July 2017

Tumor Response Evaluation Criteria: Standardization Ensures Success

Maseeh Uz Zaman  
Aga Khan University, maseeh.uzzaman@aku.edu

Nosheen Fatima  
Aga Khan University, nosheen.fatima@aku.edu

Mahwish Sajid

Follow this and additional works at: https://ecommons.aku.edu/pakistan_fhs_mc_radiol

Part of the Radiology Commons

Recommended Citation
Available at: https://ecommons.aku.edu/pakistan_fhs_mc_radiol/275
EDITORIAL

Tumor Response Evaluation Criteria: Standardization Ensures Success

Maseeh uz Zaman¹, Nosheen Fatima¹ and Mahwish Sajid²

Cancer-related mortality is growing rapidly. It is expected that in years to come, cancer will become the most common cause of death worldwide. Cancer-related therapies are not only expensive but associated with morbidities; and the tumor shrinkage with recently introduced therapies is seen in only 10-15% patients.¹ However, in recent days various therapies have improved survival rate and outcome of patients with certain cancers like childhood malignancy and lymphomas. Imaging plays an important role by helping the oncologists and guides them whether to continue, change or abandon a treatment, depending upon the response of the tumor. Response could be evaluated by physical examination or by using anatomical or functional imaging, and it could be assessed either qualitatively like the visual analysis or quantitatively using digital images and computer algorithms.

In early 70's when computerized tomography (CT) was not freely available, Moertel and Henley² presented a study in which they employed 16 experienced oncologists and assessed the consistency in palpation of 12 pairs of fake masses. For 50% reduction in perpendicular diameter of tumor as response, the inter-observer variability was 7-8%. However, for a cut-off of 25% reduction in tumor size as response, they observed an inter-observer variability of 19-25%. Therefore, they recommended that greater than 50% decline in the perpendicular tumor diameter should be considered as response. In late 70's with the wider availability of CT scan, the World Health Organization (WHO) presented tumor response criteria based on the product of 2 perpendicular diameters (bi-dimensional criteria).³ According to these criteria, the tumor responses were categorized as complete response when there was disappearance of tumor at least 4 weeks after treatment; partial response when ≥ 50% decrease in some of products; and no change when neither partial nor complete response. Progressive disease was defined as ≥ 25% increase in sum of product with or without appearance of new lesion. However, WHO criteria had severe limitation of overestimation of the progressive disease because an increase of 11% in each dimension gives a product of 25% (false positive). Other limitations were that there was no mention of the number of lesions that should be measured or the smallest lesion that should be considered as a measurable one.

In the year 2000, a comprehensive response evaluation criteria in solid tumor (RECIST) was published which was revised (RECIST 1.1) in 2009.⁴,⁵ These criteria in contradistinction to WHO criteria recommended using the single longest diameter of tumor (uni-dimensional) and sum of longest dimensions was recommended to categorize response evaluation. RECIST and RECIST 1.1 have addressed most of the limitations of WHO criteria as well. RECIST 1.1 recommends measuring maximal 5 target lesions and 02 lesions per organ (target non-nodal lesion: ≥ 10 mm in the longest dimension; target node: ≥ 15 mm short axis dimension); and ≥30% declining the sum of the longest dimension is considered as partial response. Revised RECIST 1.1 criteria also recommend an increase in 20% in sum of the longest dimension with at least 5 mm absolute increase in size of a lesion to qualify for progressive disease.

The basic limitations of RECIST criteria are variability in the measurement of tumor sizes by various readers and its application in non-measurable diseases like mesothelioma or bone marrow metastasis. Furthermore, with the introduction of the cytostatic therapy (Imatinib in GIST and Sorafenib in Hepatoma) tumor shrinkage is not expected in the early phase of the treatment. Having a residual mass at completion of the chemotherapy is not an unusual finding and poses difficulty to assess the response solely on anatomical criteria.

Histopathologically, these residual tumor masses may be composed of non-viable fibrotic tissue or having subclinical viable tumor tissue. This has opened vista for metabolic response criteria. History of metabolic response assessment dates back to 1988 when Israel et al. claimed disappearance of Gallium-67 as an indicator of remission of Non-Hodgkin lymphoma (NHL).⁶ In 1993, Richard Wahl shared his experience of using serial 18-fluorodeoxyglucose (¹⁸FDG) in a patient with breast cancer who responded to chemo-hormone therapy with a progressive decline in ¹⁸FDG uptake.⁷ Based on these evidences, it was found that there is a strong correlation between ¹⁸FDG uptake and metabolic tumor burden and, therefore, changes in the ¹⁸FDG uptake rather than the tumor size must be considered as

¹ Section of PET/CT and NM Imaging, Department of Radiology, The Aga Khan University Hospital, Karachi, Pakistan.
² Department of Radiotherapy, KIRAN, Karachi, Pakistan.
³ Correspondence: Dr. Maseeh uz Zaman, Associate Professor and Section Head NM and PET/CT Imaging, Department of Radiology, AKUH, Karachi.
E-mail: maseeh.uzzaman@aku.edu

Received: March 29, 2017; Accepted: May 27, 2017.
a metabolic indicator for tumor response. Presently, \(^{18}\)FDG because of its wider availability is the most commonly used agent for evaluation of metabolic tumor response. \(^{18}\)FDG based assessment could be made either visually or qualitatively, which requires static image but no blood sampling; or quantitatively, which needs acquisition of the dynamic imaging and arterial blood sampling (cumbersome); and despite being accurate, is not used by most of the clinical centers worldwide. Another method, is the semiquantitative method which requires static whole body imaging without blood sampling and the ratio of tumor \(^{18}\)FDG uptake with injected \(^{18}\)FDG dose normalized to the body weight (BW) or lean body mass (LBM) or body surface area (BSA) is calculated. This ratio is called standardized uptake value (SUV), which along with qualitative (visual) assessment is the most commonly used strategy for assessment of metabolic tumor response.

