Achieving success in the development of a cancer drug continues to be challenging. Given the increasing costs (1) and the small number of drugs that gain regulatory approval (2), it is crucial to understand these failures. In this issue of the Journal, Gan et al. (3) reviewed 235 recently published phase III randomized clinical trials (RCTs). They report that 62% of the trials did not achieve results with statistical significance. Trying to explain the high failure rate, they note the actual magnitude of benefit achieved in a clinical trial (designated B) is nearly always less than what was predicted at the time the trial was designed (designated 8) and conclude, “investigators consistently make overly-optimistic assumptions regarding treatment benefits when designing RCTs.”

But really should we be surprised that phase III trials, the venue for detecting “small” differences, so often disappoint? Almost by definition, phase III studies are designed to detect small differences (4,5). The problem is that small has given way to “marginal” as a definition, phase III studies are designed to detect small differences (4,5). The problem is that small has given way to “marginal” as a definition, phase III studies (11). Similarly, the rate of stable disease and the “clinical benefit rate,” two measures increasingly reported in early phase studies, have never been shown to correlate with outcomes yet are regarded by many as measures of efficacy (12–14). Hence, we would argue that inaccurate assessments of limited data and reliance on endpoints that have not been validated are likely more important than overoptimism.

To be sure, Gan et al. (3) recognize it is not just about statistical validity when they acknowledge “significant benefit could also result from overpowered studies that detect differences that are not clinically meaningful.” Noting that their data showed many positive studies in which the observed difference was less than predicted but still statistically significant, they wonder, as have others, whether these “positive” studies merit regulatory approval in the absence of additional supporting data (15).

We agree with the authors’ opinion that more research is needed to determine how to better define 8; a goal they suggest might be achieved by “using statistical modeling rather through empiricism.” But pending the outcome of that research, they advocate more frequent use of interim analyses with options of early study termination for futility or efficacy and adaptive trial designs. In regards to the latter, they note, “up to 50% of RCTs that do not show a statistically significant benefit might actually be false-negative trials.” Their suggestion that these studies did not enroll enough patients and are underpowered because of unreliable 8 values has the inherent assumption that marginal benefits matter. But with 100275 patients enrolled in 158 negative trials, and an average trial size of 635 patients, many would not consider the magnitude of benefit missed to be “clinically meaningful” or worth the enrollment of hundreds of additional patients to confirm. Furthermore, their assertion that “if 8 is set unrealistically high, the trial will be underpowered to detect a smaller but still clinically meaningful benefit, resulting in a negative trial” assumes that marginal differences can confer clinically meaningful benefit, an assumption with which we disagree (14,16).

It is interesting how so much is reported and can be analyzed about what constitutes a “statistically significant benefit”; yet, so little is reported or devoted to assess the statistical validity of toxicity and the risk to benefit ratio of a therapy. Indeed, one would be hard-pressed to find an experimental arm deemed statistically superior in terms of efficacy described as anything but tolerable. Unfortunately, increasingly it appears that any toxicity is tolerable or acceptable provided some gain, no matter how marginal, is achieved. But as efficacy gains become increasingly smaller, toxicity becomes increasingly important. And as Gan et al. (3) remind us, unfortunately for cancer patients, toxicity is all too often more than a grade 1 rash. Toxicity can be both severe and life altering and unfortunately at times is accompanied by a statistically inferior outcome. Remarkably, among 158 negative studies, Gan et al. found a trend toward detriment in the experimental arm in 42

**Why Do Phase III Clinical Trials in Oncology Fail so Often?**

Laleh Amiri-Kord estani, Tito Fojo

Correspondence to: Tito Fojo, MD, PhD, Medical Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bldg 10, Rm 12N226, 9000 Rockville Pike, Bethesda, MD 20892 (e-mail: fojott@mail.nih.gov).
studies and a statistically significant detriment in eight RCTs that enrolled 5287 patients.

The observation that trials with industry funding were more likely to be positive and associated with a small but statistically significant increased risk of detriment in the experimental arm is of concern. Their greater likelihood to achieve a positive outcome may well reflect their larger trial size because of better funding. We can only speculate as to why there would be an increased risk of detriment in the experimental arm, but the possibility that toxicity emerging in phase I/II trials would be less likely to derail a product in which a substantive investment has been made cannot be discounted. Furthermore, targeted agents—the principal if not exclusive components of the portfolios of all major companies—were no better when toxicity in the experimental arm was concerned. Because targeted agents will dominate the oncology enterprise for the rest of this decade, these observations are discouraging.

Although our goals may initially be lofty, they eventually meet reality. And in cancer drug development, reality is all too often failure (2). The challenge in oncology is to be sure that we remain focused on true clinical benefit—prolonging life. Our goals must remain lofty, and we must remember that marginal benefit should never be that goal (14,16). We need to be vigilant and as soon as it becomes apparent that any benefit will be marginal, we must discard that strategy and move on, ensuring we do not redefine failure as success. The data of Gan et al. (3) warn us that we are at risk of losing our focus. Conducting larger trials, doing more interim analyses, or using adaptive trial designs are not the solutions (17–19). We do not need more marginal results that are then pronounced “new treatment paradigms” or a “new standard of therapy.” What we need are meaningful goals and better drugs—much better drugs aimed at targets that are really important!

References


Notes

The authors have no conflicts to disclose.

Affiliation of authors: Medical Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD.