Evolution of end points for cancer immunotherapy trials

A. Hoos*

Bristol-Myers Squibb, Global Clinical Research, Wallingford, USA†

The effect of cancer immunotherapies is on the immune system and not directly on the tumour. The kinetics of immunotherapy are characterised by a cellular immune response followed by potential changes in tumour burden or patient survival. To adequately investigate immunotherapies in clinical trials, a new development paradigm including reconsideration of established end points addressing this biology is needed. Over the last 7 years, several initiatives across the cancer immunotherapy community were facilitated by the Cancer Research Institute Cancer Immunotherapy Consortium. They systematically evolved an immunotherapy-focused clinical development paradigm and proposed to redefine trial end points. On that basis, analysis of several large datasets generated throughout the immunotherapy community supports three novel end point proposals. First, results from T-cell immune response assays are highly variable and often nonreproducible. Harmonisation of assays can minimise this variability and support the investigation of the cellular immune response as a biomarker and testing it for clinical surrogacy. Secondly, immunotherapy induces novel patterns of the antitumour response not captured by World Health Organisation criteria or Response Evaluation Criteria in Solid Tumours. New immune-related response criteria were defined which more comprehensively capture all response patterns. Thirdly, survival curves in randomised immunotherapy trials can show a delayed separation, which can impact study results. Altered statistical models are needed to describe the hazard ratios as a function of time, and differentiate them before and after separation of curves to improve planning of phase III trials. Taken together, these recommendations may improve our tools for cancer immunotherapy investigations.

Key words: assay harmonisation, hazard ratios, immune system, immunotherapy, survival analysis, tumour response

introduction

There is widespread recognition that using the patient’s own immune system to target and destroy cancer cells may offer an effective therapy with less toxicity than conventional chemotherapy. The cancer immunotherapy field has been recognised for over 10 decades, but the clinical success of immunotherapy has been limited by challenges associated with integrating the knowledge gained from methodological advances to ensure the clinical success. The failure of many cancer immunotherapy programmes can be attributed to three reasons: (i) ineffective products; (ii) insufficient understanding of the mechanism of action of the immune system and (iii) inadequate methodologies, which have prevented the accurate interpretation of immunotherapy mechanisms in patients [1, 2].

Although there have been many advances in our understanding of the immune response in cancer, the challenge arises in the ability to communicate and share the learning to establish best working practice. To address this challenge, the dissemination of collective learning within the field of cancer immunotherapy has, over the last 8 years, been largely facilitated by the Cancer Immunotherapy Consortium (CIC; formerly Cancer Vaccine Consortium), a programme of the nonprofit Cancer Research Institute in the United States, and by the Association for Cancer Immunotherapy (CIMT; Mainz, Germany) in Europe. This collaboration of academic, nonprofit, biotech and pharmaceutical sectors has resulted in the creation of a methodological framework that aims to define a better path for the development of new therapies and to inform future practitioners in the field, thereby laying the foundations for reproducible success in the development of cancer immunotherapies. Ultimately, this collaborative approach has resulted in a shift in the research and development paradigm from conventional chemotherapy towards immunotherapy.

The clinical kinetics of immunotherapy can be considered as three distinct, progressive clinical end points. First, the initiation of treatment leads to activation of the immune system, which results in a cellular immune response. Secondly, an antitumour response becomes evident weeks to months following the start of immunotherapy. Thirdly, after several months, there is an impact on patients’ overall survival. Each of these three distinct clinical end points is associated with their own challenges that have been the subject of much discussion within the immunology community in recent years (Table 1). The outcomes of these discussions are described below.

immune biomarker development: immune assay harmonisation

The first biological event to occur following treatment with an immunotherapeutic agent is activation of the immune system.

*Correspondence to: A. Hoos, Bristol-Myers Squibb, Global Clinical Research, Wallingford, CT 06492, USA. E-mail: axel.hoos@blackberry.com

†At time of manuscript preparation.
Hence, the measurement of the immune response (T-cell or antibody response) is an attractive ‘proof of principle’ biomarker that could be assessed before clinical end points are reached. Immunological biomarkers have broad applications, from determining whether an immune intervention has achieved its biological effect to predicting clinical outcomes as surrogates for clinical benefit. The prerequisite for effective biomarker measurement through immune-monitoring assays is the use of reliable and reproducible assays. T-cell immune response assays are notorious for their high degree of variability. As a result, they are often nonreproducible, which has implications for clinical trials in which assays must be used across multiple centres [3]. Unfortunately, this has contributed to the field’s failure to accurately correlate immunological biomarker findings with clinical end points such as survival or response [4]. Therefore, harmonisation of assays is critical to reduce this variability and support the investigation of cellular immune responses as both a clinical biomarker and for clinical surrogacy.

