

# Overall Survival As the Outcome for Randomized Clinical Trials With Effective Subsequent Therapies

Edward L. Korn, Boris Freidlin, and Jeffrey S. Abrams

Edward L. Korn and Boris Freidlin, Biometric Research Branch; and Jeffrey S. Abrams, Cancer Therapy Evaluation Program, National Cancer Institute, Bethesda, MD.

Submitted January 3, 2011; accepted March 18, 2011; published online ahead of print at [www.jco.org](http://www.jco.org) on May 9, 2011.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Edward L. Korn, PhD, Research Branch, EPN-8129, National Cancer Institute, Bethesda, MD 20892; e-mail: [korne@ctep.nci.nih.gov](mailto:korne@ctep.nci.nih.gov).

© 2011 by American Society of Clinical Oncology

0732-183X/11/2917-2439/\$20.00

DOI: 10.1200/JCO.2011.34.6056

## A B S T R A C T

We review how overall survival (OS) comparisons should be interpreted with increasing availability of effective therapies that can be given subsequently to the treatment assigned in a randomized clinical trial (RCT). We examine in detail how effective subsequent therapies influence OS comparisons under varying conditions in RCTs. A subsequent therapy given after tumor progression (or relapse) in an RCT that works better in the standard arm than the experimental arm will lead to a smaller OS difference (possibly no difference) than one would see if the subsequent therapy was not available. Subsequent treatments that are equally effective in the treatment arms would not be expected to affect the absolute OS benefit of the experimental treatment but will make the relative improvement in OS smaller. In trials in which control arm patients cross over to the experimental treatment after their condition worsens, a smaller OS difference could be observed than one would see without cross-overs. In particular, use of cross-over designs in the first definitive evaluation of a new agent in a given disease compromises the ability to assess clinical benefit. In disease settings in which there is not an intermediate end point that directly measures clinical benefit, OS should be the primary end point of an RCT. The observed difference in OS should be considered the measure of clinical benefit to the patients, regardless of subsequent therapies, provided that the subsequent therapies used in both treatment arms follow the current standard of care.

*J Clin Oncol* 29:2439-2442. © 2011 by American Society of Clinical Oncology

## INTRODUCTION

Prolongation of overall survival (OS) is generally the most relevant measure of clinical benefit in a randomized clinical trial (RCT) of an experimental treatment. However, the availability of active therapies that are given with worsening patient condition after the randomized treatment complicates the interpretation of OS differences. We examine in detail the clinical benefit of an experimental treatment evaluated in an RCT when there are effective subsequent therapies. We first consider the situation in which the same subsequent therapy is used in both treatment arms and is differentially effective or equally effective in the treatment arms. Next we discuss the special challenges presented by trials in which patients can cross over from the standard to the experimental treatment. We then provide some examples and conclude with recommendations for OS as a clinical trial end point.

## SUBSEQUENT THERAPY DIFFERENTIALLY EFFECTIVE IN THE RANDOMLY ASSIGNED TREATMENT ARMS

With no subsequent therapies, assume hypothetically that the median survival would be 9 months for

the experimental treatment  $E$  and 6 months for the standard treatment  $S$ . Now suppose that the same subsequent therapy  $X$  is given to patients in either randomly assigned treatment arm when their condition worsens (eg, at progression), and with this subsequent therapy, that the median survival for both treatment arms is 10 months. For shorthand notation, we will refer to the treatment arms as  $E \rightarrow X$  and  $S \rightarrow X$ . Thus, the subsequent treatment improves the median survival by 1 month in the experimental arm and by 4 months in the standard arm. One can easily imagine this happening if the subsequent treatment and experimental treatment have similar mechanisms of action, with the tumor becoming resistant to this particular type of therapy. What can we say about the clinical benefit of the experimental treatment if this is the true state of affairs? First, the experimental treatment  $E$  would have additional activity over the standard treatment if given alone without subsequent treatments. However, in practice,  $E$  has no clinical benefit over the standard treatment as it would be used in this setting, that is, followed by subsequent treatment  $X$ . This is because using the standard treatments  $S \rightarrow X$  is as good as using the experimental treatment, that is,  $S \rightarrow X$  is as good as  $E \rightarrow X$ .<sup>1</sup> It is possible that

treatment  $E$  would have clinical benefit if it were given at a different dose or schedule or length of administration, in an earlier disease setting, or in combination with other treatments. But the clinical benefits of  $E$  in those settings would have to be demonstrated in their own RCTs.

