

Guidelines for the Evaluation of Immune Therapy Activity in Solid Tumors: Immune-Related Response Criteria

Jedd D. Wolchok,¹ Axel Hoos,² Steven O'Day,³ Jeffrey S. Weber,⁴ Omid Hamid,³ Celeste Lebbé,⁵ Michele Maio,⁶ Michael Binder,⁷ Oliver Bohnsack,⁸ Geoffrey Nichol,⁹ Rachel Humphrey,² and F. Stephen Hodi¹⁰

Abstract Purpose: Immunotherapeutic agents produce antitumor effects by inducing cancer-specific immune responses or by modifying native immune processes. Resulting clinical response patterns extend beyond those of cytotoxic agents and can manifest after an initial increase in tumor burden or the appearance of new lesions (progressive disease). Response Evaluation Criteria in Solid Tumors or WHO criteria, designed to detect early effects of cytotoxic agents, may not provide a complete assessment of immunotherapeutic agents. Novel criteria for the evaluation of antitumor responses with immunotherapeutic agents are required.

Experimental Design: The phase II clinical trial program with ipilimumab, an antibody that blocks CTL antigen-4, represents the most comprehensive data set available to date for an immunotherapeutic agent. Novel immune therapy response criteria proposed, based on the shared experience from community workshops and several investigators, were evaluated using data from ipilimumab phase II clinical trials in patients with advanced melanoma.

Results: Ipilimumab monotherapy resulted in four distinct response patterns: (a) shrinkage in baseline lesions, without new lesions; (b) durable stable disease (in some patients followed by a slow, steady decline in total tumor burden); (c) response after an increase in total tumor burden; and (d) response in the presence of new lesions. All patterns were associated with favorable survival.

Conclusion: Systematic criteria, designated immune-related response criteria, were defined in an attempt to capture additional response patterns observed with immune therapy in advanced melanoma beyond those described by Response Evaluation Criteria in Solid Tumors or WHO criteria. Further prospective evaluations of the immune-related response criteria, particularly their association with overall survival, are warranted. (Clin Cancer Res 2009;15(23):7412–20)

The direct cytotoxic mode of action of chemotherapeutic agents often translates into meaningful measurable effects [e.g., tumor shrinkage in baseline (index) lesions] within a few weeks of initial administration. Studies indicate that achieving a response after the initial cycles of chemotherapy is predictive of complete remission (CR) and improved survival (1, 2). Response criteria for solid tumors were developed by the WHO in an attempt to standardize the characterization of chemotherapeutic efficacy and to facilitate comparisons between

studies as well as comparisons with historical data (3, 4). More recently, following the guidelines of a large, international collaboration, simplified and standardized response definitions were published by the RECIST Group in 2000 (5). RECIST guidelines have since been revised and version 1.1 was published in January of 2009 (6). For cytotoxic agents, these guidelines assumed that an early increase in tumor growth and/or the appearance of new lesions signaled progressive disease (PD), such that the term “progression” became synonymous

Authors' Affiliations: ¹Ludwig Center for Cancer Immunotherapy and Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York; ²Bristol-Myers Squibb Company, Wallingford, Connecticut; ³Medical Oncology, The Angeles Clinic and Research Institute, Santa Monica, California; ⁴Department of Oncologic Sciences, University of South Florida, Tampa, Florida; ⁵Department of Dermatology, Saint-Louis Hospital, Paris, France; ⁶Division of Medical Oncology and Immunotherapy, University Hospital of Siena, Istituto Toscano Tumori, Siena, Italy; ⁷Department of Dermatology, Medical University of Vienna, Vienna, Austria; ⁸Perceptive Informatics, Berlin, Germany; ⁹Medarex, Inc., Princeton, New Jersey; and ¹⁰Dana-Farber Cancer Institute, Boston, Massachusetts

Received 6/24/09; revised 8/14/09; accepted 8/18/09; published OnlineFirst 11/24/09.

Grant support: Bristol-Myers Squibb Company.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: J.D. Wolchok and A. Hoos share first authorship. R. Humphrey and F.S. Hodi share senior authorship.

Requests for reprints: Jedd D. Wolchok, Department of Medicine, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, Room Z-1462, New York, NY 10021. Phone: 646-888-2395; Fax: 646-422-0453; E-mail: wolchokj@mskcc.org.

© 2009 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-09-1624

Translational Relevance

Investigators have relied on Response Evaluation Criteria in Solid Tumors or modified WHO criteria to evaluate antitumor responses to chemotherapeutic agents. However, the responses that are seen with immunotherapeutic agents may extend beyond those of cytotoxic agents and could include responses after disease progression that are not captured by Response Evaluation Criteria in Solid Tumors or WHO criteria. The current report presents novel criteria, designated immune-related response criteria (irRC), which can better capture the response patterns observed with some immunotherapeutic agents. Use of the irRC may allow more comprehensive evaluation of immunotherapeutic agents in clinical trials and, potentially, may offer guidance in clinical care. As such, the irRC offer a new tool for clinical investigation of immune therapy in cancer patients.

with drug failure. Cessation of the current treatment is recommended once PD is detected.

Increasing clinical experience indicates that traditional response criteria may not be sufficient to fully characterize activity in this new era of targeted therapies and/or biologics. For example, stable disease (SD) is characterized as either an increase or a decrease in tumor burden insufficient in magnitude to qualify as PD or a partial response (PR), respectively. With chemotherapy, SD is often transient and therefore not considered indicative of true antitumor activity. In contrast, with tyrosine kinase inhibitors (e.g., targeting epidermal growth factor receptor in non-small cell lung cancer), achieving SD has been identified as a potential surrogate end point for improved clinical outcome (median time to progression; ref. 7). Interpretation of this end point under the WHO and RECIST criteria, therefore, has been revisited in recent years, and durable modest regressions or prolonged SD achieved by these agents is, in some cases, now viewed as evidence of activity (8).