The first metabolic response criteria were published by European Organization of Research and Treatment of Cancer (EORTC), which was published by Young et al. These criteria emphasized upon the patient preparation, time of scans after the last chemotherapy and use of SUV normalized to the BSA; and for tumor sampling, SUV\(_{\text{max}}\) and SUV\(_{\text{mean}}\) were recommended to be used. Complete metabolic response (CMR) was defined as complete disappearance of tumor activity to the background level while the partial metabolic response (PMR) was defined as >15% decline in SUV\(_{\text{max}}\) after the first or more >25% decline after 2nd chemotherapy while the reduction in the extent of the uptake is not required. Stable metabolic disease (SMD) was defined as increase in SUV\(_{\text{max}}\) <25% or declining in SUV\(_{\text{max}}\) >15% decline but no increase in extent of uptake. Progressive metabolic disease (PMD) was defined as increase in SUV\(_{\text{max}}\) >25%, increase in extent >20%, or appearance of a definite new \(^{18}\)FDG avid lesion. However, the limitation of EORTC criteria included significant changes in the individual tumor SUVs and the use of the retrospective data having limited number of patients, tumor types and treatment strategies.

Richard Wahl presented a comprehensive Positron Emission Response Criteria In Solid Tumor (PERCIST) in 2009. PERCIST criteria stress upon the utilization of SUV which is normalized to LBM (SUV\(_{\text{lbm}}\) = SUL) and use SUL\(_{\text{peak}}\) (measured over 1.3 cm\(^3\) circle) for tumor sampling and not SUV\(_{\text{max}}\) (calculated upon single pixel and subject to image noise) as recommended by EORTC criteria. PERCIST also defined the measurable lesion having a SUL\(_{\text{peak}}\) >1.5 x Liver SUL\(_{\text{mean}}\) +2 standard deviation. It recommends measuring the single hottest lesion for response evaluation. However, measurement up to 5 lesions may be performed to build a prospective data. PERCIST strongly emphasized upon standardization of imaging protocol and recommends that hepatic SUL\(_{\text{mean}}\) uptake time and injected \(^{18}\)FDG doses of 2 studies must be ±20%, ±15 minute and ±10%, respectively. It also stresses upon the utilization of the same scanner as difference between calibrations of two scanners could result in variation in SUL\(_{\text{peak}}\) values. CMR is defined as visual disappearance of all of metabolically active tumors; PMR as decline of >30% (0.8 unit) of SUL\(_{\text{peak}}\) between intensities of hottest lesions pre- and post-therapy (which may not be the same lesion) and no new lesion; SMD as no CMR/PMR/RMD; PMD as increase of >30% (0.8 unit) of SUL\(_{\text{peak}}\) or appearance of a definite \(^{18}\)FDG avid malignant lesion. Limitations of PERCIST include getting SUL\(_{\text{peak}}\) (as calculation of LBM is difficult) and lack of standardization of imaging protocols. According to one study published in 2011, <50% the studies could not meet standardization criteria due to the use of different scanners at Johns Hopkins University. Because of these limitations, PERCIST has been struggling to gain popularity in oncological functional imaging.

However, in \(^{18}\)FDG-avid lymphomas, metabolic response criteria have been widely used worldwide. Because of the high diagnostic accuracy of \(^{18}\)FDG PET/CT in Hodgkin’s lymphoma and many diffuse large B-cell lymphomas (DLBCL), bone marrow biopsy can be avoided. \(^{18}\)FDG PET/CT is used for staging and interim PET (iPET) (after 2 cycles) and end of treatment [ePET] \(^{18}\)FDG PET/CTs are used for response evaluation. However, routine surveillance scan is not recommended due to high incidence of false positive findings resulting in unjustified investigations. Although there are many response evaluation criteria published by many societies, Deauville-5-point score (D-5PS) is the most commonly used criteria worldwide. Score 1 means no uptake and score 2 defined as uptake equal or less than mediastinal blood pool activity. Score 1 and 2 are construed as CMR for both iPET and ePET. Score 3 is defined as uptake greater than mediastinal blood pool activity but equal or less than liver. For iPET, it is interpreted as CMR but for ePET good response. Score 4 is defined as activity moderately greater than liver (1 liver SUVmax) and score 5 as markedly greater than liver (>2 liver SUVmax) uptake with or without a new definite lesion. If uptake is less than baseline scan, score 4 and 5 are interpreted as PMR for iPET and RMD for ePET. If uptake is greater than baseline or no change or appearance of a new lesion, then score 4 and 5 are interpreted as PMD and/or treatment failure.

Tumor response may be evaluated on the basis of anatomical or metabolic metrics. Anatomical changes although appear late and having limitation of reproducibility, RECIST is the most commonly used criteria globally. Metabolic changes which appear quite early in phase of the treatment are more sensitive than morphological changes. \(^{18}\)FDG uptake has a strong correlation with viable tumor burden; and change in

18FDG uptake has significantly high diagnostic accuracy for predicting tumor response. PERCIST criteria must be used for a precise and early prediction of tumor response. However, lack of standardization in PET/CT imaging protocols is the fundamental reason of dearth of its popularity. For 18FDG avid lymphomas, Deauville 5-PS is the most reliable response evaluation criteria which are being used for both Hodgkin and Non-Hodgkin lymphomas worldwide. It is imperative that all existing PET/CT facilities must work together to adopt same imaging protocols to make subtle changes in 18FDG tumor uptake more meaningful for better disease outcome.

REFERENCES