whereas RECIST does not include a MR category, the definition of SD should remain similar to RECIST (−29% to +20%)

Progression of disease. The definition of PD should be consistent with loss of benefit of therapy, requiring stopping or changing treatment. The current Lugano Criteria defines PD as having any of the following categories: (i) appearance of any new lesion of >1.5 cm in longest diameter, (ii) an increase of at least 50% from nadir in the SPD of any previously involved lymph nodes, (iii) at least a 50% increase in the longest diameter of any single previously identified node >1 cm in its long axis. This definition is in conflict with clinical practice as it calls for stopping or changing therapy when a single lymph node increases in size from 1.0 to 1.6 cm, even though the overall SPD may have shown significant reduction. Moreover, as several novel agents have been shown to induce a local immune response or a "flare," an increase in lymph node sizes temporally corresponding with the initial cycles of administration of therapy may not necessarily be due to progression of the disease [36, 37].

Recommendations: Consistent with RECIST 1.1, and using a unidimensional tumor measurement, PD after initiating a new therapy is defined as an increase in the sum of longest diameters of target lesions by >20%, and/or appearance of a new lesion (lymph node ≥ or a soft tissue mass ≥10 mm of the longest diameter), irrespective of FDG-PET results (Table 1). Whenever possible, questionable small FDG-PET avid lesions should be confirmed by a histologic or cytologic analysis. Appearance of a new FDG-PET avid lesion that is smaller than the above thresholds should be closely monitored, and whenever possible, a biopsy should be performed to determine its nature. An increase in the size of previously involved small lymph nodes by >20% while other lesions are decreasing, especially at the beginning of treatment with investigational agents, may represent a tumor flare and should not be designated a PD, unless there is continued increase in size on subsequent imaging studies. Patients should be allowed to remain on trial at investigators and patient discretion until the response or lack thereof is clarified on subsequent imaging.

Progression after an initial response. The current Lugano response criteria use nadir tumor measurements as the new baseline for defining PD, and therefore, even a small regrowth of one or more lymph nodes is defined as PD, requiring stopping or changing therapy. This definition may paradoxically result in shorter PFS times in patients who achieve the best response. Unlike patients with relapsed and refractory solid tumors, patients with lymphoma frequently achieve CRs and very good PRs, where the nadir tumor measurement can be very low compared with baseline. To avoid premature termination of therapy due to minor fluctuations or a small increase in tumor measurements from the nadir, especially in patients with no available curative options, patients may be allowed to continue receiving therapy beyond the strict definition of PD as long as (i) the patient does not have prohibitive toxicity and (ii) the patient remains free from significant disease-related symptoms. By doing so, the time for changing therapy or discontinuation of therapy can be a more useful measure of determining treatment success compared with PFS. However, if used, this end point should be prospectively and clearly stated in the objectives of clinical trials.

Recommendation: After an initial response, and in the absence of appearance of new lesions, PD is defined as an increase of the nadir sum of diameters by >20%. Consistent with RECIST 1.1, patients who achieve a CR (normalization of all lymph node

measurements and disappearance of extranodal lesions), at least one previously involved lymph node should increase in size to measure ≥15 mm in the long diameter, with a minimum absolute increase of at least 5 mm from nadir [9, 12] Accordingly, an increase in a lymph node longest diameter from 8 to 13 mm is not considered a PD, even though there is 38% increase in the measurement, since the lesion did not exceed 15 mm. Similarly, a change from 12 to 16 mm does not qualify as a PD even though the new measurement exceeds 15 mm, since the absolute increase was <5 mm. In the absence of alternative treatment options, lymphoma-related symptoms, and no new lesions, treatment may continue beyond PD with periodic follow up imaging studies, to prolong patient's clinical benefit. Such a plan, which should be prospectively included in the study design and should not redefine of PFS, may allow capturing data to calculate time to next therapy or time to discontinuation of therapy. If used in clinical trials, this aim should be prospectively described in the study design and in the consent form.

Time to progression, PFS, event-free survival, and overall survival

Time to progression is defined as the time from study entry until disease progression. PFS is defined as the time from start of study entry until disease progression or death. Event-free survival (EFS) should be reserved to define specific events that are intended to be prevented or delayed by therapy. Events are prespecified for each study and may include implementing a change of therapy, disease progression, disease relapse, second malignant neoplasms, and death of any cause. EFS is measured from the time from study entry to the event. Overall survival is defined as the time from study entry or initial diagnosis until death from any cause.