Numerous T-cell immune response assays exist; the most commonly used are the enzyme-linked immunosorbent spot (ELISPOT), intracellular cytokine staining (ICS) and human leukocyte antigen (HLA)-peptide multimer staining assays [5–7]. Although they are often described as classical assays, existing data clearly indicate that to maintain consistency, they need to be performed by specially trained laboratory staff [6–8]. The extent of assay variability was highlighted in a report by Janetzki et al. [8]. For all three immune-assays, the degree of variability between laboratories was striking (Figure 1).

Overcoming the technical hurdle of inter-laboratory assay variability would improve the accuracy of the results obtained, thereby adding value to their use within clinical trials [8, 9]. In response to this predicament, a series of international proficiency panels (quality control experiments across multiple centers) were initiated by CIC and CIMT in 2005 [10–12]. Over 100 participating laboratories from 14 countries, the US Department of Defense and the German regulatory agency Paul-Ehrlich-Institute (Langen, Germany) assessed the ELISPOT, ICS and HLA-multimer staining assays. The goals of this assessment were to provide an external quality assurance process for laboratories conducting immune monitoring in clinical trials, and to harmonise assay performance.

The reduction of variability observed when assay harmonisation criteria were applied demonstrates the requirement for consistency between laboratories when performing the same assay [11, 12]. In the case of ELISPOT, assay harmonisation reduced variability from 50% to <10%, which is within the realm of acceptability required for assays used in clinical trials. This important finding led to the preparation of preliminary ELISPOT harmonisation guidelines, which were published in 2008 [12]. The application of harmonisation does not require the standardisation of assay protocols, which could adversely affect the scientific creativity needed for new assay development. Instead, the process involves standard operating procedure-based improvements to identify and remove factors that can introduce major assay variability between laboratories.

### Table 1. Challenges and recommendations for assessment of cancer immunotherapy [4]

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Cellular immune response</th>
<th>Antitumour response</th>
<th>Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex assays exist</td>
<td>Conventional and novel response patterns are observed</td>
<td>Translation of immune and antitumour response into a survival effect takes time</td>
<td></td>
</tr>
<tr>
<td>Results are highly variable and not reproducible across trials</td>
<td>Translation of the immune response into an antitumour response takes time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No systematic criteria exist to capture new response patterns</td>
<td>Proportional hazards assumptions are not applicable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assay procedures are not harmonised</th>
<th>No systematic criteria exist to capture new response patterns</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Recommendations</th>
<th>Harmonise assay use through standard operating procedures that accompany individual assay protocols</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Identify relevant response patterns</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Use systematic immune-related response criteria to reproducibly capture new patterns</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Employ statistical models that account for the delayed effect</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Consider use of early interim and futility analyses</th>
</tr>
</thead>
</table>

**Figure 1.** Challenges of immune monitoring: inter-laboratory variability in immune response measurements [8]. Reprinted from Janetzki et al. [8], with permission from Elsevier.
A successful model for improving clinical development procedures has previously been established, namely the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use—Good Clinical Practice (ICH-GCP) [13]. The creation of a similar framework to increase the quality of immunological monitoring is warranted. Such a framework would enable the use of assay results to guide the clinical development of new immunotherapies and provide a better understanding of exact therapeutic mechanisms of action. Ultimately, the successful management of data variability between assays used for immune monitoring would contribute to methodological validation and clinical qualification in biomarker development.

antitumour response

Following treatment with immunotherapy, changes in tumour burden may be the consequence of three scenarios: (i) in patients that have an objective response, the tumour is invaded by immune cells and inflammatory cells and metastatic lesions decrease in size; (ii) in patients with tumour progression, metastatic lesions increase in size and (iii) in patients where tumour lesions become heavily infiltrated by immune and inflammatory cells, there may be an initial increase in tumour burden. In the case of the latter scenario, the increased burden is an indicator of an antitumour response, but would be defined as progressive disease according to existing Response Evaluation Criteria in Solid Tumours (RECIST) or World Health Organisation (WHO) criteria.

