Failing to improve overall survival because post-protocol survival is long: fact, myth, excuse or improper study design?

Vinay Prasad · Andrae Vandross

Received: 26 December 2013 / Accepted: 17 January 2014 / Published online: 29 January 2014
© Springer-Verlag Berlin Heidelberg 2014

Introduction

Recently, overall survival results were published for the AXIS study (Motzer et al. 2013), a large clinical trial comparing axitinib to sorafenib as second line treatment for patients with advanced or metastatic renal cell carcinoma. Although an earlier publication (Rini et al. 2011) showed a median progression-free survival (PFS) benefit that favored axitinib (6.7 vs 4.7 months, hazard ratio 0.665; $p < 0.0001$), the final results showed no difference in overall survival (OS) [median 20.1 months (95 % CI 16.7–23.4) versus 19.2 months (17.5–22.3) (hazard ratio [HR] 0.969, $p = 0.3744$]. Since the study did not permit crossover, the authors attempted to explain the negative outcome with this disclaimer, “being able to show a survival advantage is especially challenging when survival post-progression is long, that is, 12 months or longer, because the variability added during a long survival post-progression might dilute the ability to detect statistically significant difference in overall survival (Motzer et al. 2013).”

There is a widespread belief in oncology that when post-progression survival is long, we cannot ask trials to state an overall survival endpoint. An abstract from the 2012 COSA-IPOS (Clinical Oncology Society of Australia–International Psycho-Oncology Society) Joint Scientific Meeting states, “although improvements in overall survival

(OS) remain the Oncologist’s treatment goal, because effective post-trial therapy exists, true OS gains are increasingly difficult to quantify (Alam and Park 2012).” Regarding a study randomizing 370 patients with multiple myeloma to 20 doses of bortezomib after autologous stem cell transplantation, the authors note, “No difference in OS was seen, and this could be due to the fact that treatment at progression today is very effective and there are many treatment options (Mellqvist et al. 2013).”

The idea here is simple: If a drug improves PFS 2, 3 or 4 months while on protocol, but subsequent survival is 12, 14 or 24 months (off protocol), it becomes difficult to preserve that benefit as a statistically significant finding. While this hypothesis is a plausible explanation for negative OS results, an alternative is that survival post-protocol (SPP) is different between groups, favoring the control arm; in this case, an OS benefit was missed not because of power issues, but because any such benefit does not exist. Here, we review the evidence for the claim that long SPP masks an OS benefit and highlight reasons why OS should remain the endpoint of clinical trials even when SPP is expected to be lengthy.

The data

A seminal 2009 investigation (Broglia and Berry 2009) demonstrates how a trial might conclude favorably regarding a PFS benefit, while failing to confirm an benefit in OS. Broglia and Berry (2009) performed multiple simulations making the following assumptions: median OS is a sum of median PFS and median SPP, PFS in one arm was 6 months and another 9 months (3 month improvement) and, finally, median SPP can vary (0, 3, 6, 9, 12, 18 and 24 months). The authors found that as the median SPP increased, the

V. Prasad (✉)
Medical Oncology Branch, National Cancer Institute, National
Institutes of Health, 10 Center Dr. 10/12N226, Bethesda, MD
20892, USA
e-mail: vinayak.prasad@nih.gov

A. Vandross
Department of Medicine, Yale University School of Medicine,
New Haven, CT, USA

power to detect a significant improvement in OS also decreases. If a 3-month PFS benefit passed the threshold of significance ($p < 0.05$), the probability of detecting an OS benefit decreases dramatically from 33 % when SPP is short (2 months) to 8 % when SPP is long (24 months). However, the central limitation to this paper, as well as, more generally, any justification derived from modeling, is that it rests on an unproven assumption: median SPP is the same in treatment and control arms.