With immunotherapeutic agents, which enhance antitumor immune responses (9), SD may also be viewed as an indicator of meaningful therapeutic effect. Beyond that, additional novel response patterns observed with these agents raise concerns about the interpretation and characterization of activity using WHO or RECIST criteria. In studies with cytokines, cancer vaccines, and monoclonal antibodies (e.g., ipilimumab), CR, PR, or SD has been shown to occur after an increase in tumor burden characterized as PD by WHO or RECIST criteria (10–13). For example, in patients with HIV-related Kaposi sarcoma on a stable antiviral regimen, anticancer responses to recombinant interleukin-12 varied from patient to patient across a broad time interval and included objective responses after apparent PD (10). Therefore, conventional response criteria may not adequately assess the activity of immunotherapeutic agents because PD (by initial radiographic evaluation) does not necessarily reflect therapeutic failure. Long-term effect on the target disease must also be captured.

In 2004 and 2005, approximately 200 oncologists, immunotherapists, and regulatory experts convened in a series of

workshops to discuss their experience with immunotherapeutic agents in cancer patients (14). These discussions resulted in the following conclusions: (a) The appearance of measurable antitumor activity may take longer for immune therapies than for cytotoxic therapies; (b) responses to immune therapies may occur after conventional PD; (c) discontinuation of immune therapy may not be appropriate in some cases, unless PD is confirmed (as is usually done for response); (d) allowance for “clinically insignificant” PD (e.g., small new lesions in the presence of other responsive lesions) is recommended; and (e) durable SD may represent antitumor activity. The workshop participants proposed a new clinical paradigm and recommended that existing response criteria be refined to address these issues. Toward this end, a novel set of response criteria based on the WHO criteria were evaluated in a series of large, multinational studies, representing a clinical trial program of 487 patients with advanced melanoma who received ipilimumab, a fully human monoclonal antibody that blocks CTL antigen-4 (CTLA-4). In this article, we define systematic criteria that enhance characterization of new response patterns observed with ipilimumab.

Patients and Methods

Clinical observations with anti-CTLA-4 immune therapy (ipilimumab). The binding of CTLA-4 to costimulatory B7 molecules inhibits activation of T cells (15, 16), and thus it is hypothesized that blocking CTLA-4 enhances immune activation and allows for the expansion of T cells with antitumor activity (16). Ipilimumab has been studied in 487 patients in three multicenter phase II trials (CA184-008, CA184-022, and CA184-007) in patients with advanced (unresectable stage III or stage IV) melanoma (17–21). The study design across the phase II clinical trial program was aimed at evaluating the activity and tolerability of ipilimumab as monotherapy in these patients (see Appendix A). All three studies prospectively captured four patterns of clinical responses to ipilimumab and categorized them using a novel set of criteria.

Patients were treated with induction therapy (ipilimumab 10 mg/kg every 3 wk \times 4) followed by maintenance therapy in eligible patients (ipilimumab 10 mg/kg every 12 wk, beginning at week 24). Tumor responses were assessed by an independent review committee (IRC) using conventional modified WHO (heretofore referred to as WHO) criteria to determine the best overall response rate (BORR) and to evaluate responses by the immune-related response criteria (irRC; see Appendix A). Overall survival (OS), 1-year survival, tolerability, and other parameters were also examined. Tumor assessments were first carried out at week 12 (end of the induction dosing period) to allow adequate time for ipilimumab-mediated immune activation and consequent antitumor responses. Data from early phase I/II studies showed that ~60% of objective responses (PR/CR) were captured at this time point (22–24).

It was expected that some patients treated with ipilimumab would experience increased objective tumor burden and/or new lesions before a response was obtained (25). Thus, patients with PD before week 12 (per WHO criteria), but without rapid clinical deterioration, continued ipilimumab treatment and were observed with a stringent imaging schedule to allow detection of an antitumor response. Ipilimumab was discontinued due to PD at week 12, drug intolerance, or withdrawal of consent. Per protocol, it was recommended that patients who experience investigator-determined PD at week 12, at the discretion of the investigator, receive additional tumor assessments before the initiation of alternative anticancer therapy. The IRC evaluated tumor assessments obtained after apparent PD but before administration of non-ipilimumab anticancer treatment. Images were captured before and after WHO-defined PD, and measurable new lesions were added to

index lesion measurements for an assessment of overall tumor burden.

Results

Approximately 30% of patients treated with ipilimumab in phase II trials had disease control (CR, PR, or SD) at week 12. In these studies, SD followed by a steady decline in tumor burden over prolonged periods of time with slow evolution of an objective response (PR or even CR) has been observed (23, 26). Furthermore, some patients characterized as PD at week 12 (by WHO criteria), either by an increase in tumor burden and/or the appearance of new lesions, subsequently experienced an objective response or SD (relative to baseline) without the addition of non-ipilimumab anticancer therapy. Therefore, tumor responses to ipilimumab may not occur until the post-induction period of therapy and may occur in some cases following WHO- or RECIST-defined PD (13, 26).

Across the phase II clinical trial program, four patterns of response to ipilimumab therapy in patients with advanced melanoma were observed (Fig. 1). Two of the response patterns are captured with conventional response criteria: (a) response in baseline lesions-evident by week 12, with no new lesions, and

(b) "stable disease" (which in some patients was followed by a slow, steady decline in total tumor burden). The other two response patterns are new and involve (c) responses after an initial increase in total tumor burden and (d) a reduction in total tumor burden during or after the appearance of new lesion(s) at time points later than week 12.

Immune-related response criteria

To systematically characterize additional patterns of response in patients with advanced melanoma, underlying WHO criteria were evolved into immune-related response criteria (irRC). The definitions of the irRC and guidelines on how they can be used in clinical practice are detailed below.