In 2008, the Cancer Vaccine Clinical Trial Working Group devised a set of principles that could be used to describe the patterns of antitumour response associated with immunotherapy [14]. Essentially, an allowance could be made for the continuation of treatment in cases of early and clinically insignificant progression. Responses in patients who experienced tumour regression after initial tumour progression were determined according to the largest tumour volume measured after the start of treatment and not necessarily from the baseline tumour volume [14]. These principles were applied prospectively to the ipilimumab development programmes. Four distinct patterns of response were observed (Figure 2) [15]: (i) a complete response, with an immediate reduction in tumour burden; (ii) stable disease, with a slow and steady decline in tumour burden; (iii) an initial increase in tumour volume over time, followed by a response; (iv) the development of new lesions during treatment, resulting in an initial increase in tumour burden that was followed by a reduction. The first two patterns of response are typical of those observed in patients receiving chemotherapy and are captured by the existing RECIST and WHO response criteria. By contrast, the other patterns of response are novel and specific to immunotherapy. Using the RECIST or WHO criteria to assess these responses is not appropriate, as they

![Figure 2. Novel response patterns observed with immunotherapies [4, 15]. Reprinted by permission from the American Association for Cancer Research: Wolchok et al. [15].](https://example.com/figure2.png)
would be defined as progressive disease and, as a result, treatment may be prematurely discontinued.

To accommodate these patterns of response associated with immunotherapy, a new set of response criteria was proposed: the immune-related response criteria (irRC). In essence, the irRC have the same characteristics as the RECIST and WHO criteria, but they also capture comprehensively all four response patterns [4, 15]. The importance of using the new irRC to assess tumour responses can be illustrated by comparing the effect of the WHO criteria and the irRC on patient overall survival. Figure 3 shows how the using the irRC can identify survivors within the population of patients who would be classified as having progressive disease if the WHO criteria are used. In this figure, the survival outcome is consistently poorest in patients with progressive disease as assessed by both the RECIST and WHO criteria. In patients with responses or stable disease classified by the conventional WHO response criteria, the survival outcomes are good. Patients with nonconventional responses classified using the irRC have survival outcomes that are comparable to those in patients with conventional responses identified using the existing RECIST and WHO criteria.

In summary, the new irRC provide a much needed tool for the accurate identification of patients with tumours that are responsive to immunotherapy in a nonconventional way. For practising oncologists, one of the most important changes to conventional practice is the recording of new lesions within the existing lesions profile to establish the total tumour burden. The total tumour burden should be assessed at numerous time points, rather than at a single point. This will ensure that tumours that initially increase in size due to an immunological response before shrinking are not mistaken for progressive disease.

**overall survival**

A delayed separation of Kaplan–Meier survival curves is observed in almost all randomised immunotherapy trials, and may occur months after the start of treatment. This is in contrast to the Kaplan–Meier survival curves obtained in clinical trials of chemotherapy, where early clinical effects are achievable [4, 16]. The delayed separation of survival curves associated with immunotherapy reduces the ability for statistical power to differentiate between them [16, 17]. Clinical trials that have a delayed response deviate from the standard model used for randomised trials that assume a proportional hazard, i.e. that no events occur before the separation of curves and that the hazard is constant over time [18].

The delayed separation of curves has been reported in many clinical trials of immunotherapeutic agents, including a placebo-controlled phase III trial of the autologous active cellular immunotherapy sipuleucel-T, where the immunotherapeutic effect on survival was not evident for 8 months [19], and the first phase III trial of ipilimumab, where separation of the survival curves for patients treated with ipilimumab alone, ipilimumab plus gp100 vaccine or gp100 vaccine alone, was not observed until 4 months after the initiation of treatment [20]. In both of these examples, although immunotherapeutic activity occurred before the separation of the Kaplan–Meier curves, it did not translate into a survival difference between the curves or agents. As such, any event that occurs before the separation of the curves is a 'lost' event in terms of the final analysis, as it does not contribute to the effect that follows the separation of the curves.