Ideally, empirical data would be able to adjudicate this issue, but empirical studies on this topic have methodological issues. In August 2012, the Decision Support Unit (DSU) of The National Institute for Health and Clinical Excellence (NICE) performed a systematic review of 19 studies that examined the relationship between progression-free and overall survival in metastatic or advanced cancer (Davis and Cantrell 2012). The group found wide variability in proffered correlation coefficients between PFS and OS, ranging in size from 0.24 to 0.89. The authors conclude that correlation between PFS and OS is generally tenuous, but may be useful for some specific cancers and some classes of agent. Yet, even this interpretation is generous. Another way to look at these data is to examine the range of variability in the PFS to OS correlation. One study (Wilkerson and Fojo 2009), comparing the PFS to OS in 66 publications, found that for marginal PFS gains (0–2 months), OS could change anywhere from –2 to +8 months. In a review of 18 bevacizumab trials (Ocaña et al. 2011), the median absolute difference in PFS ranged from +0.4 to +5.9 months (median 1.9), while OS ranged from –1.7 to +7.8 months (mean 1.2). In short, even if the proffered correlations are true, they are just that: correlations (Baker and Kramer 2003). In any given case, we have no idea what OS results would be. Under what circumstances is the correlation robust, and when we should view PFS gains more cautiously? These remain crucial and open questions.

There is an additional and overarching deficiency in investigations of the correlation between PFS and OS. Specifically, the trials included in these studies did not always achieve significant findings for both metrics. Thus, correlations are being made between one significant outcome (typically PFS) and another outcome, which is not significant (typically OS). Nonsignificant outcomes may be trending toward significance and ultimately valid, but alternatively, they may be distortions of the true outcome. Were the trial to continue or be adequately powered, nonsignificant OS results may remain nonsignificant, or trend toward harm. Because these correlation studies typically include trials with one (but not usually 2) nonsignificant outcomes, these analyses may be further affected by publication or selective reporting bias (Kyzas et al. 2005). For instance, trials with small PFS benefit, but no or detrimental changes in OS

may not be published, published in lesser journals or published only after significant delay (Ioannidis 1998). Trials with little or no change in PFS may be halted, unpublished or missed by meta-analysts, and the effect of these medications on mortality may be unknown. Null median PFS results do not necessarily imply null OS, as was seen in trials of Sipuleucel-T (Kantoff et al. 2010) and ipilimumab (Hodi et al. 2010). In short, the trials included in correlation studies of PFS and OS likely represent only a select slice of studies that could capture both outcomes.

The E2100 trial

If the argument of long SPP sounds familiar, it may be because of the bevacizumab E2100 study. The study randomized 722 women with metastatic breast cancer, who had not received cytotoxic therapy, to paclitaxel with or without bevacizumab. The trial famously found a markedly prolonged progression-free survival favoring bevacizumab (median, 11.8 vs. 5.9 months; hazard ratio for progression, 0.60; $p < 0.001$) but no improvement in overall survival (median, 26.7 vs. 25.2 months; hazard ratio, 0.88; $p = 0.16$) (Miller et al. 2007). To make sense of the data, the authors evoked SPP, “Patients with metastatic breast cancer frequently receive multiple therapies during the course of their disease. Data on treatment administered after progression were not collected in this trial, precluding an exploratory analysis of the influence of subsequent therapy on overall survival (Miller et al. 2007).”

The dramatic findings of the E2100 led the US Food and Drug Administration (FDA) to grant bevacizumab accelerated approval in combination with paclitaxel for first-line treatment of metastatic breast cancer in 2008. Two years later, the results of two additional randomized trials were reported. The Avastin Plus Docetaxel (AVADO) study compared docetaxel with or without bevacizumab at two different doses (7.5 and 15 mg/kg) (Miles et al. 2010). AVADO found no difference in OS among the three arms, and median PFS was only marginally improved. RIBBON-1 evaluated bevacizumab in combination with the investigator’s choice of one of several chemotherapy regimens. RIBBON-1 found a slight improvement in PFS favoring the arms containing bevacizumab (~1–3 months), but no difference in OS (Robert et al. 2011). A pooled analyses of these 3 trials (O’Shaughnessy et al. 2010) concluded that bevacizumab along with chemotherapy improved PFS 6.7–9.2 mo (HR = 0.64, $p < 0.0001$), but did not change OS. Improving PFS, while adding toxicity, and failing to improve OS does not define a marginally effective therapy, but constitutes a harmful one. As such the FDA revoked bevacizumab’s approval in metastatic breast cancer.

