Exploration of a Novel Intermediate Response Endpoint in Immunotherapy Clinical Studies
Xin Gao, Lijun Zhang, and Rajeshwari Sridhara

Abstract

Purpose: Both objective response rate (ORR) and progression-free survival as defined by RECIST are weakly associated with overall survival (OS) in trials evaluating immunotherapy drug products. We proposed a novel intermediate response endpoint (IME) for evaluating immunotherapies.

Experimental Design: We defined IME response as having no nontarget lesion progression, no new lesion appearance, and reaching a target lesion response determined by baseline tumor burden, tumor reduction depth, and tumor change dynamics within one year after randomization. Database used consisted of data from randomized active-controlled immunotherapy trials. Criterion for IME was developed on the basis of patient-level data from a training dataset, and further evaluated using an independent testing dataset. A patient-level responder analysis comparing OS between patients with and without an IME response was conducted using combined data. Association between trial-level OS hazard ratio (HR) and IME odds ratio (OR) was analyzed using a weighted linear regression model.

Results: A total of 5,806 patients from 9 randomized studies were included in the database. At patient level, patients with IME response had improved OS compared with nonresponders (HR = 0.69). At trial level, association between OS and IME was moderate ($R^2 = 0.63$).

Conclusions: The IME was moderately associated with OS, and the association appeared to be stronger than the association observed between RECIST-defined ORR and OS. However, the analyses conducted in this research are exploratory and further evaluation is needed before using this endpoint in future studies.

Clin Cancer Res; 24(10); 2262–7. ©2017 AACR.

Introduction

Immunotherapies have been developed for many cancers in recent years, generating excitement in the cancer community. Since 2014, the FDA has approved immune checkpoint inhibitors (anti–PD-1 and anti–PD-L1 antibodies) for the treatment of patients with advanced melanoma (1–3), metastatic non–small cell lung cancer (4–7), head and neck squamous cell carcinoma (8), urothelial carcinoma (9, 10), renal cell carcinoma (11), classical Hodgkin lymphoma (12), Merkel cell carcinoma (13), and microsatellite instability–high cancers (14). However, from the experience with immune checkpoint inhibitors (anti–PD-1 and anti–PD-L1 antibodies), we learnt that patients treated with immunotherapies display unique patterns of antitumor response, due to the indirect effect of immunotherapies on the immune system rather than on tumors directly. For example, delayed tumor shrinkage is sometimes observed in patients treated with immunotherapies after an initial increase in tumor size.

Traditional endpoints based on imaging data following RECIST criteria, such as progression-free survival (PFS) and overall response rate (ORR), are widely accepted endpoints in trials of solid tumors. However, with the novel response patterns observed in trials of immunotherapies, these traditional RECIST-based endpoints could be problematic when used to evaluate treatment effects. In particular, overall survival (OS) has been used as the primary endpoint in most trials submitted to the FDA as either initial or supplemental Biologics License Applications between years 2014 and 2016. Randomized, multicenter, and active-controlled studies to evaluate checkpoint inhibitors using OS as either the sole primary endpoint or one of the co-primary endpoints, and with adequate follow-up time...
Novel Intermediate Endpoint in Immunotherapy Studies

Translational Relevance

Both objective response rate (ORR) and progression-free survival as defined by RECIST are weakly associated with overall survival (OS) in trials evaluating immunotherapy drug products. We proposed a novel intermediate response endpoint (IME) for evaluating immunotherapies. The IME response was defined on the basis of nontarget lesion progression, new lesion, and target lesion information determined by baseline tumor burden, tumor reduction depth, and tumor change dynamics within one year after randomization.

Database used consisted of data from nine randomized active-controlled immunotherapy trials. A patient-level responder analysis comparing OS between patients with and without an IME response showed that IME responders had improved OS compared with nonresponders. Association between trial-level OS HR and IME OR showed that the association between OS and IME was moderate, and the association appeared to be stronger than the association observed between RECIST-defined ORR and OS.