It is clear that the delayed response, which is often substantial, has a huge impact on the dynamics of the trial. To describe the implications of the delayed separation of Kaplan–Meier curves, a model scenario can be applied. The curves are separated into two components: component 1 is no separation of the curves (i.e. no difference), where the hazard ratio (HR) is equal to 1, and component 2 is a separation of curves, which requires a large delta value to power a statistically significant difference between the two curves (Figure 4). The large delta value is essential in the second component of the curve to compensate for the lack of separation during the first

![Figure 3](image-url) Immune-related response criteria identify survivors among 227 patients enrolled in phase II studies of ipilimumab 10 mg/kg monotherapy that would have had progressive disease according to modified World Health Organisation criteria [15]. Reprinted by permission from the American Association for Cancer Research: Wolchok et al. [15].
component of the curve. For example, consider a trial of 800 patients; if 200 events occur during the first component of the curve (i.e. where HR: 1), then these are in effect lost events due of the nature of the proportional hazard assumptions. This is important because, for example, 600 events were required for the final analysis of survival, but only 400 events were available, the final analysis would be under powered. In the event that a treatment effect was observed, it may not be statistically significant if the power of the study is only 50%. This was not the case for the two phase III trials mentioned earlier, as these had strong treatment effect responses that overcame the delay in responses. Conversely, there are examples where the treatment effect responses were not as strong and the trial had a negative result as a consequence. In a phase III trial of tremelimumab, for example, an early interim analysis for survival showed no survival benefit, and as a result, the study was terminated. An extended follow-up, however, revealed an eventual separation of the survival curves [3].

Immunotherapeutic agents may differ with respect to the presence and timing of delayed separation of the survival curves. Therefore, exploiting our current knowledge to apply adequate statistical modelling to describe HRs as a function of time, and to differentiate them before and after the separation of curves (ideally in randomised phase II trials), may improve our ability to plan for pivotal phase III trials.

**role of regulatory authorities in development process**

The acceptance of new methodologies, such as those described above, by the regulatory authorities is a prerequisite for inclusion into credible drug development programmes. To facilitate this, regulatory authorities were invited to participate in CIC workshops that provided forums where the new proposals being presented by the immunotherapy community could be openly communicated and discussed. As a result, the Food and Drug Administration produced a regulatory guidance document in 2009 on the Clinical Considerations for Therapeutic Cancer Vaccines [21], which included many topics of the current discussion. The guidance underwent public consultation and is currently being finalised. Expectation is that the guidance document will support prospective immunotherapy trials using the newly described methodologies. In line with this, the European Medicines Agencies (EMA) released a concept paper in 2010 to request public feedback for revision of its guidance on ‘evaluation of anticancer medicinal products in man’ [22]. The focus of this paper is on clinical end points for biological therapies, and it includes a section on cancer vaccines. The EMA has now received feedback from CIC and CIMT and is in the process of providing further guidance to the immunotherapy community. The continued dialogue between the regulatory authorities and community-based associations will no doubt strengthen the expansion of regulatory guidance to better serve the development of new immunotherapeutic agents.

**summary**

The field of immuno-oncology has advanced over the years through the facilitation of collaborative forums among stakeholders (via the CIC and CIMT). These forums have encouraged greater communication and awareness of the unique characteristics associated with immunotherapeutic agents. They have also led to the creation of a tailored methodological framework for the development of immunotherapy that is distinct from that widely used for the development of chemotherapeutic agents. This framework defines a better path for the development of new immunotherapies and creates the foundation for a clinical subspecialty of immuno-oncology within the universal oncology discipline. The credibility of the immuno-oncology field has the potential to be strengthened by the framework, which offers new tools, development principles and structure.

Overall, the outlook for immuno-oncology has improved significantly as a result of the development of this framework; conversely, the framework itself will no doubt continue to be expanded and refined in parallel with advances in this field. Importantly, the framework has assisted the field of immuno-oncology to establish itself within the broader field of oncology. This brings an opportunity for all treating oncologists to be exposed to both the use of immunotherapeutic agents, and to the value of using more appropriate clinical end points to establish the efficacy of this class of therapy when administered to patients with cancer.

**disclosures**

At the time of manuscript preparation, the author was an employee of and held stock in Bristol-Myers Squibb.

**references**