Response assessment in patients receiving immune modulating agents, including immune checkpoint inhibitors

Immunomodulating agents, such as lenalidomide, and new immunotherapies, such as immune check point inhibitors, in addition to cell therapy with chimeric antigen receptor engineered T cells can be associated with a "pseudo-progression" that may be related to recruitment of immune cells to disease site [36, 38–46]. After initial recruitment of activated T cells, the tumor lesion may transiently increase in size before shrinking. To avoid premature termination of such therapies, an immune-related response criteria were developed which required confirmation of PD on two consecutive scans at least 4 weeks apart, and inclusion of new lesion measurements to the total tumor burden [47–50]. These criteria are distinct from RECIST and the Lugano Criteria, which define PD at tumor burden increase above the specified threshold (20% for RECIST and 50% for Lugano Criteria) or at the appearance of new lesions, without the need for confirmation on subsequent imaging. When serial imaging studies confirms that the prior increase in tumor measurements was related to an early manifestation of disease progression rather than a tumor flare, the time of progression should be back-dated to the initial scans that documented PD.

Response assessment in patients receiving agents that mobilize lymphoma cells from lymph nodes and bone marrow into the blood

Some agents can inhibit adhesion mechanisms in tumor cells causing redistribution of tumor cells from lymph nodes and/or bone marrow into the blood. Thus, while lymph node size decrease in response to therapy, the tumor cell count increases in the blood, creating another form of "pseudo-progression". With continued therapy, blood lymphocytosis decrease as tumor cells start to die. This phenomenon has been observed with BTK and PI3K inhibitors [51–58]. Accordingly, increased lymphocytosis in the setting of a decrease in lymph node measurement is not considered PD, and response designation should depend on lymph nodes and extra-nodal disease measurement. Lymphocytosis can be included as annotation. For example, PR with increased lymphocytosis.

Appearance of a new extranodal lesion

With the use of PET imaging, a new small PET avid lesion may appear during or after therapy, but there are no guidelines on how to interpret such lesions to define PD. Ideally, such lesions should be biopsied where clinically feasible to clarify their nature, but frequently they are too small to biopsy. Eventually, with observation, the nature of such lesions is clarified. Pulmonary or skin infection and arthritis may result in false positive PET activity and may be confused with an early progression. A retrospective approach may not be appropriate as it creates confusions as patients' records and source documents will require corrections. On the other hand, a delayed declaration of PD may also create concerns of data integrity for regulatory oversight of clinical trials, especially when a drug is undergoing an approval process by a regulatory authority. Therefore, a uniform and transparent approach should be implemented.

Recommendation. A minimum of 1 cm in largest diameter of new extra-nodal lesions is required to assign PD directly. New smaller but suspicious lesion should be designated as equivocal, and if later confirmed (by CT or biopsy) as being due to lymphoma, the documented date of progression should be the date of when it was first identified as equivocal

Integrating target and nontarget lesions in the response assessment

In case of disseminated disease, the status of non-target lesions should be taken into account before formulating the final response status. A recommended approach is shown in Table 3.

Measurement of splitting lesions

Frequently, effective therapy may convert a large confluent mass to several smaller constituent lymph nodes (Figure 2). In this case, and consistent with RECIST, the measurement of each lymph node should be carried out and entered in the calculation of sum of diameters. However, to avoid an overall increase in the number of target lesions, sub-designations of A, B, C, etc. for any target lesion that has undergone splitting should be created.

Table 3. Response designation incorporating best response of target lesions (Table 1) and nontarget lesions

Target lesion	Nontarget lesion	New lesion	Response designation
CR	CR	No	CR
CR	PR, MR, or SD	No	PR[a]
CR	UE	No	UE
PR	UE	No	UE
PR	CR	No	PR
PR	PR, MR, or SD	No	PR
MR	UE	No	UE
MR	CR	No	MR
MR	PR, MR, or SD	No	MR
SD	UE	No	UE
SD	CR, PR, or MR	No	SD
SD	SD	No	SD
PD	Any	Yes/no	PD
Any	PD	Yes/no	PD
Any	Any	Yes	PD
CR	No	No	CR
PR	No	No	PR
MR	No	No	MR
SD	No	No	SD

CR, complete response; MR, minor response; PD, progression of disease; PR, partial response; SD, stable disease; UE, unevaluable.

