ISSUE BRIEF

Conference on Clinical Cancer Research
November 2015

Modernizing Measurement of Tumor Response to Therapy: Application to Immunotherapeutics

Steven Averbuch, Bristol-Myers Squibb
Gideon Blumenthal, FDA Office of Hematology and Oncology Products
Wendy Hayes, Bristol-Myers Squibb
Erich Huang, National Cancer Institute
Bruce Johnson, Dana-Farber Cancer Institute
Gary Kelloff, National Cancer Institute
Patricia LoRusso, Yale Cancer Center
Geoffrey Oxnard, Dana-Farber Cancer Institute
Louise Perkins, Melanoma Research Alliance
Andrea Perrone, Merck
Nancy Roach, Fight Colorectal Cancer
Samantha Roberts, Friends of Cancer Research
Eric Rubin, Merck
Larry Rubinstein, National Cancer Institute
Larry Schwartz, Columbia University Medical Center
Lalitha Shankar, National Cancer Institute
Robert Sweetman, Novartis
Marc Theoret, FDA Office of Hematology and Oncology Products

Summary

It is intuitive that changes in tumor size can be indicative of the activity, or lack of activity, of an anti-cancer regimen. Thus, objective metrics of tumor shrinkage (response) and growth (progression) have been codified and are often used in clinical and regulatory decision making: a patient who is classified as having progressive disease may be taken off one therapy and placed on another, an experimental new drug may be advanced from early to late stage clinical trials if it produces a high enough response rate, and a new drug may receive accelerated or even full approval based on the response rate or prolonged progression-free survival. However, existing response criteria are based primarily on historical precedent, and were developed primarily as supportive tools in clinical trials, but may be suboptimal for the accurate determination of benefit from therapy. In particular, these criteria fall short in measuring the activity of some of the most innovative and effective therapies being developed today, those which engage the immune system to recognize and kill cancer cells. Immunotherapies hold the potential for long-term, sustained antitumor activity, and promising results from recent clinical trials have generated tremendous excitement within the oncology community. However, these agents act on a prolonged timescale - delayed responses or even transient progression preceding tumor regression may be observed. Metrics that take this phenomenon into account and reflect the mechanism of immunotherapies are needed. We will discuss here how the cancer research community can come together to develop improved metrics for gauging tumor response and progression for use in drug development "go/no-go" decisions as well as potentially predictive tools for regulatory decisions. More accurate metrics could improve the efficiency of drug development, reduce the risk to companies from advancing an investigational agent into late-stage trials, and expedite patient access to effective therapies. Such new metrics are particularly needed for efficient development of immunotherapies.

Benefits and Limitations of Existing Criteria

Objective criteria for evaluating the change in tumor size in response to anti-cancer therapy were first developed by the World Health Organization (WHO) in the 1970s.[1] The widespread adoption of these criteria enabled a standard approach to tumor evaluation and response reporting by different research organizations. WHO criteria define tumor measurements as the sum of bidimensional values (the cross-

products of the two longest perpendicular diameters) from visible lesions. They also compartmentalize the types of responses to anti-cancer therapy into four categories: complete response, partial response, stable disease, and progressive disease; and further establish thresholds of tumor shrinkage or growth to define these four categories (Table 1). The selection of these thresholds was based not on any correlation with treatment outcomes, but rather on a rudimentary assessment of the variability of tumor measurements made by direct tumor palpation.[2] Although technological methods for measuring tumors have advanced considerably since that time, these crude thresholds remain the basis for subsequent adaptations of these criteria.

The WHO criteria were re-evaluated in the 1990s due to variability in their application as well as the advent of improved imaging technologies. The resulting new guideline, RECIST (Response Evaluation Criteria in Solid Tumors), modified WHO criteria by specifying the minimum size of a lesion considered to be "measurable" and by specifying the maximum number of lesions per organ as well as total number of lesions (target lesions) to be assessed.[3] It simplified tumor measurement by defining it as the sum of unidimensional values (longest diameters) of target lesions, with the rationale that unidimensional measurements correlate as well as bidimensional measurements to the amount of tumor cells killed. The threshold defining response was adjusted to a 30% decrease in the longest diameter, which is mathematically equivalent (assuming a spherical tumor) to the previous threshold of a 50% decrease in the products of perpendicular diameters. RECIST also provided specifications for radiologic imaging. RECIST was updated in 2009 (RECIST1.1) with a reduction in the number of lesions to assess, further clarification of disease progression, and additional imaging recommendations.[4]

