RESEARCH ARTICLE

Comparison of WHO and RECIST Criteria for Evaluation of Clinical Response to Chemotherapy in Patients with Advanced Breast Cancer

Samina Khokher1*, Muhammad Usman Qureshi2, Naseer Ahmad Chaudhry3

Abstract

When patients with advanced breast cancer (ABC) are treated with neoadjuvant chemotherapy (NACT), efficacy is monitored by the extent of tumor shrinkage. Since their publication in 1981, World Health Organization (WHO) guidelines have been widely practiced in clinical trials and oncologic practice, for standardized tumor response evaluation. With advances in cancer treatment and tumor imaging, a simpler criterion based on one-dimensional rather than bi-dimensional (WHO) tumor measurement, named Response Evaluation Criteria in Solid Tumors (RECIST) was introduced in 2000. Both approaches have four response categories: complete response, partial response, stable disease and progressive disease (PD). Bi-dimensional measurement data of 151 patients with ABC were analysed with WHO and RECIST criteria to compare their response categories and inter criteria reproducibility by Kappa statistics. There was 94% concordance and 9/151 patients were re-categorized with RECIST including 6/12 PD cases. RECIST therefore under-estimates and delays diagnosis of PD. This is undesirable because it may delay or negate switch over to alternate therapy. Analysis was repeated with a new criteria named RECIST-Breast (RECIST-B), with a lower threshold for PD (≥10% rather than ≥20% increase of RECIST). This showed higher concordance of 97% with WHO criteria and re-categorization of only 4/151 patients (1/12 PD cases). RECIST-B criteria therefore have advantages of both ease of measurement and calculations combined with excellent concordance with WHO criteria, providing a practical clinical tool for response evaluation and offering good comparison with past and current clinical trials of NACT using WHO guidelines.

Keywords: Response criteria - WHO - RECIST - breast cancer - clinical response - RECIST-breast

Asian Pacific J Cancer Prev, 13, 3213-3218

Introduction

A vast majority of breast cancer patients present with advanced disease in the developing countries (Chopra et al., 2001; Yip et al., 2006; Agarwal et al., 2007; Eniu et al., 2008; Khokher et al., 2012). These patients are treated with primary chemotherapy/Neoadjuvant chemotherapy (NACT) to achieve operability, to improve surgical options in favour of breast conservation surgery and to stop the natural evolution of the disease. Traditionally the efficacy of these drugs has been monitored by the extent of tumor shrinkage. Standardization of the extent of tumor shrinkage by using a common language is important for comparing the efficacy of different drug regimens used and for the comparison of results from clinical studies/trials. Evaluation of tumor size changes in response to treatment is a critical issue in the management of advanced breast cancer in clinical practice especially in the developing countries. About thirty years ago World Health Organization (WHO) guidelines for standardized tumor response evaluation were published (Miller et al., 1981). These guidelines were widely accepted and became known as “WHO criteria” for reporting results of cancer treatment. This criterion was based upon measuring the two maximum perpendicular dimensions of the tumor mass. Depending upon the percentage change in the product of these dimensions, four response categories were identified; Complete Response (CR), Partial Response (PR), Stable Disease (SD) and Progressive Disease (PD). These guidelines have been widely practiced in clinical practice as well as in the research settings. Different research groups and investigators have applied it for the assessment of response to chemotherapy in almost all solid tumors whether assessed by clinical examination or by diagnostic imaging. The prefix “c” with these response categories denotes the use of clinical examination for assessment of tumor size. This is important especially in clinical complete response (cCR) which needs to be differentiated from the pathologic complete response (pCR).