[a]As in Table 6, computerized tomography scan-based PR with complete normalization of [^{18}F]2-fluoro-2-deoxy-D-glucose positron emission tomography activity is considered CR.

Spleen measurement

Spleen size may vary considerably in size and shape in healthy individuals and typically is not selected as a target lesion. In the Lugano Criteria, splenomegaly is defined as greater than 13 cm in "vertical length." This definition might be confusing for some radiologists since it does not specify the plane that is commonly used by radiologist. To clarify this recommendation, vertical length measurement should be carried out in the coronal image as shown in Figure 3. Alternatively, the spleen vertical length can be calculated by multiplying the number of spleen slices in transverse CT views by the thickness of each slice, or by measuring splenic coronal diameter on a PET maximum intensity projection image

Reporting response results in waterfall plots

Actual percentage changes in sum of diameters should be reported, with special consideration of the definition of CR (Figure 1). If PET status is used to designate a CR (i.e. PR with negative PET), then these patients should be identified by a separate color of the bars, or by an asterisk in a waterfall plot. In calculating sum of diameters, target lymph node lesions that decrease in size to <1.0 cm and became PET negative, may be recorded as resolved or "0" (Table 2). Splitting lesions should be included in the sum of diameters measurement when reporting results in waterfall plots.

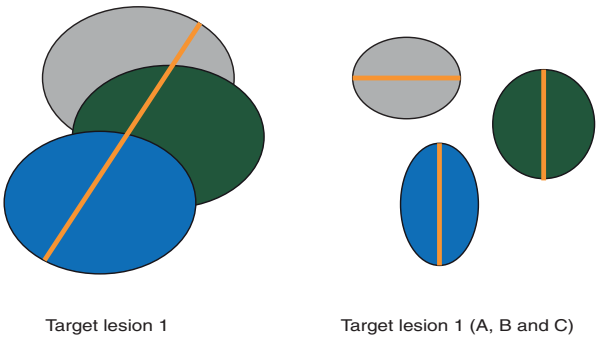

Figure 2. Measurement of a splitting lesion in response to therapy.

Frequency of response assessment

Several investigational agents have been shown to be safely administered for a prolonged period of time requiring several repeated imaging studies, raising concerns about a potential increased risk of radiation exposure. While an accurate assessment of PFS is critical during the first 6 months of initiating therapy with investigational agents, less frequent imaging assessments are reasonable with ongoing prolonged treatment. A uniform incorporation of surveillance intervals in clinical studies will be important to allow comparison of PFS response assessment across different trials [59].

Recommendation. In phase I/II clinical trials in previously treated patients, it is recommended that response assessment be carried out every 2–3 months during the first year of therapy. In the absence of new symptoms or clinical concerns, subsequent imaging studies can be carried out every 3–4 months during the second year, and every 6 months thereafter, for the duration specified in the clinical trial. Imaging assessment may be carried out less frequently during and after therapy of newly diagnosed patients, and in the settings of randomized phase III studies. In some countries, local health authorities and ethics committees may demand longer imaging intervals for response assessments.

Conclusions and future directions

The proposed new RECIL criteria are aligned with RECIST, and are applicable for both adult and pediatric patients with lymphoma (Table 4). While most of our recommendations are supported by large data analysis, some remain based on consensus recommendations, including the requirement for tissue biopsy to confirm CR of PET negative disease in the setting of minor reduction in tumor measurement, the proposed minimum increase in the size of lymph nodes to qualify as PD, the optimal method of evaluating splenomegaly, the optimal intervals of imaging studies to monitor response to therapy, and the optimal staging categories that may better predict treatment outcome. Furthermore, the proposed new category of MR will need to be prospectively validated to determine its usefulness in guiding clinical practice and clinical research.