Antitumor response based on total measurable tumor burden. For the irRC, only index and measurable new lesions are taken into account (in contrast to conventional WHO criteria, which do not require the measurement of new lesions, nor do they include new lesion measurements in the characterization of evolving tumor burden). At the baseline tumor assessment, the sum of the products of the two largest perpendicular diameters (SPD) of all index lesions (five lesions per organ, up to 10 visceral lesions and five cutaneous index lesions) is calculated. At each subsequent tumor assessment, the SPD of the index lesions and of new, measurable lesions ($\geq 5 \times 5$ mm; up to 5 new lesions per organ: 5 new

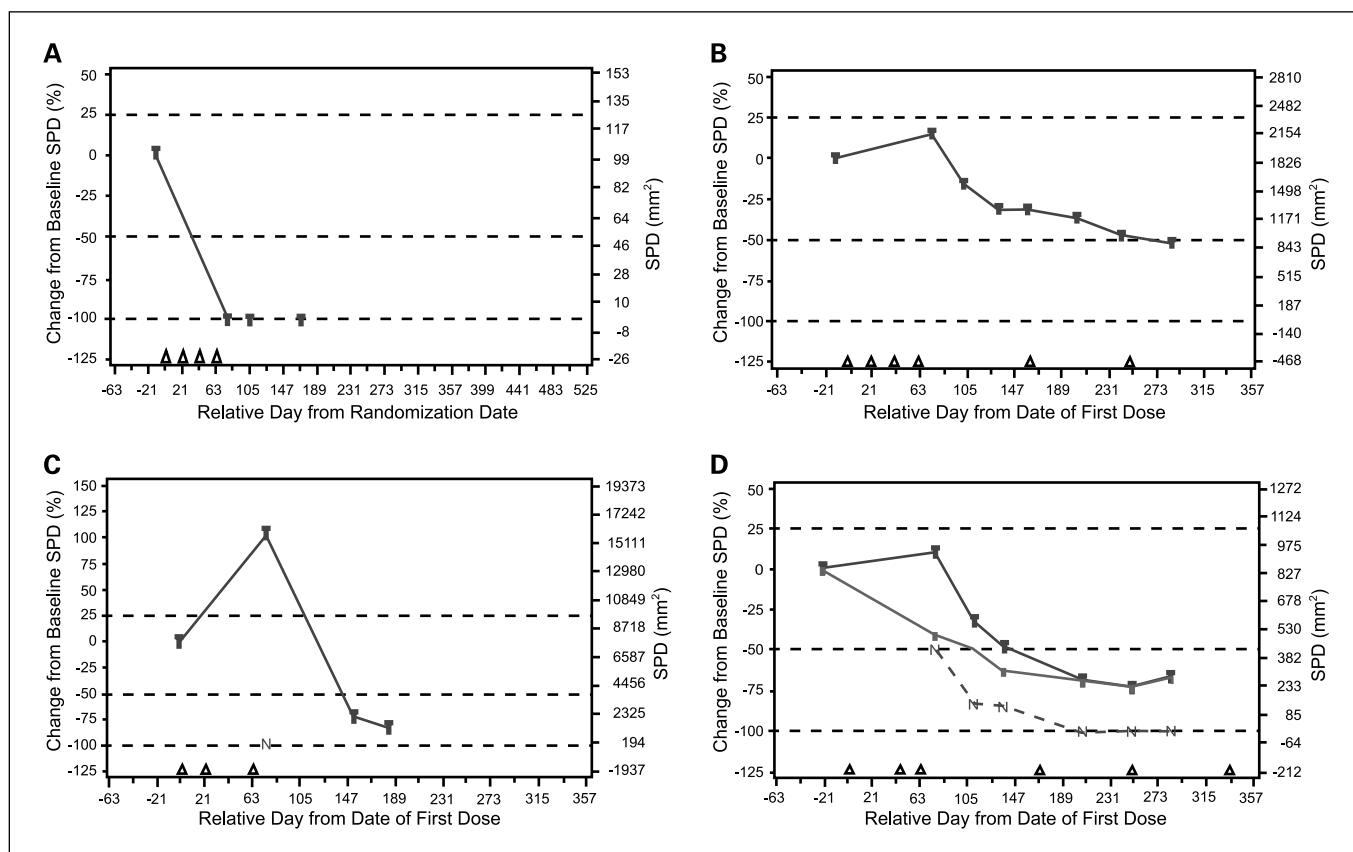


Fig. 1. Patterns of response to ipilimumab observed in advanced melanoma patients treated with ipilimumab at 10 mg/kg in the CA184-008 and CA184-022 studies. A, response in baseline lesions; B, "stable disease" with slow, steady decline in total tumor volume; C, response after initial increase in total tumor volume; D, reduction in total tumor burden after the appearance of new lesions. SPD, sum of the product of perpendicular diameters. N, tumor burden of new lesions (C and D). D, top line, total tumor burden; middle line, tumor burden of baseline lesions; bottom line, tumor burden of new lesions. Triangles, ipilimumab dosing time points; dashed lines, thresholds for response or PD/irPD.

Table 1. Comparison between WHO criteria and the irRC

	WHO	irRC
New, measurable lesions (i.e., $\geq 5 \times 5$ mm)	Always represent PD	Incorporated into tumor burden
New, nonmeasurable lesions (i.e., $< 5 \times 5$ mm)	Always represent PD	Do not define progression (but preclude irCR)
Non-index lesions	Changes contribute to defining BOR of CR, PR, SD, and PD	Contribute to defining irCR (complete disappearance required)
CR	Disappearance of all lesions in two consecutive observations not less than 4 wk apart	Disappearance of all lesions in two consecutive observations not less than 4 wk apart
PR	$\geq 50\%$ decrease in SPD of all index lesions compared with baseline in two observations at least 4 wk apart, in absence of new lesions or unequivocal progression of non-index lesions	$\geq 50\%$ decrease in tumor burden compared with baseline in two observations at least 4 wk apart
SD	50% decrease in SPD compared with baseline cannot be established nor 25% increase compared with nadir, in absence of new lesions or unequivocal progression of non-index lesions	50% decrease in tumor burden compared with baseline cannot be established nor 25% increase compared with nadir
PD	At least 25% increase in SPD compared with nadir and/or unequivocal progression of non-index lesions and/or appearance of new lesions (at any single time point)	At least 25% increase in tumor burden compared with nadir (at any single time point) in two consecutive observations at least 4 wk apart

cutaneous lesions and 10 visceral lesions) are added together to provide the total tumor burden:

$$\text{Tumor Burden} = \text{SPD}_{\text{index lesions}} + \text{SPD}_{\text{new, measurable lesions}}$$

A comparison of the use of SPD in WHO criteria versus the use of tumor burden in irRC is presented in Table 1.