With the advancement of cancer treatment and tumor imaging modalities new issues arose and to deal with them, Response Evaluation Criterion In Solid Tumors (RECIST) was introduced in 2000 (Therasse

1Surgical Division, INMOL Hospital, 2Nauman Associates, 3Pathology Department, University of Health Sciences, Lahore, Pakistan
*For correspondence: drsamkhokher@yahoo.com

et al., 2000). The RECIST criteria is based upon One-dimensional measurement of the tumor although the four categories of response described by WHO are retained with different cut-off values. The RECIST criterion is easier in application as well as in calculations and has been widely applied for assessment of response of solid tumors. Diagnostic imaging is considered indispensable for the application of this criterion and modern imaging techniques further increase the opportunity for objectivity and standardization in response evaluation. The WHO criterion however is still being practiced for clinical evaluation of response to chemotherapy in patients with breast cancer (Penault-Llorca et al., 2008; Von Minckwitz et al., 2008; Miglietta et al., 2009; Khokher et al., 2010; Cao et al., 2012). It has been argued that there is no major discrepancy in the response groupings based upon the WHO or RECIST criteria (James et al., 1999; Therasse et al., 2000: 2002: 2006; Park et al., 2003) and owing to easier measurement and calculations, RECIST is more convenient and simpler for response evaluation of breast tumors in the clinical setting. Here we attempt to evaluate the level of agreement/concordance between WHO and RECIST criteria and the inter criteria reproducibility by comparing the results obtained with these two systems, when applied to the same set of patients with advanced breast cancer.

Materials and Methods

All female patients registered at the Institute for Nuclear Medicine and Oncology Lahore (INMOL) with advanced Breast Cancer (ABC) at initial diagnosis, between 1st July 2005 to 30th June 2007, having 5 cm or more tumor size (Clinically evaluable tumor) with plan of Neoadjuvant chemotherapy were included in the study group. The study was designed to investigate the predictors for response to chemotherapy in breast cancer and was approved by the ethical committee and advanced research board of University of Health Sciences and INMOL hospital, Lahore. Tumor measurements were made in two dimensions by the breast surgeon in centimetres using callipers and a tape measure according to the standard procedure already described (Kuerer et al., 2000; Khokher et al., 2010). Measurements were recorded prior to the first cycle of chemotherapy and 3 weeks after the third cycle (prior to 4th cycle) of chemotherapy. Clinical response to NACT was evaluated by the WHO criteria. cCR was defined as complete disappearance of tumor mass, clinical Partial Response (cPR) when there was ≥50% reduction in the product of two perpendicular dimensions of the tumor mass, clinical Progressive Disease (cPD) when there was ≥25% increase in the product of two perpendicular dimensions of tumor and clinical Stable Disease (cSD) when the change did not meet the criteria for other categories. cCR and cPR were grouped as responders and cSD and cPD as non-responders. In case of multiple or bilateral lesions, measurements of the largest lesion alone was recorded. The results have already been published (Khokher et al., 2010: 2011).

The study data was retrieved and analyzed after excluding the patients with multiple or bilateral tumor masses (not eligible for evaluation by RECIST criteria as measurements of tumor mass other than the largest tumor mass were not available) and response evaluations were done by using the WHO and the RECIST criteria. The largest dimension (LD) of the tumor recorded prior to the first and the 4th cycle of chemotherapy was used for response evaluation by the RECIST criteria. The definition for cCR remained unchanged, cPR when there was ≥30% reduction in the maximum dimension of the tumor mass, cPD when there was ≥20% increase in the maximum dimension of tumor mass and cSD when the change did not meet the criteria for other categories. The results of response evaluation with WHO criteria were compared with those obtained by applying RECIST criteria, using $k$ statistic to test concordance for overall response. The analysis was then repeated with a modified RECIST criteria labelled as RECIST-Breast (RECIST-B). The definitions for cCR, cPR and cSD remained the same as for RECIST criteria but the cut off for cPD was fixed at ≥10% increase in the LD. The results of response evaluation with WHO criteria were compared with those obtained by applying RECIST-B criteria, using Kappa ($k$) statistics to test concordance for overall response.