The overarching goal of these efforts was to create a simple, more reproducible metric that could be compared between different studies. In that respect, they have been extremely useful in oncology trials. The close alignment of WHO and RECIST criteria enables comparison of tumor evaluations by the two methods, ultimately providing researchers today with more than thirty years of clinical trial results for use as historical comparisons. Nevertheless, RECIST1.1 and its predecessors have significant limitations, many of which were recognized even at the time of their publication.[5,6] Some of these limitations relate to the inherent difficulty of measuring the size of tumor lesions. For example, some anatomic sites and types of lesions are difficult to assess by radiologic imaging. In addition, many tumors are irregularly shaped or have diffuse boundaries, making linear measurements imperfect surrogates for tumor burden. These anatomic limitations, and others, have spurred efforts to develop molecular or functional imaging techniques to assess drug activity. For example, FDG-PET (^{18}F-fluorodeoxyglucose positron emission tomography) technology measures tumor metabolism, which can potentially be measured before tumor regression, and is being evaluated in multi-center studies.[7] Another limitation of our current criteria is that they dichotomize what is essentially a continuous variable, and the thresholds defining the discrete categories of response are arbitrary: clinically, there may be little difference between a 19% increase in tumor burden and a 21% increase in tumor burden, but these would be categorized by RECIST as stable disease and progressive disease, respectively. Response rates also do not reflect the depth or duration of patient responses. The proliferation of waterfall plots and other visual, qualitative representations of individual patient tumor responses in phase 2 clinical trial results can be seen as a testament to the inadequacies of our current approach to quantifying drug activity.[8]

Another limitation of RECIST and WHO criteria is that they were designed for the study of cytotoxic chemotherapies, before the entrance of targeted and immunotherapeutics into the clinic and the associated unique radiographic needs of these drug classes. The RECIST categories of progression and response may not accurately capture the activity of some targeted therapies, which may stabilize disease while inducing only minor shrinkage, or immunotherapies, which may have a delayed but profound impact on tumors. For example, some therapies, such angiogenesis inhibitors, are effective in terms of extending survival but produce low RECIST-defined responses. A cut-off of 10% reduction in tumor diameter (or “minor response”) has been proposed as one alternate predictor of long-term benefit than the RECIST-defined

cutoff of 30% decrease in tumor diameter.[9,10,11] Further, in some situations, the presence of objective progression may not indicate treatment failure. Some tyrosine-kinase inhibitors have been shown to provide patient benefit even after RECIST-defined progression, such as sunitinib in renal cell carcinoma and EGFR inhibitors in non-small cell lung cancer.[12] In these situations, continued post-progression treatment may be appropriate, and tumor growth rates have been proposed as a more accurate metric of drug activity.[13,14]

Immune checkpoint inhibitors, such as those targeting PD-1 and CTLA-4, have also been notable for their unusual response and progression kinetics. In some patients, some tumor progression may be observed before tumor regression, and some patients may experience "pseudo"-progression, in which lesions appear enlarged due to the presence of immune cell infiltrates and inflammation at tumor sites.[15] To address this, immune-related response criteria (irRC) were proposed in 2009 based on the body of phase 2 clinical trial data from ipilimumab development that was available at the time.[16] These criteria modified the WHO criteria to describe 4 different potential patterns of response, the latter two of which are not traditionally considered responses: immediate tumor regression with no new lesions, durable stable disease, tumor regression following an initial increase in tumor volume, and regression of the index lesion in the presence of new lesions. A variation of the irRC (immune-RECIST) was subsequently proposed to be more in line with RECIST unidimensional measurements.[17] These immune-based criteria have been shown to more accurately predict survival outcome than RECIST in melanoma patients (based on progression criteria).[14,18] Recent data from nivolumab in non-squamous non-small cell lung cancer have indicated similar findings – patients who were treated beyond RECIST-defined disease progression had median overall survival (OS) similar to those with RECIST-defined stable disease.[19] While this study reported a hazard ratio of 0.73 for OS, a clear dramatic benefit, the hazard ratio for PFS was only 0.92 – compelling data to support the hypothesis that RECIST-defined progression does not accurately reflect benefit for these agents. However, neither irRC nor immune-RECIST have been uniformly adopted for the study of novel immunotherapies; likewise tumor growth rates or alternative cut-offs for tumor growth have not been incorporated as alternative metrics of drug activity.