Time-point response assessment using irRC. Percentage changes in tumor burden per assessment time point describe the size and growth kinetics of both conventional and new, measurable lesions as they appear. At each tumor assessment, the response in index and new, measurable lesions is defined based on the change in tumor burden (after ruling out irPD). Decreases in tumor burden must be assessed relative to baseline measurements (i.e., the SPD of all index lesions at screening). The irRC were derived from WHO criteria and, therefore, the thresholds of response remain the same (Table 2). However, the irRC response categories have been modified from those of WHO criteria as detailed in Tables 1 and 2.

Overall response using the irRC. The overall response according to the irRC is derived from time-point response assessments (based on tumor burden) as follows:

- irCR, complete disappearance of all lesions (whether measurable or not, and no new lesions) confirmation by a repeat, consecutive assessment no less than 4 wk from the date first documented
- irPR, decrease in tumor burden $\geq 50\%$ relative to baseline confirmed by a consecutive assessment at least 4 wk after first documentation
- irSD, not meeting criteria for irCR or irPR, in absence of irPD
- irPD, increase in tumor burden $\geq 25\%$ relative to nadir (minimum recorded tumor burden) confirmation by a repeat, consecutive assessment no less than 4 wk from the date first documented

Table 2. Derivation of irRC overall responses

Measurable response	Nonmeasurable response		Overall response Using irRC
	Non-index lesions	New, nonmeasurable lesions	
Index and new, measurable lesions (tumor burden),* %			
↓100	Absent	Absent	irCR [†]
↓100	Stable	Any	irPR [†]
↓100	Unequivocal progression	Any	irPR [†]
↓ ≥ 50	Absent/Stable	Any	irPR [†]
↓ ≥ 50	Unequivocal progression	Any	irPR [†]
↓<50 to <25 [†]	Absent/Stable	Any	irSD
↓<50 to <25 [†]	Unequivocal progression	Any	irSD
≥25?	Any	Any	irPD [†]

*Decreases assessed relative to baseline, including measurable lesions only ($> 5 \times 5$ mm).

[†]Assuming response (irCR) and progression (irPD) are confirmed by a second, consecutive assessment at least 4 wk apart.

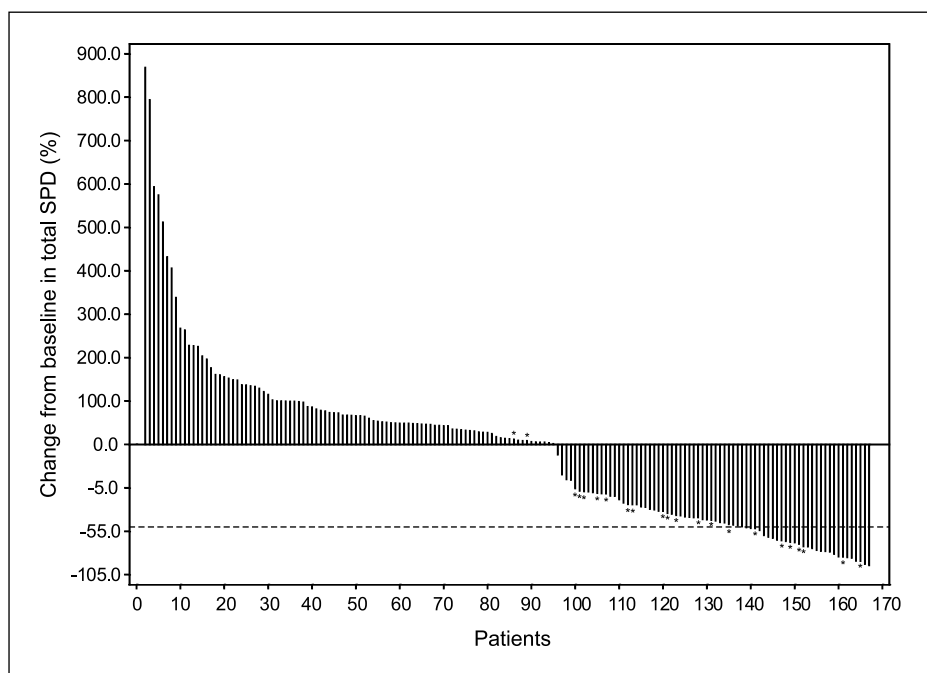


Fig. 2. Waterfall plot of maximum percentage reduction from baseline in total tumor burden. Included are advanced melanoma patients treated with, or randomized to, ipilimumab at 10 mg/kg in the CA184-008 and CA184-022 studies; the tumor responses of 167 evaluable patients were assessed using the irRC. Twenty-two patients were characterized as irPR ($n = 5$) or irSD ($n = 17$), who otherwise would have been labeled “PD” by conventional WHO criteria. These patients are indicated by an asterisk. In addition, one patient characterized as SD by WHO criteria was evaluated as irPR (patient #148).

Patients were considered to have irPR or irSD even if new lesions were present, as long as they met the respective thresholds of response as described above. Furthermore, patients were not considered to have irPD if new lesions were present and the tumor burden of all lesions did not increase by $\geq 25\%$. In contrast to irCR, irPR, and irPD, a response of irSD does not require confirmation. It is important to note that irCR, irPR, and irSD include all patients with CR, PR, or SD by WHO criteria as well as those patients that shift to these irRC categories from WHO PD. Patients with irSD, particularly those with slow-declining tumor burden $\geq 25\%$ from baseline at the last tumor assessment, are considered clinically meaningful because they show an objectively measurable reduction in tumor burden without reaching the 50% threshold that defines irPR (it represented an objectively measured reduction not commonly observed in the natural history of advanced melanoma patients).