Results

During the period of 1st July 2005 to 30th June 2007, 215 patients were registered with clinically evaluable tumor, having 5 cm or more tumor size, with plan of NACT at INMOL. Fifty patients were excluded from the study as final response data was not available and fourteen patients were excluded because they had multiple or bilateral breast masses without the complete measurement data. The remaining 151 patients were therefore eligible for response evaluation by both WHO and RECIST criteria. Among the 94 patients categorized as cPR according to WHO criteria, one was categorized as cSD and among the 35 patients categorized as cSD according to WHO criteria, two were categorized as cPR, by the RECIST criteria. Among the 12 patients categorized as cPD by WHO criteria, 6 were categorized as cSD by the RECIST criteria. Overall, 9/151 (6%) patients were therefore re-categorized by the use of RECIST criteria. The number of cSD patients was higher and number of cPD patients was lower with the RECIST criteria. The concordance level between the two criteria by applying $k$ statistics was 0.939 (94%) for overall response groups. The largest re-categorization was from the cPD group of WHO to cSD group of RECIST.

![Figure 1. Distribution of Response Outcome with the Three Criteria](image-url)
Table 1. Clinical Response Categories and Inter Criteria Concordance

<table>
<thead>
<tr>
<th></th>
<th>WHO</th>
<th>RECIST</th>
<th>RECIST-B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cCR</td>
<td>cPR</td>
<td>cSD</td>
</tr>
<tr>
<td>cCR</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cPR</td>
<td>0</td>
<td>93</td>
<td>1</td>
</tr>
<tr>
<td>cSD</td>
<td>0</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>cPD</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>95</td>
<td>40</td>
</tr>
</tbody>
</table>

\[ k = 0.94 \quad k = 0.973 \]

*a*C: clinical complete response, cPR: clinical partial response, cSD: clinical stable disease, cPD: clinical progressive disease,
WHO: World Health Organization criterion, RECIST: Response Evaluation Criteria in Solid Tumors, RECIST-B: RECIST- Breast criteria, k: Kappa concordance

This accounted for 6/12 (50%) patients categorized as cPD by the WHO criteria. The tumor dimensions of 9 patients with discordance (Group A) were compared with the other group of patients (Group B), with regard to the ratio between the two dimensions in centimetres to assess their shape. The ratio between longest diameter and its longest perpendicular diameter of a sphere is 1:1 while it is 1.5:1 for an ellipsoid. Applying this formula there were only 11/151 (7%) ellipsoid tumors in this study population and all of them were in group B (group with discordant results of WHO criteria with RECIST criteria).

The analysis was repeated after lowering the threshold of cPD from ≥20% increase to ≥10% increase in the LD (the improvised RECIST-B criteria). Only 4 patients among the 151 were now re-categorized; one categorized as cPR by WHO was categorized as cSD by RECIST-B, two categorized as cSD by WHO were categorized as cPR by RECIST-B and one categorized as cPD by WHO was categorized as cSD by RECIST-B. The overall disagreement was in 2.6% (4/151) patients and concordance with k statistic was now 0.97 (97%).

Table 1 shows the clinical response groups and overall concordance between the three criteria when applied to this set of patient population along with their respective levels of concordance. Figure 1 shows the clinical response groups, with the use of the three criteria in the same set of patient population showing good concordance between WHO and RECIST and excellent concordance between WHO and RECIST-B criteria.

Discussion

In this study we utilized the prospectively collected data in the context of prediction of response to chemotherapy in patients with advanced breast cancer of a previously published study (Khokher et al., 2010). We have found high level of agreement between WHO and RECIST criteria in assigning the overall response and the clinical response category except for the PD group of patients. We have demonstrated that the concordance can be enhanced by the use of RECIST-B criteria which sets the cut off for PD at ≥10% increase in the largest tumor dimension. The application of RECIST-B criteria in clinical practice and in the clinical trials can provide the ease and simplicity of RECIST criteria with the added advantage of higher level of agreement and concordant response categories with the historically standard WHO criteria. The basic assumption in RECIST criteria is that all solid tumors are nearly spherical and the responding tumors have equivalent reductions in all the dimensions. The measurement of one dimension is therefore enough to estimate the changes in its size. The assumption of WHO criteria on the other hand is that the solid tumors may be spherical or ellipsoidal and there may not be equivalent reductions in all the dimensions. The measurements of two dimensions are therefore required to estimate changes in its size. The finding that only 7% tumors in the present study were ellipsoids and that they were all in the group with discordant results by WHO and RECIST criteria proves the validity of RECIST criteria. This also validates that the group with concordant and discordant results, had nearly similar shapes of tumors. The reason of discordance therefore is unlikely to be the shape of tumor mass or the use of one Vs two dimensions of tumor mass for calculations and assessment of response. Assuming that the reason of discordance is the cut off value for PD, the cut off was arbitrarily reduced to ≥10% and repeat analysis showed excellent concordance of 97% with WHO criteria as compared to 94% of the RECIST criteria.