The candidate IME was defined on the basis of radiographic tumor measurement data similar to RECIST as a composite endpoint consisting of three components: target lesion, nontarget lesion, and new lesion.

For the target lesion component of the candidate IME, we considered baseline tumor burden, post-baseline tumor reduction depth, and post-baseline tumor change dynamics. Baseline tumor burden was represented by the baseline sum of the longest dimension (SLD) of target lesions. Post-baseline tumor reduction depth was represented by the nadir of target lesion SLD change compared to baseline within 1-year post-randomization. Post-baseline tumor change dynamics were represented by the area between baseline and percentage change curves (AUC) of target lesion SLD within 1-year post-randomization. Figure 1 illustrates how the AUC was calculated for post-baseline tumor change dynamics. The area below the x-axis (reduction from baseline) was calculated as a negative value, whereas the area above the x-axis (increase from baseline) was calculated as a positive value. For the observed tumor measurement data on one patient shown in Fig. 1, overall AUC within 1 year was $-3,623 + 2,018 = -1,605$.

We fit the target lesion data (baseline SLD, nadir, and AUC as defined above) into a Cox regression model for OS using the training dataset. Using the regression coefficients determined from this Cox model as weights, we calculated a patient’s target lesion score as the exponential of the weighted sum of these three factors.

Using the target lesion score, we defined each patient’s target lesion response status by dichotomizing this score into either target lesion responder or target lesion nonresponder. The optimal cutoff to determine target lesion response status was chosen as the value that yielded the smallest hazard ratio (HR) of OS comparing responders with nonresponders based on the training dataset whereas a reasonable number of patients were included in each category. A patient was defined as a target lesion responder when the patient’s target lesion score was smaller than the identified optimal cutoff value.

Nontarget lesion and new lesion status as per RECIST 1.1 criteria (16) were also considered in the composite endpoint, IME. The proposed IME was defined as a binary endpoint: response or nonresponse. An IME response was defined as satisfying all of the three following criteria (Fig. 2).

1. The patient needed to be a target lesion responder, meaning the patient’s target lesion score was less than the optimal cutoff value; and
2. The patient had no unequivocal nontarget lesion progression as determined per RECIST 1.1 criteria within 1 year; and
3. The patient had no new unequivocal lesion as determined per RECIST 1.1 criteria within 1 year.

Figure 1.
Illustration of determination of tumor change dynamics.

Statistical analysis

The initial development of the IME was based on the training dataset only. The same response criteria were applied to the independent testing dataset. Survival per IME status was summarized by Kaplan–Meier plot and Cox Proportional Hazards (PH) model for the training dataset and testing dataset separately.
**Patient-level responder analysis.** A responder analysis was performed to compare OS between IME responders and IME nonresponders, irrespective of treatment assignment, using the combined dataset with trial as a stratification factor. We estimated HRs of OS from the Cox proportional hazards (PH) model stratified by trial and obtained Kaplan–Meier estimates of OS by IME response status.

**Trial-level analysis.** The association between treatment effects on IME and OS was evaluated using a weighted linear regression model with analyses performed on a logarithmic scale and weights equal to the sample size of each randomized comparison. The coefficient of determination ($R^2$) and the associated 95% confidence intervals (CI) from the weighted linear regression model were used to measure the association between treatment effects on IME and OS. Treatment effects on OS were presented as HRs estimated from Cox PH models and treatment effects on IME were presented as odds ratios (OR) estimated from logistic regression models.

**Results**

We identified 9 randomized immunotherapy (checkpoint inhibitors; anti–PD-1 and anti–PD-L1 antibodies) trials (N = 5,806) submitted to the FDA between 2014 and 2016 in support of initial or supplemental Biologics License Applications for advanced or metastatic solid tumors (Table 1). Each of the 4 trials with 3 treatment arms was considered as 2 randomized comparisons; therefore, 13 randomized comparisons were included in the trial-level analysis. Among these 9 randomized trials, 4 trials were indicated for melanoma, 3 trials were indicated for nonsmall-cell lung cancer (NSCLC), and the remaining 2 trials were indicated for renal cell carcinoma (RCC) and head and neck cancer (HNC), respectively.