With >60 lymphoma histologic subtypes, and many different clinical presentations related to organ site involvement and bulk

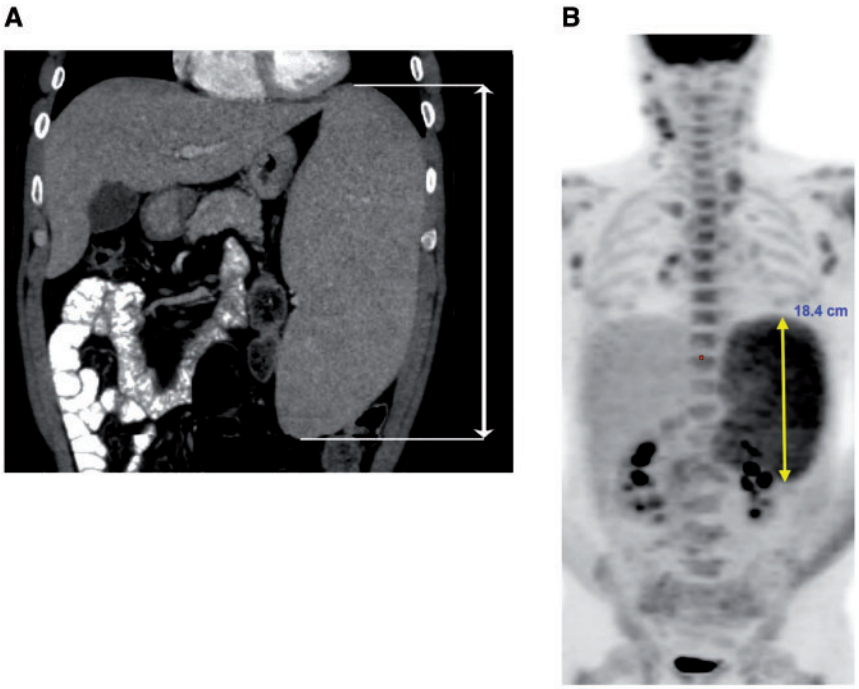

Figure 3. Recommendation for measuring spleen long diameter. (A) Coronal view of a computerized tomography (CT) scan image, (B) maximum intensity projection image of a positron emission tomography/CT.

Table 4. Comparison between RECIST 1.1, Lugano lymphoma classification, and RECIL 2017

	RECIST 1.1	Lugano	RECIL 2017
Number of target lesions	Up to 5	Up to 6	Up to 3
Measurement method	Uni-dimensional: long diameter of non-nodal lesions, short diameter of lymph nodes	Bi-dimensional: perpendicular diameters	Uni-dimensional: long diameter of any target lesion
Incorporates PET results to describe CR	May be considered to confirm CR and/or to declare PD based on detecting new lesions	Yes	Yes
Minor response	No	No	Yes, reduction in sum of long diameters between $\geq$10% and <30%
Stable disease	−29% to + 20%	−50% to + 50%	decrease <10% to increase $\leq$20%
PD	Increase in sum of diameters by 20%	Increase in the sum of products of perpendicular diameters by > 50%, or any single lesion by > 50%	Increase in sum of the longest diameters by 20%. For relapse from CR, at least one lesion should measure 2 cm in the long axis with or without PET activity

CR, complete response; PD, progression of disease; PET, positron emission tomography.

of the disease, it is impossible for a response criteria to cover all possible scenarios. Accordingly, it is possible that certain clinical scenarios may require a slight modification of the proposed RECIL 2017. For example, although our proposed criteria is applicable for bulky and non-bulky target lesions, in certain clinical presentations such as primary mediastinal diffuse large B-cell lymphoma, and classical Hodgkin lymphoma with bulky mediastinal disease, the definition of CR may require modification. In some patients with these two lymphoma types, a CR is typically defined by a PET negative status at the end of therapy, regardless of the percentage of tumor size reduction by CT scan. However, the core principle of the RECIL 2017, including using the sum of longest diameters and the inclusion of up to three target lesions should remain the same.

Future directions should evaluate the role of molecular depth of response (minimal residual disease and circulating DNA) in predicting treatment outcome, and to guide future studies aimed at evaluating shorter duration of therapy. Our new RECIL

proposal should also be evaluated by the RECIST investigators to consider further harmonization of the criteria, with a special attention to the measurement of the long axis of lymph nodes, optimal number of target lesions, and introduction of MR category.

Acknowledgements

The authors would like to thank Debra Friedman, MD (Study Chair of COG AHOD0031) and Kathleen McCarten, MD (COG AHOD0031 Radiologist who measured all of the individual lesions for the COG trial). The authors thank the European Organization for Research and Treatment of Cancer for permission to use the data from EORTC trials 20921, 20981 and 20021 for this research, and Martin Barrett and Claude Berge of ROCHE/Genentech for early statistical discussions of the project and providing data.