Use of a common language for degree of response to chemotherapy in breast cancer is important for medical management decisions in practice as well as in the clinical trials. Any criteria for response evaluation should be objective, quantifiable, reproducible and a surrogate marker for patient’s survival or some other clinical benefit. The land mark publication “Reporting results of cancer treatment” introduced the WHO criteria for this purpose to the medical community (Miller et al., 1981). The WHO criteria was widely accepted and used for reporting results of cytotoxic drug treatment in solid tumors. However the categorization of four response groups based upon measurement of two tumor dimensions, calculation of their product and calculation of the percentage change in the product of two tumor dimensions was considered laborious with inherent risk of error. It was suggested that changes in tumor diameter relate more closely to the efficacy of chemotherapy rather than the changes in the bi-dimensional product (James et al., 1999). It was further argued that assessment of tumor response based upon WHO criteria was poorly reproducible between investigators and the criteria for selection and number of target lesions was poorly defined. To clarify and simplify the tumor response assessment rules, RECIST group consisting of European Organization for Research on Treatment of...
Cancer (EORTC), the National Cancer Institute of USA (NCI) and the National Cancer Institute of Canada (NCIC) was constituted. Through this international collaboration new guidelines for evaluation of response in solid tumors were published (Therasse et al., 2000). The RECIST and WHO criteria were applied to the same patients recruited in 14 different trials including five for breast cancer (Table 4 in Therasse et al., 2000) and no significant difference was found in the percentage responders. The primary focus of the authors developing RECIST guidelines was response evaluation in early clinical trials (Phase I and II), where one needs to determine whether the particular drug/regimen demonstrates sufficient response to warrant further testing. The use of these guidelines as a guide on decisions of continuation or discontinuation of any drug/regimen in the context of clinical practice was not intended. Moreover they focused more on responders rather than the non-responders as is evident by the fact that in 8/14 trials evaluated for comparison with the WHO criteria, the data of only the responders was compared and no comparison is made for the SD and PD group.

Some form of tumor imaging is recognized as indispensable for the application of RECIST. Although the element of subjectivity cannot be completely eliminated from any evaluation method, diagnostic imaging provides an excellent opportunity for objectivity and standardization as compared to the clinical evaluation. Clinical examination however is considered fairly reliable for measuring breast tumors (Herrada et al., 1997; Fiorentino et al., 2001; Sperber et al., 2006) especially in the advanced cases treated with NACT. Clinical evaluation is technology independent as well as readily available and affordable at all levels of health care facility. The basic requirement for application of RECIST is measurement of the longest diameter of any solid tumor. Breast cancer is a type of solid tumor in an organ which is easily accessible and can be simply measured by clinical examination. Majority of breast cancer patients present with big, sometimes ulcerating masses in the breast, in Pakistan and other developing countries (Chopra et al., 2001; Bhurgri et al., 2006; Yip et al., 2006; Agarwal et al., 2007; Eniu et al., 2008; Khokher et al., 2012). Tumor measurement by imaging in these patients is not only un-necessary but also superfluous and unaffordable. Measurement of only one dimension (the longest diameter) with easier and simpler calculations for response evaluation using RECIST guidelines, suits the clinician of any developing country very well. The oncologic clinical practice in these countries is characterized by disease presentation in advanced stages and enormous work load with limited resources and technology for the diagnosis and management (Aziz., 2008). One should however be clear on the inter criteria reproducibility while comparing results of studies using RECIST or WHO criteria. With the widespread use of multi-detector computed tomography and functional and molecular imaging modalities used for the diagnosis and follow up of malignant tumors, need for further modifications in RECIST were realized and new version (Version 1.1) was introduced (Eisenhauer et al., 2009). RECIST has been found inadequate for evaluation of response in many tumors like malignant lymphoma (Cheson et al., 1999; Cheson et al., 2007) Gastrointestinal Stromal tumors (Choi et al., 2004; Choi et al., 2007), pleural mesothelioma (Van Klaveren et al., 2004; Careshi et al., 2007), prostate cancer (Scher et al., 2005), hepatocellular carcinoma (Faivre et al., 2011; Spira et al., 2011; Edeline, 2012) and renal cell carcinoma (Hutson, 2011) etc and therefore new criteria have been introduced for these tumors. Further modifications or new criteria are continuously being worked upon as the debate on further standardization of response in all types of tumors continues (Husband et al., 2004) and as the new imaging modalities like Positron Emission Tomography (PET) scans are being used for the response evaluation (Wahl et al., 2009). It is expected that with the dawn of the era of molecular medicine, many new cancer specific and therapy specific response criteria will be developed to complement the pitfalls of the RECIST criteria (Nishino et al., 2012).