If a patient is classified as having irPD at a post-baseline tumor assessment, then confirmation of irPD by a second scan in the absence of rapid clinical deterioration is required. The definition of confirmation of progression represents an increase in tumor burden $\geq 25\%$ compared with the nadir at two consecutive time points at least 4 wk apart. It is recommended that this be done at the discretion of the investigator because follow-up with observation alone may not be appropriate for patients with a rapid decline in performance status. Confirmation of irPD allows for the capture of all observed responses using the irRC (Table 2), as most of these late-responding patients have a trend toward response within 4 wk after initial irPD. Whereas WHO criteria consider any new measurable lesion to indicate PD, determination of immune-related best overall response (irBOR) is based on changes in total tumor burden from the baseline (nadir, for irPD) tumor assessment, regardless of any initial increase in baseline lesions or the appearance of new lesions.

Evaluation of irRC in the ipilimumab phase II program

A total of 227 patients were treated with, or randomized to, ipilimumab at 10 mg/kg in the single-arm study CA184-008 and the dose-ranging study CA184-022, respectively (see Appendix A). The BORR, as assessed by the IRC, was 7.5% (17 of 227). CR and PR were achieved in 0.9% (2 of 227) and 6.6% (15 of 227) of patients, respectively, with SD in 20.3% (46 of 227) of patients. Forty-one patients had no follow-up scan due to early PD (“unknown response”), and 123 patients had a BOR of PD at week 12. Of these 123 patients, 57 were followed beyond WHO PD (at weeks 16, 20, and later) before the institution of other anticancer therapy. After response status per WHO criteria was known, all patients were also evaluated using irRC. Among the patients with WHO PD, the irRC identified 22 with objective responses (irBOR): 5 had an irPR and 17 had irSD.

These analyses show that 9.7% (22 of 227) of treated patients, who were initially characterized as PD by WHO criteria, have evidence of activity consistent with a response to ipilimumab. This suggests that a measurable clinical effect can be present in a subset of patients with an early increase in tumor burden and/or the appearance of new lesions. Follow-up beyond PD (by WHO criteria) could not be mandated by the clinical trial protocols, which limited the data set available for evaluation after PD. Thus, the actual number of patients with responses after PD in these studies may be underestimated. Overall, the data emphasize the need to use evaluation criteria such as irRC to identify patients with activity among those with WHO PD at week 12 or beyond to determine who should continue ipilimumab therapy.

The importance of measuring overall tumor burden is shown by the waterfall plot in Fig. 2. The plot shows that even when the measurements of new lesions are included in the calculation of tumor burden, the net effect is an overall decline in some patients. Kaplan-Meier curves for OS, which include

one group of patients with CR, PR, or SD by WHO criteria and another group of patients that shifted from PD by the WHO criteria to response per irRC, are shown in Fig. 3. These data suggest that patients in these two populations have comparable survival, and that the irRC can identify at least an additional 10% of patients with favorable survival among those characterized as WHO PD. With the irRC, all of these patients would be categorized together under a single heading of irRC responders. As such, there seems to be a disparity between WHO BOR and survival data, which is likely explained, at least in part, by the new response patterns. The data further illustrate the importance of SD in ipilimumab-treated patients, as survival for patients with SD or irSD is similar to that of patients with a response per WHO.

Discussion

The core novelty of the irRC is the incorporation of measurable new lesions into "total tumor burden" and comparison of this variable to baseline measurements (before and after WHO PD, but not after confirmed irPD). The apparent increases in tumor burden that sometimes precede responses in patients receiving immune therapy may reflect either continued tumor growth until a sufficient immune response develops or transient immune-cell infiltrate with or without edema. Examination of tumor biopsies from ipilimumab-treated patients with apparent PD before response is consistent with both hypotheses

(27, 28). For patients who seem to develop new lesions, it may be difficult to differentiate these from baseline, nonmeasurable lesions, as the latter may be due to T-cell infiltration into established, radiographically undetectable tumor deposits. As a result, inflammation in baseline lesions may be misinterpreted as PD (a version of the "tumor flare reaction"). In a case study of an ipilimumab-treated patient that seemed to have PD at the 12-week tumor assessment, histologic analyses showed that the increase in lesion size was likely due to T-cell infiltration rather than tumor cell proliferation (shown in Fig. 4).

Clinical investigation of cancer immune therapies is hampered by the absence of response criteria that can comprehensively describe all patterns of antitumor activity associated with such agents. This was evidenced by reported clinical experiences with cancer vaccines that induced responses of SD or PR (11, 12). In these studies, some responses originally evaluated as SD or PD showed evidence of subsequent tumor regression (11), whereas others showed "mixed responses," consisting of regression in some lesions while others remained stable, progressed, or appeared simultaneously (12). Whereas such patterns have been described by many investigators, the clinical significance of these observations has not been adequately studied due to the lack of suitable response criteria to capture the patterns. Overall, such observations indicate the need for novel criteria in the evaluation of responses to immunotherapeutic agents.

Modifying WHO or RECIST criteria to capture the unique response patterns of immunotherapeutic agents was proposed previously (14) and is increasingly recognized as important for their proper evaluation (29). In 2004 and 2005, a series