The aggregated key baseline patient demographics and disease characteristics of these 9 trials were summarized (Supplementary Data). Over half of the patients were younger than 65-years-old, and 65% of patients were male. Majority of the patients were White (90%), and only 1% of patients were Black. The patient characteristics were considered homogeneous, and data from the 9 trials were combined for analysis.

Trials 1 to 4 (N = 2810) were randomly selected to be included in the training dataset with different indications and immunotherapies. The remaining trials 5 to 9 (N = 2996) were used as an independent testing dataset in the analyses.

As described in Materials and Methods section above, regression coefficients β1, β2, and β3 were obtained for baseline SLD, nadir and AUC, respectively (β1 = 0.006, β2 = 0.01, and β3 = 0.0001). These regression coefficients were used to calculate each patient's target lesion score.

To explore an optimal cutoff value to dichotomize a patient into a target lesion responder or a nonresponder on the basis of the target lesion score, we evaluated different cutoffs (ranging from 0.05 to 5). HRs obtained from the analyses versus various cutoff values were plotted (Supplementary Data). The results showed that HR was smallest with the cutoff value of 0.7. However, using the cutoff value of 0.7, the number of patients classified as target lesion responders was minimal (approximately 10%). On the basis of the HR estimates, and with a reasonable number of patients included in each group, we chose the optimal cutoff value as 1. A patient with the target lesion score less than 1 was considered a target lesion responder, or was considered a nonresponder otherwise.

Using the chosen cutoff value of one, 554 patients in the training dataset were target lesion responders. Of the 554 patients, 165 patients had unequivocal nontarget lesion progression or an unequivocal new lesion within a year. According to the IME definition criteria described in Materials and Methods section above, 389 (14%) patients were IME responders in the training dataset. Applying the same IME criteria to the independent testing dataset, 541 (18%) patients were IME responders (Fig. 3).

As shown in Fig. 4, IME responders had longer survival (HR, 0.10; 95% CI, 0.07–0.14) compared with nonresponders in the training dataset, and a similar association was observed in the independent testing dataset (HR, 0.08; 95% CI, 0.05–0.11).

---

**Table 1. Summary of trials included**

<table>
<thead>
<tr>
<th>Cancer types</th>
<th>Primary endpoints</th>
<th>Number of arms</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>NSCLC</td>
<td>2</td>
<td>582</td>
</tr>
<tr>
<td>Study 2</td>
<td>HNC</td>
<td>2</td>
<td>361</td>
</tr>
<tr>
<td>Study 3</td>
<td>Melanoma</td>
<td>2</td>
<td>1033</td>
</tr>
<tr>
<td>Study 4</td>
<td>NSCLC</td>
<td>3</td>
<td>854</td>
</tr>
<tr>
<td>Study 5</td>
<td>NSCLC</td>
<td>3</td>
<td>272</td>
</tr>
<tr>
<td>Study 6</td>
<td>RCC</td>
<td>2</td>
<td>821</td>
</tr>
<tr>
<td>Study 7</td>
<td>Melanoma</td>
<td>3</td>
<td>418</td>
</tr>
<tr>
<td>Study 8</td>
<td>Melanoma</td>
<td>3</td>
<td>945</td>
</tr>
<tr>
<td>Study 9</td>
<td>Melanoma</td>
<td>3</td>
<td>540</td>
</tr>
</tbody>
</table>

Abbreviations: HNC, head and neck cancer; NSCLC, non-small cell lung cancer; RCC, renal cell carcinoma.
Examining the overall combined dataset, including all 9 trials, the HR of OS was 0.09 (95% CI, 0.07–0.11).

Figure 5 showed a scatterplot of the treatment effects on the log-scale, illustrating trial-level association between IME and OS based on the combined dataset. As shown in this scatterplot, IME and OS were moderately associated ($R^2 = 0.68$; 95% CI, 0.02–0.91) at the trial level. A weaker trial-level association was observed when IME were evaluated on the basis of shorter observation time ($R^2 = 0.54$ for 9 months after randomization).