Funding

Supported in part by the MSK SPORE in lymphoma P50 CA192937-01A1 and MSK Cancer Center Support Grant P30CA008748 (AY, ADZ, and VES), and P50 CA97274 University of Iowa/Mayo Lymphoma SPORE (TW and SA), and a Leukemia and Lymphoma SCOR-9483-16 grant (AY).

Disclosure

The authors have declared no conflicts of interest.

References

1. Cheson BD. The International Harmonization Project for response criteria in lymphoma clinical trials. Hematol Oncol Clin North Am 2007; 21(5): 841–854.
2. 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.
3. 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(5): 571–578.
4. 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.
5. Younes A, Ansell S, Fowler N et al. The landscape of new drugs in lymphoma. Nat Rev Clin Oncol 2016. Dec 29 [Epub ahead of print], doi: 10.1038/nrclinonc.2016.205.
6. Younes A, Hagenbeek A, Coiffier B. Optimising the lymphoma response criteria in the era of targeted therapy. Lancet Oncol 2011; 12(7): 616–617.
7. Oxnard GR, Morris MJ, Hodi FS et al. When progressive disease does not mean treatment failure: reconsidering the criteria for progression. J Natl Cancer Inst 2012; 104(20): 1534–1541.
8. Fournier L, Ammari S, Thiam R, Cuenod CA. Imaging criteria for assessing tumour response: RECIST, mRECIST, Cheson. Diagn Interv Imaging 2014; 95(7-8): 689–703.
9. Eisenhauer EA, Therasse P, Bogaerts J et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009; 45(2): 228–247.
10. Cheson BD, Pfistner B, Juweid ME et al. Revised response criteria for malignant lymphoma. J Clin Oncol 2007; 25(5): 579–586.
11. Assouline S, Meyer RM, Infante-Rivard C et al. Development of adapted RECIST criteria to assess response in lymphoma and their comparison to the International Workshop Criteria. Leuk Lymphoma 2007; 48(3): 513–520.
12. Schwartz LH, Bogaerts J, Ford R et al. Evaluation of lymph nodes with RECIST 1.1. Eur J Cancer 2009; 45(2): 261–267.
13. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response and other comparisons of time-to-event by outcome variables. J Clin Oncol 2008; 26(24): 3913–3915.
14. Bowman AW, Azzalini A, Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations. London: Oxford University Press 1997.
15. Friedman DL, Chen L, Wolden S et al. Dose-intensive response-based chemotherapy and radiation therapy for children and adolescents with newly diagnosed intermediate-risk hodgkin lymphoma: a report from the Children's Oncology Group Study AHOD0031. J Clin Oncol 2014; 32(32): 3651–3658.
16. Aurer I, Eghbali H, Raemaekers J et al. Gem-(R)CHOP versus (R)CHOP: a randomized phase II study of gemcitabine combined with (R)CHOP in untreated aggressive non-Hodgkin's lymphoma–EORTC lymphoma group protocol 20021 (EudraCT number 2004-004635-54). Eur J Haematol 2011; 86(2): 111–116.
17. Hagenbeek A, Eghbali H, Monfardini S et al. Phase III intergroup study of fludarabine phosphate compared with cyclophosphamide, vincristine, and prednisone chemotherapy in newly diagnosed patients with stage III and IV low-grade malignant Non-Hodgkin's lymphoma. J Clin Oncol 2006; 24(10): 1590–1596.
18. Offner F, Samoilova O, Osmanov E et al. Frontline rituximab, cyclophosphamide, doxorubicin, and prednisone with bortezomib (VR-CAP) or vincristine (R-CHOP) for non-GCB DLBCL. Blood 2015; 126(16): 1893–1901.
19. Salles G, Seymour JF, Offner F et al. Rituximab maintenance for 2 years in patients with high tumour burden follicular lymphoma responding to rituximab plus chemotherapy (PRIMA): a phase 3, randomised controlled trial. Lancet 2011; 377(9759): 42–51.
20. van Oers MH, Van Glabbeke M, Giurgea L et al. Rituximab maintenance treatment of relapsed/resistant follicular non-Hodgkin's lymphoma: long-term outcome of the EORTC 20981 phase III randomized intergroup study. J Clin Oncol 2010; 28(17): 2853–2858.
21. Evens AM, Balasubramanian S, Vose JM et al. A Phase I/II Multicenter, Open-Label Study of the Oral Histone Deacetylase Inhibitor Abexinostat in Relapsed/Refractory Lymphoma. Clin Cancer Res 2016; 22(5): 1059–1066.
22. Gopal AK, Kahl BS, de Vos S et al. PI3Kdelta inhibition by idelalisib in patients with relapsed indolent lymphoma. N Engl J Med 2014; 370(11): 1008–1018.
23. Coiffier B, Osmanov EA, Hong X et al. Bortezomib plus rituximab versus rituximab alone in patients with relapsed, rituximab-naive or rituximab-sensitive, follicular lymphoma: a randomised phase 3 trial. Lancet Oncol 2011; 12(8): 773–784.
24. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476): 307–310.
25. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc.1979; 74(368): 829–836.
26. Federico M, Bellei M, Marcheselli L et al. Follicular lymphoma international prognostic index 2: a new prognostic index for follicular lymphoma developed by the international follicular lymphoma prognostic factor project. J Clin Oncol 2009; 27(27): 4555–4562.
27. 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(27): 3048–3058.
28. Pelosi E, Penna D, Douroukas A et al. Bone marrow disease detection with FDG-PET/CT and bone marrow biopsy during the staging of malignant lymphoma: results from a large multicentre study. Q J Nucl Med Mol Imaging 2011; 55(4): 469–475.

