Pre-Judging Data: Benchmarking Clinical Significance Before Study Results Are Known

With major clinical trial, usually randomized phase III studies, determining the worthiness of the findings involves several layers of consideration. The first is the statistical evaluation and demonstration that a primary study end point has achieved the essential (if arbitrary) P value of less than 0.05. The second is review by regulatory authorities such as the FDA. FDA approval is usually predicated on a statistically significant finding in a specific setting that meets predetermined criteria discussed privately between the FDA and the drug developer or study organizer. These analyses are obviously critical for new drug approvals.

But these assessments do not necessarily incorporate clinical judgment: is the treatment worthwhile? Does it do enough to justify widespread use? How does it compare with other options? Should side effects or other aspects of the treatment affect consideration of the regimen? These are the matters of clinical impact that are often adjudicated by experts, including those who comprise the many NCCN treatment panels.

Typically, these decisions are made after the data have been amassed and presented, either at major international conferences or published in peer-reviewed journals. This ensures that the data are in the public domain and that the presentation has some accountability because it has been through the scrutiny of additional eyes beyond those of the study sponsors themselves.

A full and honest reckoning of a treatment can only be made when all the data are available; for that reason, it makes perfect sense that experts withhold judgment until major public reports are completed. A wrinkle in this timetable, however, is that the release of important data is often highly orchestrated, and certain results can create momentum of their own, based on quick impressions of limited information.

As an adjunct to this current model of critiquing data, consider a strategy used by many other sectors: pre-judging the data and handicapping the outcomes. For many business analysts, trying to anticipate results and make judgments based on estimated benchmarks is quite common. For instance, stock analysts will often estimate quarterly profits for a company, pegging a number as being “good” or, conversely, suggesting that failing to reach that goal would be a disappointment. The same is true for the entertainment business, where estimates of revenue for a soon-to-be-released movie are frequently available, and in the realm of politics, where pundits are constantly suggesting that a candidate needs to perform at a certain target level to be a winner.

The common element here is that a benchmark is declared before data are available, and this standard is held up as “worth,” Once data are available, additional interpretation is then made. If the data exceed expectations, people notice. If data fall short, those involved work to explain. Mitigating circumstances are discussed—was the benchmark set too high; was it, in retrospect, too low?

Now consider how this might work with results from important clinical trials. Imagine if a community of experts articulated what a benchmark should be to make the study result meaningful or worthwhile. The data would then be declared important if they exceed that benchmark, or evaluated negatively if they fell short. This strategy has the virtue of stating in advance what is considered a positive or negative outcome, and the adjudication for positivity and negativity is done in a transparent, hopefully unbiased, fashion.

Let’s apply this strategy to two important trials in early-stage breast cancer; the HERA study, which compared 0 versus 1 versus 2 years of trastuzumab as adjuvant therapy for HER2-positive breast cancers (the comparison of 1 vs. 2 years is not
yet available), and ECOG 5103, which compared chemotherapy alone versus chemotherapy with bevacizumab as adjuvant treatment. These large multicenter studies have long since closed to accrual, and we are waiting for the clinical outcomes data. The near-term toxicity data should all be available for consideration. Imagine if a blue-ribbon panel of experts and advocates met right now, and defined what would be considered a meaningful clinical benefit from each of these trials. Such a panel might assess, for instance, whether a difference in progression-free survival of 3% at 5 years, in the absence of a significant survival difference, would be sufficient. Should it be 5% at 5 years? 2%? 10%? Should we require a survival difference? If so, how much? What if life-threatening toxicity occurs in 1% of patients? In 2%? How would such toxicity counterbalance the need for positive clinical benefits? You get the idea.

My point is not to suggest what the benchmarks should be, but rather to suggest that it is entirely possible to have a meaningful dialogue right now about the importance of these results. In particular, it is possible to create a public, fair, and reasoned benchmark defining what a study really ought to show to justify a new clinical standard for a treatment. Doing it ahead of time would give the clinical and advocacy communities the opportunity to weigh the likely significance of any study findings without the drum-beat of publicity and marketing that often accompanies the highly orchestrated release of major data. It would allow for a sober and reasoned assessment of the data, liberated from the protocol-driven needs of the P value. Prejudging the study results would not be a perfect process. Each trial finds new things to say and new twists in what we anticipate. But it would provide a clear roadmap that might facilitate regulatory and guideline recommendations and be less subjective in its ultimate assessment of a new treatment.

**Join the Debate**

What do you think? We welcome your comments about this article, any of the materials in JNCCN, or any topic relating to cancer prevention, detection, treatment, supportive care, or survivorship. Please e-mail correspondence (include contact information if you would like your comments considered for possible publication) to JNCCN@ncen.org or mail to JNCCN, 275 Commerce Drive, Suite 300, Fort Washington, PA 19034.