Summary

Patients treated with immunotherapy agents (e.g., anti–PD-1 and anti–PD-L1) have shown unique patterns of response, such as delayed decrease in tumor size after an initial increase and prolonged duration of response. Therefore, traditional tumor measurement-based response criteria, RECIST, may not be able to capture the delayed benefit of immunotherapies and the persistence of responses for a long time. Furthermore, the area under the curve reflects both the depth and duration of tumor size reduction. Baseline tumor burden and tumor reduction nadir (depth of response) were reported as being correlated with long term clinical benefit by others (17–19). An intermediate endpoint that includes these characteristics, given that the objective response rate as defined by RECIST 1.1 is weakly associated with OS ($R^2 = 0.13$; ref. 15), may be a potentially better surrogate marker for OS.

We developed and evaluated a radiographic tumor measurement data-based IME, which considered three components, information of target lesion, nontarget lesion, and new lesion within the first-year post randomization. Target lesion information was summarized using baseline tumor burden, post-baseline tumor reduction depth, and post-baseline tumor change dynamics. The preliminary analyses showed that, at the patient level, a patient having IME response had a more favorable survival outcome compared to a patient having IME nonresponse. At the trial level, the preliminary results showed a moderate association with OS.
There are several limitations of the IME that has been proposed here. First, unlike the traditional RECIST-based endpoints, the proposed IME can only be determined after a pre-specified period of time, for example, 1-year post randomization, because information within the first year is needed to determine the IME. Therefore, the proposed IME is a potentially more useful endpoint in the regulatory setting rather than during real-time monitoring of patient response status. Second, for the proposed IME, a patient with new lesion(s) in the first year was considered a nonresponder. However, as observed from several clinical trials of immunotherapies, patients could have further clinical benefit after the appearance of new lesions. Therefore, some researchers suggested that the appearance of new lesions should no longer be counted as definitive indication of disease progression, and instead should be included in the change of total tumor burden (20). Among the 1,309 patients with target lesion response as defined in this paper, 219 (17%) were classified as IME nonresponders due to new lesion appearance only (Fig. 3). We further explored a modified IME in which the new lesion is not considered in determination of IME response. The association analyses between the modified IME and OS showed similar results to those reported for the proposed IME in this article at both the trial level and the patient level, which indicated that new lesions had minimum impact on the association analyses. Third, the trials included in the combined database were indicated for various solid tumor types, including NSCLC, melanoma, RCC, and HNC. Given the differences in the natural history of these diseases and disease growth rates, the threshold for the IME may vary from disease-to-disease. These data were combined together in this research work notwithstanding the differences, so that a number of trials and comparisons were available for the association analyses. Finally, most of the trials were designed to compare an immunotherapy with chemotherapy. This may affect the evaluation of the association between IME and OS in future clinical trials with different comparative treatments, such as a targeted therapy or another immunotherapy.

In conclusion, we developed a novel IME for clinical trials with immunotherapy products. Under the current definition, the IME had a moderate trial-level association with OS, and this association appeared to be stronger than the association observed between RECIST-defined ORR and OS. However, the analyses conducted in this research are exploratory, and further evaluation needs to be performed before using this endpoint in future studies.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

This article reflects the views of the authors and must not be construed to represent FDA’s views or policies.

Authors’ Contributions

Conception and design: X. Gao, L. Zhang, R. Sridhara
Development of methodology: X. Gao, L. Zhang, R. Sridhara
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): X. Gao, L. Zhang
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): X. Gao, L. Zhang, R. Sridhara
Writing, review, and/or revision of the manuscript: X. Gao, L. Zhang, R. Sridhara
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): X. Gao, L. Zhang, R. Sridhara
Study supervision: R. Sridhara

Received July 21, 2017; revised October 24, 2017; accepted December 12, 2017; published first December 26, 2017.

References


Clinical Cancer Research

Exploration of a Novel Intermediate Response Endpoint in Immunotherapy Clinical Studies

Xin Gao, Lijun Zhang and Rajeshwari Sridhara