29. Adams HJ, Kwee TC, de Keizer B et al. FDG PET/CT for the detection of bone marrow involvement in diffuse large B-cell lymphoma: systematic review and meta-analysis. Eur J Nucl Med Mol Imaging 2014; 41(3): 565–574.
30. El-Galaly TC, d'Amore F, Mylam KJ et al. Routine bone marrow biopsy has little or no therapeutic consequence for positron emission tomography/computed tomography-staged treatment-naive patients with Hodgkin lymphoma. J Clin Oncol 2012; 30(36): 4508–4514.
31. Meignan M, Itti E, Gallamini A, Younes A. FDG PET/CT imaging as a biomarker in lymphoma. Eur J Nucl Med Mol Imaging 2015; 42(4): 623–633.
32. Younes A, Sureda A, Ben-Yehuda D et al. Panobinostat in Patients With Relapsed/Refractory Hodgkin's Lymphoma After Autologous Stem-Cell Transplantation: Results of a Phase II Study. J Clin Oncol 2012 Apr 30 [Epub ahead of print].
33. Younes A, Gopal AK, Smith SE et al. Results of a pivotal phase II study of brentuximab vedotin for patients with relapsed or refractory Hodgkin's lymphoma. J Clin Oncol 2012; 30(18): 2183–2189.
34. Younes A, Oki Y, Bociek RG et al. Mocetinostat for relapsed classical Hodgkin's lymphoma: an open-label, single-arm, phase 2 trial. Lancet Oncol 2011; 12(13): 1222–1228.
35. Advani RH, Buggy JJ, Sharman JP et al. Bruton tyrosine kinase inhibitor ibrutinib (PCI-32765) has significant activity in patients with relapsed/refractory B-cell malignancies. J Clin Oncol 2013; 31(1): 88–94.
36. Witzig TE, Wiernik PH, Moore T et al. Lenalidomide oral monotherapy produces durable responses in relapsed or refractory indolent non-Hodgkin's lymphoma. J Clin Oncol 2009; 27(32): 5404–5409.
37. Chanan-Khan A, Miller KC, Lawrence D et al. Tumor flare reaction associated with lenalidomide treatment in patients with chronic lymphocytic leukemia predicts clinical response. Cancer.117(10): 2127–2135.
38. Armand P, Shipp MA, Ribrag V et al. Programmed death-1 blockade with pembrolizumab in patients with classical Hodgkin lymphoma after brentuximab vedotin failure. J Clin Oncol 2016; 34: 3733–3739.
39. Zinzani PL, Ribrag V, Moskowitz CH et al. Phase 1b Study of PD-1 Blockade with Pembrolizumab in Patients with Relapsed/Refractory Primary Mediastinal Large B-Cell Lymphoma (PMBCL). Blood 2015; 126(23): 3986.
40. Younes A, Santoro A, Shipp M et al. Nivolumab for classical Hodgkin's lymphoma after failure of both autologous stem-cell transplantation and brentuximab vedotin: a multicentre, multicohort, single-arm phase 2 trial. Lancet Oncol 2016; 17(9): 1283–1294.
41. Goy A, Sinha R, Williams ME et al. Single-agent lenalidomide in patients with mantle-cell lymphoma who relapsed or progressed after or were refractory to bortezomib: phase II MCL-001 (EMERGE) study. J Clin Oncol 2013; 31(29): 3688–3695.
42. Witzig TE, Vose JM, Zinzani PL et al. An international phase II trial of single-agent lenalidomide for relapsed or refractory aggressive B-cell non-Hodgkin's lymphoma. Ann Oncol 2011; 22(7): 1622–1627.