WHO criterion, being the first criterion introduced for the standardization of tumor response evaluations was followed for almost twenty years prior to the publication of RECIST, in thousands of clinical trials and land mark studies of NACT in breast cancer (Ellis et al., 1998; Fisher et al., 1998; Bonadonna et al., 1999; Wolmark et al., 2001; Bear et al., 2003). It continues to be used in the clinical trials of breast cancer patients treated with primary chemotherapy even a decade after the introduction of RECIST (Penault-Llorca et al., 2008; Rastogi et al., 2008; Miglietta et al., 2009; Von Minckwitz et al., 2009; Cao et al., 2012). WHO criterion, therefore shall appear as a reference for many historical comparisons in the foreseeable future and there is a need to continue exploration of similarities and differences on prospective data for patients with breast cancer. Any change in the tumor size can be better assessed by a change in the tumor volume rather than a change in its linear measurements. The four categories of tumor response described in the WHO guidelines, have been retained by the RECIST working group. When their calculations are translated into volume changes, they give almost similar values for the responders but differ in defining the PD group. The 50% decrease in the product of two dimensions (2r²) by WHO as well as 30% decrease in LD (2r) by RECIST, is equivalent to 65% decrease in volume (4/3πr³). As regards the PD however, 25% increase in the product of two dimensions (2r²) by WHO is equivalent to 40% increase in the volume, while 20% increase in the largest dimensions (2r) by RECIST is equivalent to 73% increase in the volume. The use of same terminology for different degrees of response related to PD appears unjustified especially when this does not apply to the other response groups of CR and PR. The under estimation of PD by RECIST has been considered advantageous in the Phase I and II clinical trials employed as the first screening of efficacy for a treatment regimen (Gehan and Tefft, 2000; Shankar et al., 2009) and in tumors like lung cancer with few therapeutic options. The under estimation of PD by RECIST however is a disadvantage in breast cancer as a number of alternative therapies are available to deal with.
the disease progression on first line cytotoxic drugs. This may be crucial for a clinician dealing with inoperable or borderline operable locally advanced breast cancer cases seen in the clinical practice of the developing countries. Categorization of such a patient as PD when there is 73% increase in its volume (RECIST) rather than the 40% (WHO) may delay diagnosis of progressive disease with continued ineffective chemotherapy which may delay or negate the opportunity of an alternative therapy/surgery with chance of good local control of the disease. The RECIST-B criterion described in the present study facilitates the working of a busy clinician dealing with advanced breast cancer in the developing countries. It provides simplicity of measurement and ease of calculations required for the clinical management decisions and the development of database of patient outcomes in a format comparable with the clinical trials in the developed world.

In conclusion, owing to its simplicity and reduced risk of error, RECIST criterion may be used for the evaluation of treatment efficacy breast cancer patients in the clinical practice. To achieve higher level of concordance and agreement with WHO criteria enabling comparison with landmark studies in breast cancer as well as for the timely diagnosis of progressive disease in patients with locally advanced breast cancer, it is suggested that the cut off for PD may be fixed at ≥10% rather than the current recommendation of ≥20% and the term RECIST-Breast (RECIST-B) may be used for this criterion.

References


