Response Rate as an Endpoint in Clinical Trials

Stephen L. George

In this issue of the Journal, Lewis et al. (1) report the results of a randomized clinical trial in operable osteosarcoma of an extremity designed to compare “standard” two-drug therapy, utilizing cisplatin and doxorubicin given in 3-week cycles, with a “dose-intensive” or “dose-dense” version of the same therapy given in 2-week cycles with granulocyte colony-stimulating factor support. The primary endpoint was overall survival with a secondary endpoint of progression-free survival. More than 500 patients were entered in a 9-year period from 1993 to 2002, the second largest clinical trial ever completed in osteosarcoma, according to the authors. The results of the primary and secondary analyses showed little difference between the treatment groups. The estimated hazard ratios were 0.94 for overall survival (P = .64) and 0.98 for progression-free survival (P = .82). However, the histologic response rate, defined as the percentage of patients in whom there was greater than 90% necrosis of the tumor, was significantly higher (P = .003) in the dose-dense treatment group (51%) than in the standard treatment group (36%).

Putting aside the nontrivial issues in the definition, assessment, and verification of responses in an open-labeled trial such as this one and the fact that response rate was not one of the primary or secondary endpoints, the evidence is fairly clear that the dose-intensive regimen produced a higher response rate than the standard regimen. Unfortunately, this difference did not translate into better survival or progression-free survival. This is not an uncommon finding in clinical trials; promising early results measured by response rates often do not translate into long-term benefits such as longer progression-free survival or overall survival. Part of the reason may be that fairly large increases in response rates are required to translate into detectable overall survival differences (2). There are examples, such as metastatic breast cancer, in which a response rate difference does appear to translate into an overall survival benefit, but even in these cases, the strength of the correlation tends to be quite low (3). Of course, the reverse can also happen, albeit rarely; treatments that do not differ with respect to response rates can have different overall survival. This situation may occur for effective treatments that have modes of actions that would not be expected to have a direct effect on response rates as traditionally defined. Either way, reliance on response rates as an early indicator of treatment benefit or lack of benefit can be misleading and the use of response rate as a primary endpoint in definitive clinical trials is problematic.

The use of response rate as an endpoint in clinical trials is widespread for the obvious reason that objective response to therapy is a clear early indication of activity in diseases for which spontaneous remissions are rare or nonexistent and because of the often tacit assumption that early objective response is a sine qua non for eventual clinical benefit. Thus, early-phase clinical trials have commonly used objective response rate as the primary endpoint and as the measure of treatment activity, although the situation has been changing somewhat in recent years for treatments without direct cytotoxic effects. In designing clinical trials in a specific disease that involve the assessment of objective response, several questions should be asked, and the existing evidence bearing on the answers to those questions should be assessed before deciding on how response rate as an endpoint might be used effectively in the trial.

Is response itself a direct measure of clinical benefit to the patient? If we define clinical benefit as something that makes a patient feel better or live longer, a loose interpretation of Food and Drug Administration (FDA) definition of the phrase, then the answer is probably “no” for most solid tumors. The RECIST criteria for response, for example, concern measurement of various target lesions with no direct clinical benefit per se, although in some cases, the reduction in tumor burden may lead to an improvement in symptoms. Many of the hematologic malignancies are different in this regard. For example, attainment of a durable complete response in acute leukemia may appropriately be assumed to be a direct clinical benefit to the patient by reducing symptoms of the disease in nearly all cases. In a regulatory setting, response rate has often been used by the FDA as the basis for full or accelerated approval for certain indications in cancer, although there has been a growing appreciation of the risks of doing so (4).

Is response a prognostic factor for some measure of clinical benefit? In many settings in cancer, the answer is “yes,” particularly if “durability of response” is added to the definition, although the strength of the association between response and clinical benefit is variable and the analyses must be carefully done (5). Simple comparison of responders and nonresponders is not a valid approach. Some statistical modeling or landmark analyses are required to address this issue. In Lewis et al., various landmark analyses indicated that response was an important prognostic variable, with responders having a more favorable prognosis with respect to both progression-free survival and overall survival. Such a finding corresponds to our intuitive sense that response is indicative of sensitivity to therapy and that such sensitivity should translate into clinical benefit. However, prognostic validity alone is insufficient reason for using response rate as a substitute, or surrogate, for clinical benefit.

Is response rate a surrogate endpoint for some established measure of clinical benefit? If it were, we could use response rate as a primary endpoint and generally get answers much sooner and without the complications of intervening treatments than we could by using...
endpoints such as survival or progression-free survival. To assess the adequacy of response rate or any other putative surrogate, several issues must be addressed as detailed in recent work on this topic (6,7). First, response rate must be prognostic as discussed above. Second, there must be a treatment effect on both response rate and the clinical benefit endpoint. This second condition was not met in the Lewis et al. study. Third, the treatment effect on response rate must explain most or all of the treatment effect on the clinical benefit endpoint. Unfortunately, examples of definitive validation of putative surrogate endpoints are rare, and these usually require a meta-analysis of several studies. It should also be emphasized that the definition of surrogacy depends on the treatment or at least the class of treatment, so that a change to a treatment with a different mode of action requires revalidation. The net result of all these considerations is that very few surrogates, including response rate, have been validated in specific clinical settings involving specific classes of treatments (8).

From the above considerations, it is clear that response rate as an endpoint is most useful in early-phase clinical trials, in which interest focuses more on assessing activity than proving clinical benefit, and as an important but secondary outcome in more definitive trials.

References