The article by Hoos et al. (1) in this issue of the Journal makes three recommendations about endpoints in immunotherapy trials: 1) cellular immune response assays are highly variable and require standardization; 2) immunotherapies may evince unusual patterns of antitumor response and therefore require different endpoints; and 3) immunotherapy may have a delayed benefit on survival, which requires a different approach to achieving statistical power in clinical trials. All three recommendations are important and indeed are worthy of consideration in oncology more generally. I will expand on the authors’ points, focusing on the importance of immune response, biomarkers, and tumor response and progression.

Historically, cancer drug development has been remarkably inefficient. Suppose we were to use the same approach in developing agricultural products, for example, corn varieties. We start with many genetically distinct varieties of corn. In the first phase of development, we would plant seeds from each variety and spray the sprouting plants with pesticide. We would drop from further development those varieties that seem adversely affected. In the next phase, we would plant seeds from those varieties still under consideration and pull up the stalks 8 weeks later to weigh the ears. Small ear varieties would be discarded. In the next phase, 14 weeks after planting, we assess the sugar content of the kernels, discarding those with low concentrations. Then, we would consider hardness and so on. We’d probably end up with lousy corn and we’d wonder if we had discarded a better overall variety along the way.

One problem with this approach to choosing corn varieties is that we would have failed to address the relationships among the various effects. Another is that we would have failed to observe the effects of all the variables at maturity. But the biggest problem is the staccato learning process. Each experiment was a discrete unit rather than one part of a continual learning process in which the answers to relevant questions got stronger over time. In the clinical development of cancer drugs, we consider toxicity early on, then antitumor effect (or immunologic effect or other biomarker effect), and then clinical effect. We fail to assess and relate these effects over the patients’ follow-up, and we fail to associate them with the primary clinical endpoint in a longitudinal fashion. We must address the entire patient, including the patient’s course of disease over time. Especially important are tumor burden and factors that may be related to tumor burden. In the immunology setting, this includes measures such as T-cell proliferation.

There are obstacles to building better drug and vaccine development. One is the notion that early endpoints should play a role in clinical trials only if they are surrogates for the primary endpoint. Another is the notion that we can use markers only if they have been “validated.” Both notions inhibit progress. We should assess all available information, associate it with the primary endpoint, and use these associations to draw stronger conclusions about the primary endpoint—recognizing that we will find some falsely positive relationships because of multiple comparisons. If there is a relationship between a particular biomarker and the primary endpoint in a clinical trial, then we should exploit it. If there is no relationship, then we will learn that and have lost little.

I turn to the issue of unusual patterns of antitumor response evinced by immunotherapies with a possible consequence being that any survival effect is delayed. Hoos et al. (1) advocate measuring tumor burden using the sum of perpendicular diameters of all lesions, both index and new. However, they advocate a similar categorization as that for RECIST, and they do not sufficiently exploit this new measure’s continuous nature. For example, Hoos et al. (1) state, “Thus, the definition of confirmation of progression represents an increase in tumor burden of at least 25% compared with baseline at two consecutive time points at least four weeks apart.” Why 4 weeks? A 25% increase that goes to 100% at 3 weeks would seem to be confirmatory. Indeed, a sufficient increase (perhaps 100%) in a single measurement may not require confirmation.

As Hoos et al. (1) indicate the possibility of delayed therapeutic benefit presents a major conundrum for clinical trials. Delayed benefit can manifest itself in several ways. One example is that there is no prolongation in time to progression, but there is longer survival postprogression (2). This is unusual for standard oncology therapies. To achieve a benefit in overall survival usually requires delaying the disease in ways that can be measured. There have been some important counterexamples that have strained our collective credibility (1). How can it be that tumor burden is not affected by a therapy and yet the patient lives longer on the therapy? Hoos et al. (1) argue that immunotherapies have kinetics different from cytotoxic agents and first induce cellular immune responses before any reduction can be seen in tumor burden. This is quite plausible. However, figures 4 and 5 of Hoos et al. (1) do not persuade me that the delayed benefit is real. Figure 1 shows that the hazard ratio function in their figure 5 and a constant hazard ratio at, say, 0.80 give rise to very similar survival curves. Hence, in an empirical setting, it will be nearly impossible to distinguish between them. Similarly, the Kaplan–Meier curves in their figure 4 may be consistent with the hazard ratio shown in their figure 5, but they are also consistent with constant hazard. Although the general shape of figure 4 was confirmed in a subsequent trial (3), even taking the two trials together does not compellingly argue for delayed benefit. The two curves are not unlike some survival curves showing benefit of cytotoxic chemotherapy.

Any delayed effect of therapy makes product development harder and more expensive than developing a drug that works by attacking the tumor directly (1). Follow-up has to be longer, and it is difficult to give up on such a therapy. This presents a major conundrum.
To fully investigate the potential of an immunotherapy, clinicians may have to stick with it beyond a patient’s progression and thereby delay switching to potentially more effective therapy. This has happened in at least one phase III trial, to the detriment of the patients. As is usual in clinical research, there is no easy answer in the face of conflicting desiderata. The Belmont Report (http://ohsr.od.nih.gov/guidelines/belmont.html) makes clear that clinical trials are experiments, in which patients may be exposed to risks. The potential for such exposure must be clearly explained to patients, along with the potential for benefit. It is also incumbent on investigators to build designs that minimize risk without sacrificing science. It is a delicate balance. Having a predictive marker of benefit is an important part of chemotherapy trials, but the possibility of a late survival benefit in immunotherapy trials makes inclusion of predictive markers of immune response even more important.

To add one more complication to the mix, accrual rate becomes more important in immunotherapy trials. Finishing a trial in a reasonable time requires rapid accrual. However, rapid accrual may mean having hundreds of patients in a trial with none having reached the point at which the survival curves are expected to separate. Outcomes that would result in stopping a trial of standard cytotoxic experimental therapy would be less negative and might even be expected in an immunotherapy trial. Even though such a trial may not be stopped outright, ethical considerations might require putting a governor on accrual, or pausing accrual, to see whether there is evidence of the curves separating with longer follow-up. This of course will slow the agent’s development.

Hoos et al. (1) suggest random assignment in phase II immunotherapy trials. I wholeheartedly concur (4). They accept that their recommendations in the context of phase III will result in larger statistically overpowered trials. An alternative that I prefer is to build an adaptive trial that learns about hazards over time, controls accrual, and stops accrual when the answer is known or is predicted to be known after additional follow-up of the presently accrued patients (5). Specifically, piecewise exponential models (6) could be used to fulfill the suggestion of Hoos et al. (1) for “. . . altered statistical models describing hazard ratios as a function of time.”

References


Notes

Donald Berry is part owner of Berry Consultants, LLC, specialists in adaptive Bayesian designs and whose clients include companies that are developing immunotherapies: Bristol-Myers Squibb and BioVest International.

Affiliation of author: Department of Biostatistics, University of Texas M. D. Anderson Cancer Center, Houston, TX.