In reality, one does not get to observe the survival experience of  $E$  and  $S$  alone, but only  $E \rightarrow X$  and  $S \rightarrow X$ . Therefore, after observing 10-month OS medians for both  $E \rightarrow X$  and  $S \rightarrow X$ , although it is possible that there would have been a 3-month difference in OS medians if there were no subsequent therapies given, it is also possible that the OS medians would have been the same. Although between-arm differences in intermediate end points determined before the subsequent therapy is given (eg, progression-free survival [PFS] when the subsequent treatment is given at progression) may be consistent with an OS improvement, this is not necessarily the case and cannot be concluded. Special statistical methods could suggest there would have been OS differences seen if no subsequent therapy had been given, but this type of evidence is necessarily weak<sup>2</sup> and, as noted above, would be irrelevant to the question of the clinical benefit of  $E$  in this clinical setting.

In summary, reliable assessment of the hypothetical improvement in OS that would have been observed if subsequent therapies were not available (“explanatory approach”<sup>3</sup>) is infeasible and is not relevant to estimating the benefit in clinical practice with subsequent therapies (“pragmatic approach”<sup>3</sup>).

#### SUBSEQUENT THERAPY EQUALLY EFFECTIVE IN THE RANDOMLY ASSIGNED TREATMENT ARMS

Interpretation of OS differences with subsequent therapies is easier if the subsequent therapy works equally well (or equally poorly) in the two treatment arms. However, even in this situation, there are complications. For example, suppose the effect of the subsequent therapy is to extend patients’ lives 2 months, on average. One would then expect the absolute difference in median survivals between the two treatments to be approximately the same as would be seen if the subsequent therapy had not been given. Therefore, a 3-month difference in median survivals (eg, 9 months  $\nu$  6 months) without subsequent therapy would translate into an approximate 3-month difference in median survival with subsequent therapy (approximately 11 months  $\nu$  8 months). Note, however, that the corresponding relative difference in survival is decreased: a 37.5% (3 of 8) improvement instead of a 50% (3 of 6) improvement. A key point is that even though the subsequent therapy is working equally well in both treatment arms, it is likely to add variability to the OS times because some patients get less than a 2-month benefit and some patients get more than a 2-month benefit. This added variability makes it harder to estimate the treatment benefit. This means in settings with effective subsequent therapies (eg, breast cancer), larger sample sizes will be required to detect (with statistical significance) whether the experimental treatment is working (or larger hypothesized effects for the experimental treatment will need to be targeted); this will not be an issue in settings with relatively ineffective subsequent therapies (eg, hepatocellular or gastric cancer). By using typical exponential distribution assumptions, the number of events required to detect a certain ratio of median survivals is proportional of the squared logarithm of the ratio. For example, approximately a 60% larger sample size would be required to detect an improvement from 8 to 11 months versus an improvement from 6 to 9 months median survival.

#### CROSS-OVER TO EXPERIMENTAL THERAPY

There are two distinct situations to consider with different implications for the interpretation of OS results. In the first, the experimental treatment has not previously been shown to be effective in the given disease in any later-line setting; in the second, it has previously been shown to be effective in a later-line setting. Examples of the second scenario are not infrequent: when an agent is tested as a first-line metastatic treatment after it has been demonstrated to be an effective second-line treatment or when an agent is tested in the adjuvant setting after showing efficacy in the metastatic disease setting.

#### Experimental Therapy Not Previously Shown to Be Effective in Later-Line Settings

In the first situation, the purpose of designing a trial with a crossover is to increase interest in participation in the trial, because patients will eventually get access to the experimental treatment regardless of their initial treatment. When a large number of patients cross over to the experimental treatment, the trial is essentially testing whether giving the experimental therapy early is better than giving it later. Therefore, the clinical benefit of the experimental agent will be underestimated. For example, if the experimental agent extends median survival by 3 months whether it is given at the time of random assignment (in the experimental arm) or at the time of progression (in the standard arm), then the OS will appear similar in the two treatment arms, even though the experimental agent has clinical benefit in this setting. This is in contrast to the situation described previously (when the same subsequent therapy is given in both arms).

Some have suggested that it is necessary for ethical reasons to include a crossover in trial designs to allow patients access to the experimental treatment.<sup>4,5</sup> Conversely, others have suggested that allowing a crossover raises ethical concerns that there is a potential for coercion in the enrollment process.<sup>6</sup> As a practical matter, a 2:1 or even a 3:1 random assignment in favor of the experimental arm may increase interest in participation in the trial. Unequal randomization ratios require a trial with a larger sample size (13% larger and 33% larger in the case of 2:1 and 3:1 randomization ratios), but the time it takes to perform the trial may not be much longer than with a 1:1 ratio if accrual is more rapid because of the promise of the experimental agent. Additionally, formal interim monitoring that will stop the trial early if the experimental treatment is working exceptionally well may make the lack of a crossover in the trial design more acceptable.