43. Kochenderfer JN, Dudley ME, Kassim SH et al. Chemotherapy-refractory diffuse large B-cell lymphoma and indolent B-cell malignancies can be effectively treated with autologous T cells expressing an anti-CD19 chimeric antigen receptor. J Clin Oncol 2015; 33(6): 540–549.
44. Park JH, Brentjens RJ. Adoptive immunotherapy for B-cell malignancies with autologous chimeric antigen receptor modified tumor targeted T cells. Discov Med 2010; 9(47): 277–288.
45. Turtle CJ, Hanafi LA, Berger C et al. Immunotherapy of non-Hodgkin's lymphoma with a defined ratio of CD8+ and CD4+ CD19-specific chimeric antigen receptor-modified T cells. Sci Transl Med 2016; 8(355): 355ra116.
46. Bollard CM, Gottschalk S, Torrano V et al. Sustained complete responses in patients with lymphoma receiving autologous cytotoxic T lymphocytes targeting Epstein-Barr virus latent membrane proteins. J Clin Oncol 2014; 32(8): 798–808.
47. Hodi FS, Hwu WJ, Kefford R et al. Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with pembrolizumab. J Clin Oncol 2016; 34(13): 1510–1517.
48. Hoos A, Wolchok JD, Humphrey RW, Hodi FS. CCR 20th anniversary commentary: immune-related response criteria–capturing clinical activity in immuno-oncology. Clin Cancer Res 2015; 21(22): 4989–4991.
49. Wolchok JD, Hoos A, O'Day S et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. Clin Cancer Res 2009; 15(23): 7412–7420.
50. Cheson BD, Ansell S, Schwartz L et al. Refinement of the Lugano classification response criteria for lymphoma in the era of immunomodulatory therapy. Blood 2016; 128: 2489–2496.
51. Byrd JC, Brown JR, O'Brien S et al. Ibrutinib versus ofatumumab in previously treated chronic lymphoid leukemia. N Engl J Med 2014; 371(3): 213–223.
52. Dreyling M, Jurczak W, Jerkeman M et al. Ibrutinib versus temsirolimus in patients with relapsed or refractory mantle-cell lymphoma: an international, randomised, open-label, phase 3 study. Lancet 2016; 387(10020): 770–778.
53. Furtado M, Wang ML, Munneke B et al. Ibrutinib-associated lymphocytosis corresponds to bone marrow involvement in mantle cell lymphoma. Br J Haematol 2015; 170(1): 131–134.
54. Herman SE, Niemann CU, Farooqui M et al. Ibrutinib-induced lymphocytosis in patients with chronic lymphocytic leukemia: correlative analyses from a phase II study. Leukemia 2014; 28(11): 2188–2196.
55. Rossi D, Gaidano G. Lymphocytosis and ibrutinib treatment of CLL. Blood 2014; 123(12): 1772–1774.
56. Wang ML, Rule S, Martin P et al. Targeting BTK with ibrutinib in relapsed or refractory mantle-cell lymphoma. N Engl J Med 2013; 369(6): 507–516.
57. Brown JR, Byrd JC, Coutre SE et al. Idelalisib, an inhibitor of phosphatidylinositol 3-kinase p110delta, for relapsed/refractory chronic lymphocytic leukemia. Blood 2014; 123(22): 3390–3397.
58. Fiorcari S, Brown WS, McIntyre BW et al. The PI3-kinase delta inhibitor idelalisib (GS-1101) targets integrin-mediated adhesion of chronic lymphocytic leukemia (CLL) cell to endothelial and marrow stromal cells. PLoS One 2013; 8(12): e83830.
59. Panageas KS, Ben-Porat L, Dickler MN et al. When you look matters: the effect of assessment schedule on progression-free survival. J Natl Cancer Inst 2007; 99(6): 428–432.