Ongoing Effort to Identify Improved Metrics of Tumor Growth: Vol-PACT

Although a variety of alternative metrics of tumor growth or shrinkage have been proposed, traditional endpoints continue to be used in most situations. Until recently, there has been no systematic or unified approach to test or validate alternative metrics. To address this, a collaborative research effort has been initiated under the auspices of The Biomarkers Consortium of the Foundation for the NIH (FNIH), a public-private partnership whose goal is to develop opportunities for innovative collaborations between industry, academia, and the philanthropic community to support the mission of the NIH. This effort, Vol-PACT (Volumetric CT for Precision Analysis of Clinical Trial results), aims to collect source imaging data and associated metadata (patient outcomes) from completed large landmark trials in several measurable solid tumors in order to study a variety of potential quantitative response metrics.[20] This effort differs from previous undertakings in the use of source digital images rather than case report forms, which are collected for the purpose of RECIST-based measurements and may not contain all the data elements necessary to perform a quantitative analysis of imaging biomarkers. The CT images, which are collected centrally on most trials and held at various imaging core laboratories, will be transferred to an academic laboratory for volumetric measurement. These images are re-analyzed in a semi-automated fashion with computer-generated contouring to determine unidimensional, bidimensional, and volumetric measurements for each lesion at each time point. Using these measurements, phase 2 trials will be simulated to assess a variety of potential metrics to determine how well they correlate with the survival results that were observed in the phase 3 trials. By studying multiple agents from multiple diseases, this research team has the potential to be able to identify metrics that may be broadly applicable across different therapeutic modalities and tumor types, as well as those that may be appropriate in specific settings.

A key challenge to the Vol-PACT effort has been establishing the infrastructure through which different pharmaceutical companies can share de-identified patient level data. Several criteria are important in selecting trials to include in this study: the images must be physically located at a central imaging repository, have undergone quality control, and have associated patient outcome data. Once candidate trials are identified, contracts must be negotiated with the pharmaceutical sponsor, who may still be trying to gain new marketing indications for the drug under study. Once contracts are in place, the trial data must be de-identified and re-coded so that images are linked to the correct outcome data. To date, several trials of targeted therapies have been selected for study, data has been acquired for most of these, and analysis of one phase 3 trial has been completed. By leveraging existing data from large randomized trials where outcomes are known, this retrospective approach is able to develop a rich and robust dataset without waiting for outcomes from a prospective study and should yield considerable insight into more appropriate measures of response to targeted therapies. Fortunately, this initiative has been able to benefit from an increased emphasis on the importance of data sharing within the entire oncology community, including drug developers. The recognition of big data as a re-usable resource has been integral to this effort and others like it. Establishment of mechanisms for data-sharing in the public domain could facilitate similar projects in the future, such as analyses of functional imaging data in relation to patient outcomes, for example. Recognizing the sensitivities of companies in this competitive space, and the need for patient privacy protections, it could be possible to develop a safe-haven for deposition of de-identified data after IRB approval and publication of trial results.

Moving Forward: Imaging Biomarkers for Immunotherapies

Although Vol-PACT has made considerable progress analyzing trials of targeted therapies, it is only just beginning to enter into data-sharing agreements for trials of immune checkpoint inhibitors. This field is relatively novel; currently three agents targeting two immune checkpoints have been approved for treatment of metastatic melanoma (ipilimumab targeting CTLA-4, and nivolumab and pembrolizumab targeting PD-1), while two of these (nivolumab and pembrolizumab) have been approved for treatment of metastatic non-small cell lung cancer. From a data collection standpoint, immunotherapies present a particular challenge for this type of retrospective analysis due to their unique progression characteristics. Documentation of progression typically results in a change in the clinical management, such that at time of progression, the collection of imaging data on a clinical trial may cease. If a patient is taken off study at RECIST-defined progression but before true clinical progression, it may be difficult to use data from that patient to help develop alternative metrics because no post-progression data has been collected. Fortunately, continued post-progression treatment and imaging is becoming more common in trials of immunotherapies. Further, due to the novelty of this field, while some trials are complete and others are emerging, many are still being designed. Thus, there is an opportunity today to analyze existing data in order to inform the design of ongoing and future clinical trials.