If one believes that the pretrial evidence on the efficacy of an experimental treatment (when compared with best available therapy for that setting) is so convincing that a crossover must be allowed, then perhaps this evidence should also preclude performing the random assignment to a control treatment. Such was the case for the initial trial of imatinib for metastatic gastrointestinal stromal tumor (GIST),<sup>7</sup> which randomly assigned patients between two doses of imatinib rather than versus a control treatment. Conversely, if the pretrial expectation of efficacy is weaker, then the perceived ethics of allowing crossovers should be weighed against their effect on the ability to assess clinical benefit of the new therapy in that setting. Because one should not ask patients to participate in a trial that cannot meet its scientific objectives, a crossover should not be permitted in trials in which OS is the most relevant primary objective.

Even when a crossover is not in the design of the trial, crossovers may become an issue in the OS evaluation of a trial when the trial results announced are based on a primary end point that is not OS (eg, PFS).

When such results are released and demonstrate that the experimental treatment arm is better with respect to the primary end point, patients on the standard arm will typically be offered the experimental treatment. One approach to handling this problem when evaluating OS is to censor the OS data for all patients on the standard treatment arm at the time crossing over is permitted. This allows an estimation of the OS differences between the treatment arms that is unaffected by the crossovers. However, the censoring will lessen the number of deaths observed in the trial, making the confidence interval for the OS differences wider than if there were no crossovers. This, combined with the fact that the trial may not have been designed with a sufficient sample size to examine OS differences, even without the censoring, may lead to inconclusive results concerning OS unless the OS difference between the treatment arms is quite large.

### **Experimental Therapy Previously Shown to Be Effective in Later-Line Settings**

In this setting, the experimental therapy has been shown to be effective as, say, a second-line treatment. Therefore, patients on the control (standard) arm in a first-line RCT testing the experimental therapy should be given the experimental treatment when second-line treatment is indicated because this is part of standard care. Although this type of crossover will again attenuate the OS difference between the treatment arms, we would argue that the observed attenuated difference is the relevant one to assess the clinical benefit of the experimental treatment as a first-line treatment. This is because the clinical question is whether moving the therapy up to a less advanced disease setting improves OS over the current standard in which the therapy is given later.

Besides crossovers, there can be trials in which different subsequent therapies are given in the treatment arms because of the nature of the experimental and standard treatments. As with crossovers, the OS difference can possibly be attenuated when compared with cases in which the same subsequent therapies are given. In this situation, we would again argue that the observed, possibly attenuated, OS difference is the appropriate measure of clinical benefit because it properly reflects the clinical reality of available subsequent treatments for the experimental and standard treatments.

#### EXAMPLES

Two randomized trials of capecitabine + ixabepilone versus capecitabine for metastatic breast cancer patients previously treated with anthracyclines and taxanes showed improved PFS with the ixabepilone but little or no improvement in OS in that population overall.<sup>8-11</sup> Although patients were not crossed over to ixabepilone, it is possible that the active therapies that a majority of the patients receive after progression lessen an OS benefit that would have been seen if active subsequent therapies did not exist or were not given, or it is possible that, even without these subsequent therapies, the PFS benefits seen would not have translated into an OS benefit. Regardless of which is the case, the lack of OS benefit seen in these trials suggests that this combination is not useful in that overall population.

A randomized trial of irinotecan + fluorouracil + leucovorin (IFL) versus IFL + bevacizumab for previously untreated metastatic colon cancer demonstrated a clinically meaningful OS benefit for IFL + bevacizumab.<sup>12</sup> Although crossover to receive bevacizumab was not allowed, approximately half the patients received second-line treatments, including 25% receiving oxaliplatin. The fact that the

experimental regimen had an OS benefit, even with standard second-line treatments, shows that it has clinical benefit in this setting.

In a trial of trastuzumab + anastrozole versus anastrozole for postmenopausal women with HER2-positive and hormone receptor-positive metastatic breast cancer showed improved PFS but no improvement in OS.<sup>13</sup> However, 70% of the patients on the anastrozole alone treatment arm crossed over to receive trastuzumab at progression. Therefore, it is impossible to say whether the trastuzumab is offering OS benefits in this setting.