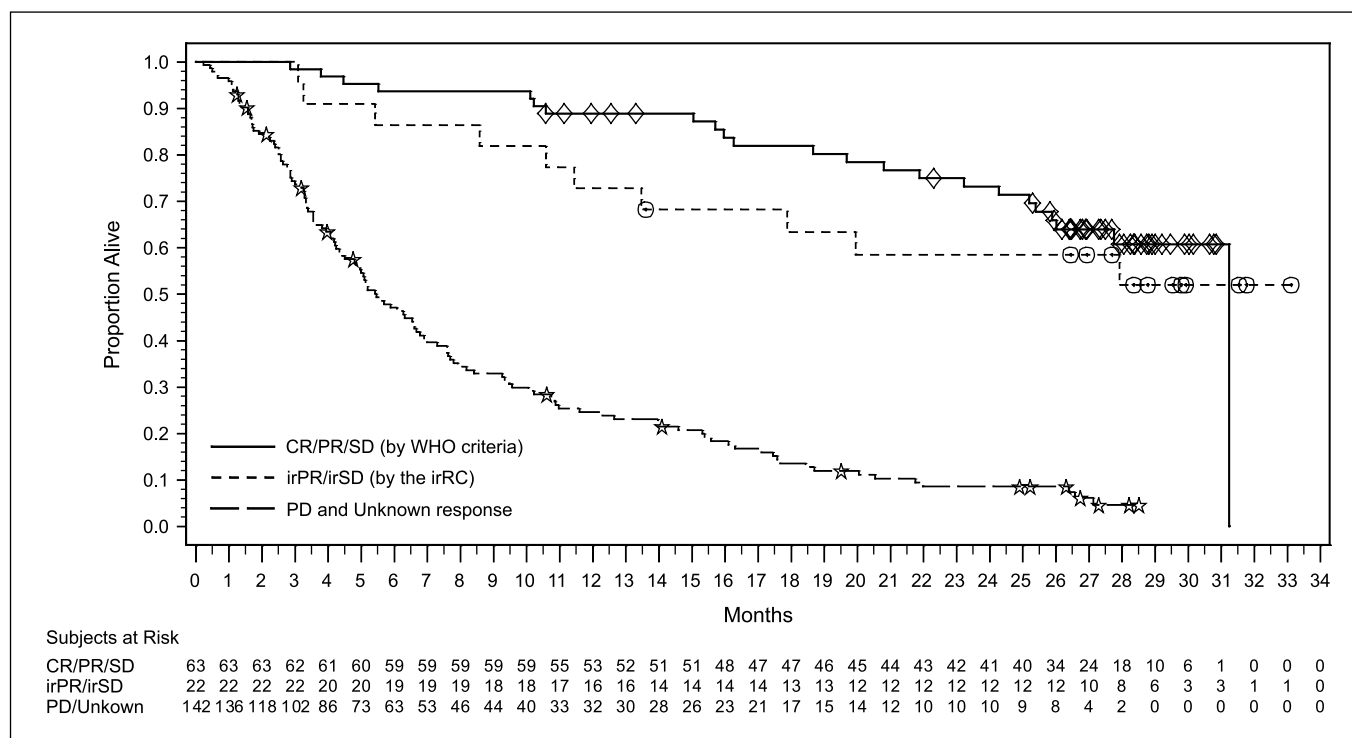


Fig. 3. Association of OS with response using WHO criteria or irRC. Data are included for all patients treated with, or randomized to, ipilimumab at 10 mg/kg in the CA184-008 and CA184-022 studies, respectively ($n = 227$). Numbers of patients by response categories were as follows: 63 with CR, PR, or SD (BOR by WHO criteria); 22 with PD (by WHO criteria) and assessment by the irRC as irPR or irSD; 142 with PD (by WHO criteria) or unknown response. Each patient is included in only one response category. Different symbols for the respective curves indicate censored patients. Median OS in months (95% confidence intervals) corresponding to each curve: CR/PR/SD, 31.2 (27.8-31.2); irPR/irSD, not reached (13.5-not reached); PD/unknown, 5.45 (4.5-6.77).

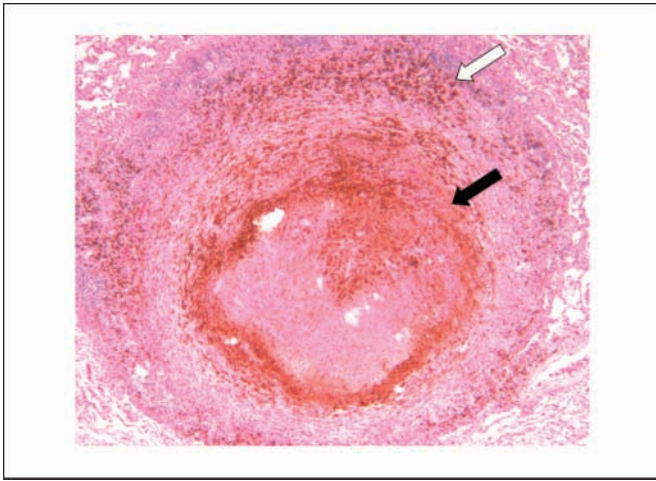


Fig. 4. Resected metastatic melanoma tumor nodule of the lung. This case is a 53-y-old male, diagnosed with melanoma of the scalp, who underwent resection and adjuvant biochemotherapy. After two cycles, imaging confirmed multiple new lung nodules consistent with recurrent disease (stage M1b). Eight months after starting ipilimumab, the dominant lung lesion was resected along with two small nodules (3 mm each). From a biopsy of one of the small nodules, note the T-cell infiltrate (*white arrow*) and extensive necrosis (*black arrow*) with no residual tumor cells. Section was stained with H&E.

of international workshops hosted by the Cancer Vaccine Consortium, in collaboration with the International Society of Biological Therapy of Cancer, reviewed the current knowledge on cancer vaccines and defined, in a consensus process, the foundations of a new clinical development paradigm for these and related agents (14). In reference to the measurement of responses to cancer vaccines, recommendations were made to revise clinical trial end points and to modify existing criteria to take into account the unique properties of these agents. The following were recommended for consideration: the potential for long-term clinical improvements and response after PD, consideration of the benefit-to-risk ratio for discontinuation of therapy, confirmation of PD and an allowance for clinically insignificant PD, and that durable SD may represent benefit.

For immunotherapeutic agents such as ipilimumab that have the ability to induce tumor shrinkage in some patients, the irRC are likely to provide a more comprehensive assessment of clinical activity and may help to explain why patients with apparent PD by the traditional response criteria experience long-term survival. For agents that are less likely to cause considerable tumor shrinkage (e.g., certain cancer vaccines), irRC may help to identify initial signals of activity. However, the importance of the irRC and confirmation of PD as a means to show activity for most immunotherapeutics, such as cytokines, checkpoint modulators, or cancer vaccines, remains to be determined and should be studied prospectively.

