

EDITORIALS

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

Laleh Amiri-Kordestani, 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: fojot@mail.nih.gov).

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 δ) 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 outcomes have fallen below our already modest expectations. And who or what is to blame? Are investigators really overly optimistic regarding experimental therapies and, as the authors suggest, responsible for the large number of negative studies? Although we agree that optimism regarding clinical benefit may lead to an underpowered trial, we disagree that optimistic investigators are those we should blame. We would ask, how do Gan et al. (3) define optimism? Where do they place the line between an optimistic and a realistic expectation? The authors demonstrated a poor correlation between the expected and observed benefits but in the majority of trials also found the “expected benefits” were less than 4 months—a duration many would argue represents a modest and defensible expected benefit for the majority of solid tumors. So that, rather than excessive optimism, we believe several factors including inaccurate assessments of “limited data from early phase trials and/or investigators’ experience” interpreted in what the authors themselves acknowledge “is usually an empirical process” lead to the differences that Gan et al. (3) found between the actual (B) and predicted (δ) benefit. Although there are models that use the results of phase I/II trials to predict the outcome of phase III studies, no model is perfect (6–8). For example, the response rate, which is part of the limited dataset available in designing phase III trials, has been correlated with survival and clinical benefit (9,10). But other factors such as the duration of response make response rate less reliable and lead to discrepancy in the results of phase II and 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 δ ; 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 δ values has the inherent assumption that marginal benefits matter. But with 100 275 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 δ 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

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

1. Collier R. Rapidly rising clinical trial costs worry researchers. *CMAJ*. 2009;180(3):277–278.
2. Bates SE, Amiri-Kordestani L, Giaccone G. Drug development: portals of discovery. *Clin Cancer Res*. 2012;18(1):23–32.
3. Gan HK, You B, Pond GR, Chen EX. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. *J Natl Cancer Inst*. 2012;104(8):590–598.

4. Wu W, Shi Q, Sargent DJ. Statistical considerations for the next generation of clinical trials. *Semin Oncol*. 2011;38(4):598–604.
5. Hoering A, Leblanc M, Crowley JJ. Randomized phase III clinical trial designs for targeted agents. *Clin Cancer Res*. 2008;14(14):4358–4367.
6. Claret L, Girard P, Hoff PM, et al. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J Clin Oncol*. 2009;27(25):4103–4108.
7. Chen TT, Chute JP, Feigal E, Johnson BE, Simon R. A model to select chemotherapy regimens for phase III trials for extensive-stage small-cell lung cancer. *J Natl Cancer Inst*. 2000;92(19):1601–1607.
8. Bruno R, Lu JF, Sun YN, Claret L. A modeling and simulation framework to support early clinical drug development decisions in oncology. *J Clin Pharmacol*. 2011;51(1):6–8.
9. Tsujino K, Shiraishi J, Tsuji T, et al. Is response rate increment obtained by molecular targeted agents related to survival benefit in the phase III trials of advanced cancer? *Ann Oncol*. 2010;21(8):1668–1674.
10. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer. *Lancet*. 2000;356(9227):373–378.
11. Pazdur R. Response rates, survival, and chemotherapy trials. *J Natl Cancer Inst*. 2000;92(19):1552–1553.
12. Vidaurre T, Wilkerson J, Simon R, Bates SE, Fojo T. Stable disease is not preferentially observed with targeted therapies and as currently defined has limited value in drug development. *Cancer J*. 2009;15(5):366–373.
13. Tolcher AW. Stable disease is a valid end point in clinical trials. *Cancer J*. 2009;15(5):374–378.
14. Ohorodnyk P, Eisenhauer EA, Booth CM. Clinical benefit in oncology trials: is this a patient-centred or tumour-centred end-point? *Eur J Cancer*. 2009;45(13):2249–2252.
15. Ocana A, Tannock IF. When are “positive” clinical trials in oncology truly positive? *J Natl Cancer Inst*. 2011;103(1):16–20.
16. Booth CM, Ohorodnyk P, Eisenhauer EA. Call for clarity in the reporting of benefit associated with anticancer therapies. *J Clin Oncol*. 2009;27(33):e213–e214.
17. Emerson SS, Fleming TR. Adaptive methods: telling “the rest of the story”. *J Biopharm Stat*. 2010;20(6):1150–1165.
18. Chow SC, Corey R. Benefits, challenges and obstacles of adaptive clinical trial designs. *Orphanet J Rare Dis*. 2011;6:79.
19. Berry DA. Adaptive clinical trials in oncology. *Nat Rev Clin Oncol*. 2011.

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.