Immunotherapies in Clinical Trials: Do They Demand Different Evaluation Tools?

By Rabiya S. Tuma

Melanoma patients and immunotherapy researchers alike had cause to celebrate in late March when the U.S. Food and Drug Administration approved ipilimumab (Yervoy) to treat metastatic melanoma. For patients, the drug is a much-needed alternative to the commonly used dacarbazine (DTIC-Dome), the first agent shown to prolong overall survival. For immunotherapy researchers, the drug, which works by binding to a receptor on the surface of T lymphocytes and releasing the brakes on the immune system, represents one more piece of evidence that they can recruit a patient’s immune system to fight cancer.

Despite ipilimumab’s success, a key question in the field remains how best to measure disease response in patients treated with immunotherapies. No one will argue with the statistically significant increase in overall survival that occurred in the pivotal ipilimumab trial (10 months in patients treated with ipilimumab with or without an experimental vaccine vs. 6.5 months for patients treated with the experimental vaccine alone). However, many researchers are concerned that response and progression-free survival criteria developed in the era of cytotoxic therapies are inappropriate measures for immunotherapeutic agents and may lead to premature closures of trials and, perhaps, loss of active therapies.

Last fall, an international group of academic and industry researchers, the Cancer Immunotherapy Consortium (CIC), proposed modified criteria for measuring response to drugs that work through the immune system. Many cancer researchers see value in amended criteria and are currently testing them prospectively, but not everyone is convinced they are necessary.

“The unique kinetics we see with immunotherapies really challenge the endpoints that we usually use for cancer clinical trials,” said Jedd Wolchok, M.D., Ph.D., a medical oncologist at Memorial Sloan-Kettering Cancer Center in New York and an associate director of the CIC. “The surrogate endpoints that we’ve become so dependent on—response rate and progression-free survival—just don’t tell the whole story.”

Delayed Response Pattern

The CIC scientists point out that whereas cytotoxic agents attack cancer cells directly and cause rapid tumor shrinkage in responsive patients, immunotherapies such as ipilimumab or the prostate cancer vaccine sipuleucel-T (Provenge), activate the patient’s immune system, which in turn attacks the tumor. That indirect effect means there may be a delayed tumor response to treatment. In fact, between 10% and 25% of the patients who participated in phase II ipilimumab trials and ultimately responded to the drug, initially showed tumor growth or had new lesions before their tumors shrank. By conventional response criteria, such as RECIST or those of the World Health Organization, those patients would be labeled non-responders and taken off the drug, even though they benefited from the agent over time. And that sort of misleading initial lack of response could cause an active drug to fail in clinical testing, Wolchok and others contend.

To avoid such problems and more accurately represent what happens when patients receive immunotherapies, CIC members have proposed amended endpoints for clinical trials, which they published last year (see J. Natl. Cancer Inst. 2010;102:1388–97). In brief, the new criteria recommend that researchers develop tools to minimize the difference between assays that measure T-cell immune responses; that evaluation criteria incorporate new clinical response patterns; and that new statistical models be considered to assess survival, or more specifically, Kaplan–Meier survival curves.

The second recommendation—new ways to measure a tumor’s response to treatment—refers to what the authors call immune-related response criteria (irRC). These guidelines resemble standard tumor shrinkage criteria but allow for some tumor growth with or without the formation of new lesions before tumor shrinkage. Instead of immediately labeling patients with early tumor growth as non-responders, the irRC recommend reevaluation and reimaging of these patients after 4–6 weeks. Physicians would consider the patient a non-responder only if that second imaging test also showed an increased tumor burden. In other words, non-responses need to be confirmed at two consecutive time points, just as for confirmation of response or stable disease under conventional response criteria.

No one has yet prospectively validated the proposed irRC, but retrospective analyses suggest that they do add something to conventional response criteria, according to Axel Hoos, M.D., Ph.D., the medical lead for ipilimumab at Bristol-Myers Squibb in Wallingford, Conn., and co-chair of the CIC Executive Committee. He said the company is prospectively testing the irRC in ongoing and future ipilimumab trials.
Ipilimumab is not the only immuno-therapy that induces unusual response patterns that the irRC are designed to cover. For example, in the April issue of *Melanoma Research*, researchers at the Vrije Universiteit Brussel in Belgium reported similar delayed responses in patients treated with dendritic cell vaccines and interferon.

Although the irRC may seem unnecessarily detailed at first, some researchers think that previously tested experimental immunotherapies may have been considered inactive because they were evaluated with traditional tumor response criteria and passed over. (Wolchok, for example, hypothesizes that tremelimumab, an antibody that binds to the same cell surface receptor as ipilimumab, failed in a phase III trial partly because physicians used conventional criteria to identify responders and non-responders, and the non-responders then took other therapies, which possibly muddied the overall survival analysis.) Hoos isn’t willing to speculate about the fate of individual therapies, such as tremelimumab, but said, “We now recognize that some therapies have activity in early clinical trials, whereas in the past we might have dismissed agents as not effective because they were not showing an immediate shrinkage of tumor.” With increased understanding about the timing of response to immunotherapies and new tools for measuring such responses, he expects the new immunotherapies will be more likely to succeed.

**Altered Statistical Models?**

Not everyone thinks new tools are necessary, however. One concern is that clinical researchers may leave a patient on an experimental therapy longer than they should because they are waiting for a delayed response, according to Donald Berry, Ph.D., head of the division of quantitative sciences at the University of Texas M.D. Anderson Cancer Center in Houston. If so, the patient would be exposed to more risks than necessary, Berry argues in an editorial accompanying Hoos and colleagues’ proposed endpoints.

Berry also disagrees with the CIC’s argument that overall survival benefit with immunotherapies doesn’t start to show immediately and that Kaplan–Meier survival curves need to be analyzed with different statistical approaches. Rather, he thinks, the survival curves look like those that occur with many cytotoxic agents and are amenable to analysis with standard approaches.

Interestingly, Diane Lejeune, Ph.D., a senior manager in Immunotherapeutics Strategic Portfolio Analysis at GlaxoSmithKline Biologicals in Rixensart, Belgium, agrees with that last objection, at least in part. She said that although her team has seen a slow separation of survival curves in some clinical immunotherapy trials, that doesn’t always happen. For example, time-to-treatment-failure survival curves separated slowly in a phase II trial that tested the MAGE A3 Antigen-Specific Cancer Immunotherapeutic vaccine in patients with metastatic melanoma. However, when the researchers split the patients into three groups to identify, test, and then validate a gene signature associated with response, the time-to-treatment-failure curves diverged quickly in the patients whose tumors expressed the gene signature and those whose tumors did not.

Lejeune acknowledges the importance of using appropriate evaluation tools but argues that choosing the right patients and the right disease setting is crucial as well. More and more, she said, evidence indicates that immunotherapies may be most effective in the adjuvant setting, when most of the disease has been removed surgically and the patient’s immune system only needs to keep a relatively small number of cancer cells in check. And in that setting, traditional response evaluation criteria work just fine.

“We don’t need different response criteria for this situation,” Lejeune said. “The irRC are only relevant in the metastatic setting.”

As for the criticisms about the new endpoints, Hoos indicates that they are part of the process and should be taken into account as research on endpoints continues. “The criteria themselves are at the beginning of new tools for immunotherapy development,” Hoos said. “Nobody says they are perfect. Nobody says this can’t be improved upon. But what is important is that they represent innovation in the right direction that we can continue to evolve.”

© Oxford University Press 2011. DOI: 10.1093/jnci/djr184

*Melanoma Research*, researchers at the Vrije Universiteit Brussel in Belgium reported similar delayed responses in patients treated with dendritic cell vaccines and interferon.

Although the irRC may seem unnecessarily detailed at first, some researchers think that previously tested experimental immunotherapies may have been considered inactive because they were evaluated with traditional tumor response criteria and passed over. (Wolchok, for example, hypothesizes that tremelimumab, an antibody that binds to the same cell surface receptor as ipilimumab, failed in a phase III trial partly because physicians used conventional criteria to identify responders and non-responders, and the non-responders then took other therapies, which possibly muddied the overall survival analysis.) Hoos isn’t willing to speculate about the fate of individual therapies, such as tremelimumab, but said, “We now recognize that some therapies have activity in early clinical trials, whereas in the past we might have dismissed agents as not effective because they were not showing an immediate shrinkage of tumor.”

Berry also disagrees with the CIC’s argument that overall survival benefit with immunotherapies doesn’t start to show immediately and that Kaplan–Meier survival curves need to be analyzed with different statistical approaches. Rather, he thinks, the survival curves look like those that occur with many cytotoxic agents and are amenable to analysis with standard approaches.

Interestingly, Diane Lejeune, Ph.D., a senior manager in Immunotherapeutics Strategic Portfolio Analysis at GlaxoSmithKline Biologicals in Rixensart, Belgium, agrees with that last objection, at least in part. She said that although her team has seen a slow separation of survival curves in some clinical immunotherapy trials, that doesn’t always happen. For example, time-to-treatment-failure survival curves separated slowly in a phase II trial that tested the MAGE A3 Antigen-Specific Cancer Immunotherapeutic vaccine in patients with metastatic melanoma. However, when the researchers split the patients into three groups to identify, test, and then validate a gene signature associated with response, the time-to-treatment-failure curves diverged quickly in the patients whose tumors expressed the gene signature and those whose tumors did not.

Lejeune acknowledges the importance of using appropriate evaluation tools but argues that choosing the right patients and the right disease setting is crucial as well. More and more, she said, evidence indicates that immunotherapies may be most effective in the adjuvant setting, when most of the disease has been removed surgically and the patient’s immune system only needs to keep a relatively small number of cancer cells in check. And in that setting, traditional response evaluation criteria work just fine.

“We don’t need different response criteria for this situation,” Lejeune said. “The irRC are only relevant in the metastatic setting.”

As for the criticisms about the new endpoints, Hoos indicates that they are part of the process and should be taken into account as research on endpoints continues. “The criteria themselves are at the beginning of new tools for immunotherapy development,” Hoos said. “Nobody says they are perfect. Nobody says this can’t be improved upon. But what is important is that they represent innovation in the right direction that we can continue to evolve.”