In a trial of sorafenib versus placebo for metastatic renal cell carcinoma,<sup>14</sup> there was not a predesigned crossover at time of progression, but at the time the positive PFS results for the trial were announced, all patients on the placebo arm were offered sorafenib, with approximately half the patients receiving it. Little or no OS benefit for the sorafenib was seen. However, an additional analysis that censored the data from the placebo patients at the time the patients were allowed to cross over showed an OS benefit, suggesting the sorafenib offers an OS benefit in this setting.<sup>14</sup>

Even with a predesigned crossover, it is possible for an experimental agent to show OS benefits when it is active, and giving it earlier is better than giving it later. For example, a trial of sunitinib versus placebo in patients with advanced GIST who were resistant to imatinib showed a large improvement in OS.<sup>15</sup> This improvement in OS was seen even though approximately 80% of the patients in the placebo arm crossed over to the sunitinib at the time of progressive disease.

Following US Food and Drug Administration approval of herceptin for treatment in HER2-positive metastatic breast cancer, National Surgical Adjuvant Breast and Bowel Project (NSABP) and North Central Cancer Treatment Group (NCCTG) conducted trials to assess the efficacy of herceptin in the adjuvant setting. The joint analysis of the two studies demonstrated OS benefit for adjuvant herceptin,<sup>16,17</sup> even though a substantial proportion of the standard-arm patients would have been expected to receive herceptin on relapse as part of their standard therapy for metastatic disease. The hazard ratio observed in the trial accurately reflects the clinical benefit of the adjuvant herceptin, even though we can imagine that the hazard ratio would have been more dramatic if herceptin had not been used as a subsequent therapy for the standard-arm patients.

#### DISCUSSION

We have focused in this commentary on OS differences as the measure of clinical benefit. In some settings, such as early-stage disease, improvement in intermediate end points such as disease-free survival or recurrence-free survival (RFS) may represent clinical benefit by themselves. This implies that one believes that there would be clinical benefit for the experimental treatment and it would be recommended for general use, even if there were absolutely no OS differences between the treatment arms. As a possible example, consider the trial that demonstrated a large RFS advantage of imatinib over placebo as adjuvant treatment for GIST.<sup>18</sup> Because the death rates were low in both arms of this trial, it is unlikely that there is any clinical benefit in terms of OS of giving this agent early versus at recurrence. Conversely, if one views the large RFS difference as directly indicating clinical benefit of this relatively nontoxic agent, then the trial has demonstrated that giving the agent early is preferred. In other situations, alternative end points that directly measure quality of life or organ

preservation (eg, larynx preservation in laryngeal cancer<sup>19</sup>) can be used to quantify clinical benefit.

The use of an intermediate end point to show direct clinical benefit should not be confused with its use as a surrogate outcome for OS. In the latter situation, one uses the results observed on the intermediate end point to predict the OS results that would be observed if the patients were followed longer.<sup>20</sup> The predictions must be based on statistical models fit to outcome data from previously completed trials of similar agents in the same disease setting.<sup>21,22</sup> The predictions should be made in the context of whatever subsequent therapies are used in the disease setting and not for what the OS results would be if such subsequent therapies did not exist; this point seems to have sometimes been misunderstood.<sup>5</sup> If the modeling assumptions are correct, one can obtain information about the clinical benefit of the experimental agent more quickly than waiting for the mature OS data. However, since the assumptions will always involve an extrapolation for application to an experimental agent, it is important to follow the patients for the OS end point, even if the intermediate end point results are definitive and released. The timing of the release of any information on the intermediate end point needs to take into account the expected number of patients that will cross over to the experimental treatment and the effect this will have on the ability to eventually determine the OS difference between the treatment arms.

In summary, when there is no intermediate end point available as a measure of direct clinical benefit, RCTs should be designed with an appropriate sample size to detect clinically meaningful differ-

ences in OS. We make this recommendation even though there may be effective subsequent therapies that the patients will receive. When a new experimental therapy is undergoing definitive evaluation and has not previously demonstrated clinical benefit in later-line treatment in the same disease, trial designs with standard arm crossover to the experimental treatment should be avoided when possible. At a minimum, subsequent therapies the patients receive should be recorded by treatment arm. Interim monitoring on intermediate end points should not be used for superiority of the experimental treatment<sup>23</sup> (but may be used for futility/inefficacy<sup>24</sup>). After a trial is completed, the OS difference observed represents the relevant clinical benefit of the experimental treatment provided that the subsequent therapies used in both treatment arms follow the current standard of care.

#### AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

#### AUTHOR CONTRIBUTIONS

**Conception and design:** Edward L. Korn, Boris Freidlin, Jeffrey S. Abrams

**Manuscript writing:** All authors

**Final approval of manuscript:** All authors

#### REFERENCES

- Cannistra SA: Evaluating new regimens in recurrent ovarian cancer: How much evidence is good enough? *J Clin Oncol* 28:3101-3103, 2010
- Korn EL, Freidlin B: Causal inference for definitive clinical end points in a randomized clinical trial with intervening nonrandomized treatments. *J Clin Oncol* 28:3800-3802, 2010
- Schwartz D, Lellouch J: Explanatory and pragmatic attitudes in therapeutical trials. *J Chron Dis* 20:637-648, 1967
- Di Leo A, Bleiberg H, Buyse M: Overall survival is not a realistic end point for clinical trials of new drugs in advanced solid tumors: A critical assessment based on recently reported phase III trials in colorectal and breast cancer. *J Clin Oncol* 21:2045-2047, 2003
- Sargent DJ, Hayes DF: Assessing the measure of a new drug: Is survival the only thing that matters? *J Clin Oncol* 26:1922-1923, 2008
- Fleming TR, Rothmann MD, Lu HL: Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol* 27:2874-2880, 2009
- Blanke CD, Rankin C, Demetri GD, et al: Phase III randomized, intergroup trial assessing imatinib mesylate at two dose levels in patients with unresectable or metastatic gastrointestinal stromal tumors expressing the kit receptor tyrosine kinase: S0033. *J Clin Oncol* 26:626-632, 2008
- Thomas ES, Gomez HL, Li RK, et al: Ixabepilone plus capecitabine for metastatic breast cancer progressing after anthracycline and taxane treatment. *J Clin Oncol* 25:5210-5217, 2007
- Thomas ES: Ixabepilone plus capecitabine for metastatic breast cancer progressing after anthracycline and taxane treatment. *J Clin Oncol* 26:2223, 2008
- Roché H, Conte P, Perez EA, et al: Ixabepilone plus capecitabine in metastatic breast cancer patients with reduced performance status previously treated with anthracyclines and taxanes: A pooled analysis by performance status of efficacy and safety data from 2 phase III studies. *Breast Cancer Res Treat* 125:755-765, 2011
- Sparano JA, Vrdoljak E, Rixe O, et al: Randomized phase III trial of ixabepilone plus capecitabine versus capecitabine in patients with metastatic breast cancer previously treated with an anthracycline and a taxane. *J Clin Oncol* 28:3256-3263, 2010
- Hurwitz H, Fehrenbacher L, Novotny W, et al: Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med* 350:2335-2342, 2004
- Kaufman B, Mackey JR, Clemens MR, et al: Trastuzumab plus anastrozole versus anastrozole alone for the treatment of postmenopausal women with human epidermal growth factor receptor 2-positive, hormone receptor-positive metastatic breast cancer: Results from the randomized phase III TAnDEM Study. *J Clin Oncol* 27:5529-5537, 2009
- Escudier B, Eisen T, Stadler WM, et al: Sorafenib for treatment of renal cell carcinoma: Final efficacy and safety results of the phase III treatment approaches in Renal Cancer Global Evaluation Trial. *J Clin Oncol* 27:3312-3318, 2009
- Demetri GD, van Oosterom AT, Garrett CR, et al: Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: A randomised controlled trial. *Lancet* 368:1329-1338, 2006
- Romond EH, Perez EA, Bryant J, et al: Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 353:1673-1684, 2005
- Perez EA, Romond EH, Suman VJ, et al: Updated results of the combined analysis of NCCTG N9831 and NSABP B-31 adjuvant chemotherapy with/without trastuzumab in patients with HER2-positive breast cancer. *J Clin Oncol* 25:6s, 2007 (suppl; abstr 512)
- Dematteo RP, Ballman KV, Antonescu CR, et al: Adjuvant imatinib mesylate after resection of localized, primary gastrointestinal stromal tumour: A randomised, double-blind, placebo-controlled trial. *Lancet* 373:1097-1104, 2009
- Forastiere AA, Goepfert H, Maor M, et al: Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. *N Engl J Med* 349:2091-2098, 2003
- Ocaña A, Amir E, Vera F, et al: Addition of bevacizumab to chemotherapy for treatment of solid tumors: Similar results but different conclusions. *J Clin Oncol* 29:254-256, 2011
- Korn EL, Albert PS, McShane LM: Assessing surrogates as trial endpoints using mixed models. *Stat Med* 24:163-182, 2005
- Burzykowski T, Molenberghs G, Buyse M (eds): *The Evaluation of Surrogate Endpoints*. New York, NY, Springer, 2005
- Cannistra SA: The ethics of early stopping trials: Who is protecting whom? *J Clin Oncol* 22:1542-1545, 2004
- Goldman B, LeBlanc M, Crowley J: Interim futility analysis with intermediate endpoints. *Clin Trials* 5:14-22, 2008