This working group identified key data elements that trials should be collecting now in order to better define progression on immunotherapy. Typically, when RECIST-defined progression is declared, the evidence for this may not be fully documented. It is important to collect detailed data at the time of progression regarding the presence of new lesions and their characteristics, including location and size. Brain metastases may signify worse clinical outcomes as opposed to metastases to other locations. Also, as described earlier, lesions in some anatomical sites are not measurable by radiographic methods. One example is cutaneous lesions, which are of particular importance in melanoma but may be documented on clinical exam rather than using imaging; these types of lesions must be fully documented and described in detail. If a patient is continued on therapy after RECIST-defined progression and is later determined to experience true progression, the clinical characteristics and symptoms that led to this determination should be clearly described. It is also is potentially valuable for trials to document pre-treatment disease trajectory. Some have posited that durable stable disease could be considered as an endpoint of efficacy for immunotherapies. Because stable disease can sometimes reflect the natural course of the disease, there is a risk of overestimating efficacy by labeling a patient with a more indolent disease as having obtained

benefit from the investigational agent. By collecting pre-treatment disease trajectory, it should be possible to differentiate true disease stabilization from naturally slow growing disease.

Conclusion

The limitations of RECIST criteria have been described in numerous publications and many potential alternatives have been proposed for specific settings. However, no concerted effort has yet been undertaken to analyze these potential metrics or determine if they provide value over the currently accepted approach. The Vol-PACT initiative described here aims to change this by evaluating the correlation of candidate response metrics with survival across multiple randomized trials, and could yield improved trial endpoints for drug development and potentially for regulatory decision-making. It remains to be seen whether different criteria will be needed for immunotherapeutic agents versus targeted agents, or whether more generalizable metrics can be identified.

The potential implications of developing improved measures of response and progression are manifold. Faulty phase 2 metrics likely contribute to the failure rate of subsequent phase 3 trials. Moreover, as increasingly adaptive and innovative approaches to drug development are implemented in oncology clinical trials, the boundaries between the distinct phases of development are becoming increasingly blurred. Thus, even "exploratory" endpoints in early phase trials may ultimately be used to support registration if the trial undergoes rapid expansion. In addition, response rates and progression statistics that are reported in clinical trials may be used in correlative studies to identify biomarkers that are predictive of or associated with therapeutic benefit. Basing these correlative studies on metrics that are not themselves based on patient outcomes will ultimately hinder the development of accurate predictive or other biomarkers. Finally, while overall survival remains the "gold standard" for FDA approval, this endpoint can be confounded by post-progression active therapies, which is a significant liability when several active similar agents are in clinical development. Consequently, measures of response and progression serve an important role in regulatory decision-making. Therefore, the development of improved metrics that do reflect patient outcomes is of high importance to all stakeholders.

Table 1: Comparison of Selected Response/Progression Criteria

	WHO	irRC	RECIST 1.1	irRECIST (Modifications may be implemented by pharma)
Measurement method	Bidimensional	Bidimensional	Unidimensional	Unidimensional
# of lesions	10 total; 5 per organ	15 total: 10 visceral and 5 cutaneous	5 lesion; 2 per organ	5 lesion; 2 per organ
Complete Response (CR)	Disappearance of all lesions	Disappearance of all lesions	Disappearance of all lesions	Disappearance of all lesions
Partial Response (PR)	≥ 50% decrease in SPD of index lesions (equivalent to ≥ 65% volumetric change) No new lesions	≥ 50% decrease in SPD of index lesions (equivalent to ≥ 65% volumetric change) New lesions are possible, added to SPD	≥ 30% decrease in SOD of index lesions (equivalent to ≥ 66% volumetric change) No new lesions	≥ 30% decrease in SOD of index lesions (equivalent to ≥ 66% volumetric change) New lesions are possible, added to SOD
Stable Disease (SD)	Neither 50% decrease nor 25% increase in SPD No new lesions	Neither 50% decrease nor 25% increase in SPD	Neither 30% decrease nor 20% increase in SOD No new lesions	Neither 30% decrease nor 20% increase in SOD
Progressive Disease (PD)	≥ 25% increase in SPD of index lesions (equivalent to ≥ 40% volumetric change) Appearance of new lesions or unequivocal PD of non-index lesions	≥ 25% increase in SPD of index lesions (equivalent to ≥ 40% volumetric change) New lesions are possible if adding to SPD does not exceed 25%	≥ 20% increase SOD of index lesions (equivalent to ≥ 73% volumetric change) Appearance of new lesions or unequivocal PD of non-index lesions	≥ 20% increase SOD of index lesions (equivalent to ≥ 73% volumetric change) New lesions are possible if adding to SOD does not exceed 20%