We therefore recommend a paradigm shift for the oncologist in the evaluation of immune therapies to ensure assessment of activity based on clinically relevant criteria and time points. Importantly, the irRC share the established thresholds of response and PD with WHO criteria and therefore do not require an adjustment in this regard but allow for the inclusion of new lesion measurements into the assessment. With cytotoxic agents, the current reaction to apparent PD by the classic WHO/RECIST definition is discontinuation of therapy. With ipilimumab or other

immune therapies, an increase in tumor burden or the appearance of new lesions before radiographic responses can be partially circumvented by appropriate follow-up at a subsequent time point to confirm PD. Treatment should be continued as tumors may begin to shrink in this interval. Patients treated with immune therapy whose performance status is stable and whose laboratory values have not significantly deteriorated, as well as those with moderate tumor growth on physical exam or radiographic imaging, should be considered for repeat confirmation scans before true PD is defined and the immunotherapeutic agent is withdrawn. This, of course, needs to be balanced against the potential toxicity associated with continued treatment.

The irRC were defined based on data from ipilimumab clinical trials, but their conceptual foundations result from consistent observations with several agents across the immune therapy community, and therefore it is expected that these criteria will have broad applicability to immunotherapeutic agents. Although potentially an improvement over conventional criteria for immunotherapeutic agents, the irRC may still not capture or fully characterize all relevant patterns of clinical activity. For example, one challenge for the irRC is that the term "irSD" is appropriate both for cases of minimal change in tumor burden over time and for large increases in tumor burden followed by a reduction to baseline levels. In clinical practice, where the current tumor assessment is often compared with the most recent one (i.e., the baseline is "reset" to reflect the latest measurements), this latter case might be more appropriately characterized as an objective response and not as SD. A second challenge for the irRC is that meaningful objective responses that are only first observed after repeated "cycles" of irPD are still categorized as irPD. Therefore, as we further investigate the clinical utility of irRC, they may need to be further optimized. Their use and refinement and the determination of the extent to which they are associated with survival are being prospectively evaluated in phase III clinical trials with ipilimumab.

Appendix

Study designs across the ipilimumab phase II clinical trial program in advanced melanoma. Ipilimumab has recently been studied in three multicenter phase II trials in previously treated and treatment-naïve patients with advanced (unresectable stage III or IV) melanoma (all trials are registered with ClinicalTrials.gov).¹¹ Study CA184-008 was a phase II, open-label, single-arm, multicenter trial in patients who had PD during or after at least one prior therapy. Eligible patients received induction ipilimumab at 10 mg/kg every 3 weeks ×4. Patients with a tumor response or SD and who tolerated treatment were eligible to receive single maintenance doses of ipilimumab at 10 mg/kg every 12 weeks beginning at week 24. Study CA184-022 was a phase II, randomized, double-blind, multi-arm, multicenter, dose-ranging trial in patients previously treated with other agents, but without a CR or PR, or with poor tolerability. Eligible patients were randomized to receive induction ipilimumab at 0.3, 3.0, or 10 mg/kg every 3 weeks ×4. Patients who tolerated ipilimumab and had SD, PR, or CR after induction were

¹¹ <http://www.clinicaltrials.gov>

eligible to receive maintenance doses of ipilimumab every 12 weeks (beginning at week 24) at their assigned dose. Only those patients receiving ipilimumab at 10 mg/kg are included in the present analyses.

Diagnosis and main criteria for inclusion. Patients were males and females, ≥ 16 years of age, who had histologically confirmed, measurable (using modified WHO criteria), stage III (unresectable) or stage IV melanoma, and had progressed during or after at least one prior therapeutic regimen containing one or more of the following: interleukin-2, dacarbazine, fotemustine, or temozolomide (also includes paclitaxel, carboplatin for study CA184-008). Patients were to have a life expectancy of ≥ 16 weeks and an Eastern Cooperative Oncology Group performance status of 0 or 1.

Criteria for evaluation and statistical considerations. Tumor response was evaluated by the investigator and by an IRC based on WHO criteria. The assessment of the IRC was considered primary. Exploratory end points were also assessed using irRC, which were developed from WHO criteria, as a preliminary approach to a systematic categorization of ipilimumab clinical activity before and after PD as defined by WHO criteria. Response using both WHO criteria and irRC was determined by the IRC. The primary analysis of activity was based on IRC BORR (number of subjects with a BOR of CR or PR, divided by the number of treated/randomized subjects). BORR and disease control rate (number of subjects with CR, PR or SD, divided by the number of treated/randomized subjects) were calculated along with corresponding exact two-sided 95% confidence intervals using the method of Clopper and Pearson. Images were captured before and after WHO-defined PD. All scans that were obtained after WHO PD and before alternative anticancer therapy were included in the IRC evaluation. Measurable new lesions were added to

index lesion measurements to obtain a more accurate picture of overall tumor burden.

OS was defined as the time between the first dose of study therapy (CA184-008) or the randomization date (CA184-022) and death; 1-year survival rate was a secondary end point. OS was calculated using the Kaplan-Meier product-limit method to provide the median estimate together with a two-sided 95% confidence interval for the median, calculated using the method of Brookmeyer and Crowley. Survival rates at 1 year were calculated using the Kaplan-Meier method together with a corresponding two-sided 95% bootstrap confidence interval. The immune-related response end points were analyzed using methods similar to those used for the main response end points. Safety was evaluated using the National Cancer Institute Common Terminology Criteria for Adverse Events, based on adverse events, physical examinations, and clinical laboratory assessments.

Disclosure of Potential Conflicts of Interest

A. Hoos and R. Humphrey, employment, Bristol-Myers Squibb Company. G. Nichol, employment, Medarex, Inc. S. O'Day, J.S. Weber, commercial research grant, Bristol Myers Squibb. O. Hamid, commercial support, Bristol Myers Squibb. J.S. Weber, ownership interest, Medarex. S. O'Day, J.S. Weber, O. Hamid, honoraria, Bristol Myers Squibb. J.D. Wolchok, S. O'Day, J.S. Weber, O. Hamid, C. Labbé, M. Maio, F.S. Hodi, consultant, Bristol Myers Squibb. J.D. Wolchok, consultant, Medarex. The other authors declare no potential conflicts of interest.

