Clinical Practice Guidelines: Still a Way to Go

This month’s issue provides the journal reader with an elegant analysis of the clinical evidence substantiating the use of aromatase inhibitors (AIs) in the adjuvant treatment of breast cancer in postmenopausal women. Based on the data from 6 randomized studies, women may be treated with an AI alone, with 2 or 3 years of tamoxifen followed by an AI, or with 5 years of tamoxifen followed by another extended period of an AI. As this issue’s special feature by Carlson et al. points out, the NCCN guidelines, the ASCO Technology Assessment, and the St Gallen consensus statement all recommend these 3 options. All 3 documents also state that existing data do not allow a choice among these approaches because they have not been compared head-to-head.

In the broadest sense, all 3 panels agree that AIs offer a benefit, but the recommendations are not identical. The St Gallen recommendations also present a fourth option, the use of standard tamoxifen in patients at minimal or intermediate risk. The NCCN, in turn, does not recommend adjuvant treatment for very small lesions. As Baum and Ravdin previously observed, patients in the same clinical category could potentially be treated differently based on which clinical guideline their practitioner consulted.

Why this disparity? Several differences in the guideline development process may account for it. The first and most obvious reason is that the panels can interpret data differently or attribute differing validity to the studies. This does not appear to be the case in this instance. A second possible reason is related to how patients are grouped for clinical decision-making, i.e., the somewhat homogeneous groups based on common clinical factors for which certain interventions are indicated or for whom similar outcomes are expected. These groupings are constructs and may vary from guideline to guideline. In making clinical branch points, St Gallen and NCCN differ on the size of lesions, the definition of estrogen receptor positivity, and the use of grading. Thus, women who present with similar favorable factors might be classified differently according to the guideline used, and therefore receive a different recommendation. For women at higher risk, the groupings overlap, and an AI-containing regimen would be recommended no matter which guideline was used.

Thirdly, the guidelines do not make the threshold needed for a recommendation explicit. For example, if the degree of hazard ratio needed to exclude a treatment was set differently, guidelines might arrive at different conclusions. In another example, 2 panels may disagree on the length of follow-up needed for data to mature enough for a recommendation. Thus, the ASCO Technology guidelines lagged behind the NCCN guidelines in presenting AIs as an option.

Finally, factors unrelated to clinical trials might impinge on the recommendations, including issues such as cost or availability. The Ontario experience specifically includes government input into the guideline process to assess whether the recommendations are within the payer’s resources.

What should be done? First, the more explicit the parameters used to make decisions, the better. This will allow users to understand why the recommendations differ and to make decisions accordingly. The second trend should be harmonization. Guidelines developers should talk to each other and smooth out some of the differences, if possible, so that guidelines will be more congruent.

References