SPD, sum of product of diameters
SOD, sum of diameters

References

[1] Miller AB, Hoogstraten B, Staquet M, Winkler A. "Reporting results of cancer treatment." *Cancer*. 1981 Jan 1;47(1):207-14.
[2] Moertel CG, Hanley JA. "The effect of measuring error on the results of therapeutic trials in advanced cancer." *Cancer*. 1976 Jul;38(1):388-94.
[3] Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. "New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada." *J Natl Cancer Inst*. 2000 Feb 2;92(3):205-16.
[4] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)." *Eur J Cancer*. 2009 Jan;45(2):228-47.
[5] Jaffe CC. "Measures of response: RECIST, WHO, and new alternatives." *J Clin Oncol*. 2006 Jul 10;24(20):3245-51.
[6] Oxnard GR, Morris MJ, Hodi FS, Baker LH, Kris MG, Venook AP, Schwartz LH. "When progressive disease does not mean treatment failure: reconsidering the criteria for progression." *J Natl Cancer Inst*. 2012 Oct 17;104(20):1534-41.
[7] Kelloff GJ, Hoffman JM, Johnson B, Scher HI, Siegel BA, Cheng EY, et al. "Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development." *Clin Cancer Res*. 2005 Apr 15;11(8):2785-808.
[8] Booth CM, Calvert AH, Giaccone G, Lobbezoo MW, Eisenhauer EA, Seymour LK. Design and conduct of phase II studies of targeted anticancer therapy: recommendations from the task force on methodology for the development of innovative cancer therapies (MDICT). *Eur J Cancer* 2008;44:25–9.
[9] Krajewski KM, Franchetti Y, Nishino M, Fay AP, Ramaiya N, Van den Abbeele AD, Choueiri TK. "10% Tumor diameter shrinkage on the first follow-up computed tomography predicts clinical outcome in patients with advanced renal cell carcinoma treated with angiogenesis inhibitors: a follow-up validation study." *Oncologist*. 2014 May;19(5):507-14.
[10] Chen HX, Rubinstein LV, Shankar LK, Abrams JS. "Are we ready for the 10% solution?" *Oncologist*. 2014 May;19(5):439-40.
[11] Zhao B, Oxnard GR, Moskowitz CS, Kris MG, Pao W, Guo P, *et al*. "A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development." *Clin Cancer Res*. 2010 Sep 15;16(18):4647-53.
[12] Lo PC, Dahlberg SE, Nishino M, Johnson BE, Sequist LV, Jackman DM, *et al*. "Delay of treatment change after objective progression on first-line erlotinib in epidermal growth factor receptor-mutant lung cancer." *Cancer*. 2015 Aug 1;121(15):2570-7.
[13] Nishino M, Dahlberg SE, Cardarella S, Jackman DM, Rabin MS, Ramaiya NH, et al. "Volumetric tumor growth in advanced non-small cell lung cancer patients with EGFR mutations during EGFR-tyrosine kinase inhibitor therapy: developing criteria to continue therapy beyond RECIST progression." *Cancer*. 2013 Nov 1;119(21):3761-8.
[14] Burotto M, Wilkerson J, Stein W, Motzer R, Bates S, Fojo T. "Continuing a cancer treatment despite tumor growth may be valuable: sunitinib in renal cell carcinoma as example." *PLoS One*. 2014 May 5;9(5):e96316.
[15] Ribas A, Chmielowski B, Glaspy JA. "Do we need a different set of response assessment criteria for tumor immunotherapy?" *Clin Cancer Res*. 2009 Dec 1;15(23):7116-8.
[16] Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbé C, et al. "Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria." *Clin Cancer Res*. 2009 Dec 1;15(23):7412-20.
[17] Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH, Hodi FS. "Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements." *Clin Cancer Res*. 2013 Jul 15;19(14):3936-43.
[18] Topalian SL, Sznol M, McDermott DF, Kluger HM, Carvajal RD, Sharfman WH, et al. "Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab." *J Clin Oncol*. 2014 Apr 1;32(10):1020-30.
[19] Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. "Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer." *N Engl J Med*. 2015 Oct 22;373(17):1627-39.
[20] http://www.biomarkersconsortium.org/projects_active.php