Acknowledgments

We thank Dr. Lawrence Schwartz (Memorial Sloan-Kettering Cancer Center) for critical review of the manuscript, and StemScientific for editorial and writing assistance.

References

1. von Minckwitz G, Sinn HP, Raab G, et al. Clinical response after two cycles compared to HER2, Ki-67, p53, and bcl-2 in independently predicting a pathological complete response after preoperative chemotherapy in patients with operable carcinoma of the breast. *Breast Cancer Res* 2008;10:R30.
2. Hutchings M, Loft A, Hansen M, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. *Blood* 2006;107:52-9.
3. WHO handbook for reporting results of cancer treatment. Geneva (Switzerland): World Health Organization Offset Publication No. 48, 1979.
4. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981;47:207-14.
5. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 2000;92:205-16.
6. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228-47.
7. Tsujino K, Kawaguchi T, Kubo A, et al. Is response rate or stable disease rate a surrogate endpoint for survival in the treatment of advanced non-small cell lung cancer using epidermal growth factor receptor tyrosine kinase inhibitors? [abstract 14600] *J Clin Oncol* 2008;26.
8. Ratain MJ, Eckhardt SG. Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST. *J Clin Oncol* 2004;22:4442-5.
9. Dougan M, Dranoff G. Immune therapy for cancer. *Annu Rev Immunol* 2009;27:83-117.
10. Little RF, Pluda JM, Wyvill KM, et al. Activity of subcutaneous interleukin-12 in AIDS-related Kaposi sarcoma. *Blood* 2006;107:4650-7.
11. van Buren N, Bonnet MC, Dréno B, et al. Tumor and immunologic response after vaccination of melanoma patients with an ALVAC virus encoding MAGE antigens recognized by T cells. *J Clin Oncol* 2005;23:9008-21.
12. Kruit WHJ, van Ojik HH, Brichard VG, et al. Phase 1/2 study of subcutaneous and intradermal immunization with a recombinant MAGE-3 protein in patients with detectable metastatic melanoma. *Int J Cancer* 2005;117:596-604.
13. Di Giacomo AM, Danielli R, Guidoboni M, et al. Therapeutic efficacy of ipilimumab, an anti-CTLA-4 monoclonal antibody, in patients with metastatic melanoma unresponsive to prior systemic treatments: clinical and immunological evidence from three patient cases. *Cancer Immunol Immunother* 2009;58:1297-306.
14. Hoos A, Parmiani G, Hege K, et al. A clinical development paradigm for cancer vaccines and related biologics. *J Immunother* 2007;30:1-15.
15. Melero I, Hervas-Stubb S, Glennie M, et al. Immunostimulatory monoclonal antibodies for cancer therapy. *Nat Rev Cancer* 2007;7:95-106.
16. Wolchok JD, Saenger Y. The mechanism of anti-CTLA-4 activity and the negative regulation of T-cell activation. *Oncologist* 2008;13:2-9.
17. O'Day SJ, Ibrahim R, DePril V, et al. Efficacy and safety of ipilimumab induction and maintenance dosing in patients with advanced melanoma who progressed on one or more prior therapies [abstract 9021]. *J Clin Oncol* 2008;26.
18. Hamid O, Chin K, Li J, et al. Dose effect of ipilimumab in patients with advanced melanoma: results from a phase II, randomized, dose-ranging study [abstract 9025]. *J Clin Oncol* 2008;26.
19. Weber JS, Berman D, Siegel J, et al. Safety and efficacy of ipilimumab with or without prophylactic budesonide in treatment-naïve and previously treated patients with advanced melanoma [abstract 9010]. *J Clin Oncol* 2008;26.
20. Hodi FS, Hoos A, Ibrahim R, et al. Novel efficacy criteria for antitumor activity to immunotherapy using the example of ipilimumab, an anti-CTLA-4 monoclonal antibody [abstract 3008]. *J Clin Oncol* 2008;26.
21. Wolchok JD, Ibrahim R, DePril V, et al. Antitumor response and new lesions in advanced melanoma patients on ipilimumab treatment [abstract 3020]. *J Clin Oncol* 2008;26.
22. Weber JS, Hersh EM, Yellin M, et al. The efficacy and safety of ipilimumab (MDX-010) in patients with unresectable stage III or stage IV malignant melanoma [abstract 8523]. *J Clin Oncol* 2007;25.
23. Hamid O, Urba WJ, Yellin M, et al. Kinetics of

- response to ipilimumab (MDX-010) in patients with stage III/IV melanoma [abstract 8525]. *J Clin Oncol* 2007;25.
24. Attia P, Phan GQ, Maker AV, et al. Autoimmunity correlates with tumor regression in patients with metastatic melanoma treated with anti-cytotoxic T-lymphocyte antigen-4. *J Clin Oncol* 2005;23:6043-53.
25. Maker AV, Phan GQ, Attia P, et al. Tumor regression and autoimmunity in patients treated with cytotoxic T lymphocyte-associated antigen 4 blockade and interleukin 2: a phase I/II study. *Ann Surg Oncol* 2005;12:1005-16.
26. Weber JS, O'Day S, Urban W, et al. Phase I/II study of ipilimumab for patients with metastatic melanoma. *J Clin Oncol* 2008;26:5950-6.
27. Hodi FS, Butler M, Oble DA, et al. Immunologic and clinical effects of antibody blockade of cytotoxic T lymphocyte-associated antigen 4 in previously vaccinated cancer patients. *Proc Natl Acad Sci U S A* 2008;105:3005-10.
28. Hodi FS, Oble DA, Drappatz J, et al. CTLA-4 blockade with ipilimumab induces significant clinical benefit in a female with melanoma metastases to the CNS. *Nat Clin Pract Oncol* 2008;5:557-61.
29. Tuma RS. New response criteria proposed for immunotherapies. *J Natl Cancer Inst* 2008;100:1280-1.