Conclusion

Another way to conceptualize how a long SPP affects statistical power is simply to realize that small gains are harder to detect when OS is long. Increasing survival by 3 months from 3 to 6 is a 100 % gain, while increasing survival from 24 to 27 months is only a 13 % improvement. While the principle long SPP is surely true, what does not follow is that all PFS increases translate into increases in OS. They may or may not, as with many intermediate and surrogates endpoints (Svensson 2013), we simply do not know. Whether PFS is a more reliable intermediate of OS when SPP is long or short also remains unknown. For the time being, regardless of the length of SPP, if OS survival is not improved statistically, we must continue to hold the null hypothesis.

A 2013 ASCO draft position statement (Defining Clinically Meaningful Outcomes 2013) asks us to raise the bar for novel cancer therapies for metastatic cancer. Novel therapies should improve OS by 20, 25 or 50 %, depending on the cancer type (Defining Clinically Meaningful Outcomes 2013). This push is in contrast to trends over the last decade. PFS (and time to progression) is increasingly used as endpoint of oncology randomized trials, constituting 11 % of all cancer trials from 1995 to 2004, and 25 % of such trials from 2005 to 2009 (Kay et al. 2012). From 2005 to 2007, 23 % of new drug approvals were based on PFS or time to progression (Sridhara et al. 2010). While many authors (Booth and Eisenhauer 2012; Driscoll and Rixe 2009; Cheema and Burkes 2013) have criticized increasing use of PFS and argued that OS should remain the preferred endpoint, for indications with long SPP, improving OS has been asserted to be difficult and not worth the effort or expense (Lebwohl et al. 2009). We disagree. The cost to society when ultimately ineffective practices gain widespread use is often an order of magnitude greater than the cost of conducting proper studies upfront (Prasad et al. 2012; Prasad and Cifu 2011). Knowing that small gains in survival are real, especially when survival is long, remains best adjudicated by adequately powered trials, and not continued faith in PFS.

Conflict of interest None.

References

- Alam MDP, Park SH (2012) Progression free survival vs overall survival: an example from randomised phase III trial with Axitinib (AXIS) in metastatic renal cell carcinoma. COSA-IPOS Joint Scientific Meeting. COSA Posters: Clinical Research
- Baker S, Kramer B (2003) A perfect correlate does not a surrogate make. *BMC Med Res Methodol* 3:16
- Booth CM, Eisenhauer EA (2012) Progression-free survival: meaningful or simply measurable? *J Clin Oncol* 30:1030–1033
- Broglio KR, Berry DA (2009) Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst* 101:1642–1649
- Cheema PK, Burkes RL (2013) Overall survival should be the primary endpoint in clinical trials for advanced non-small-cell lung cancer. *Curr Oncol* 20:e150–e160
- Davis STP, Cantrell A (2012) A review of studies examining the relationship between progression-free survival and overall survival in advanced or metastatic cancer. (Accessed Aug 30, 2013, at <http://www.nicesdu.org.uk/PFSOS%20Report.FINAL.06.08.12.pdf>)
- Defining Clinically Meaningful Outcomes: ASCO recommendations to raise the bar for clinical trials. (Accessed June 11, 2013, at <http://www.asco.org/advocacy-practice/clinically-meaningful-outcomes>)
- Driscoll JJ, Rixe O (2009) Overall survival: still the gold standard: why overall survival remains the definitive end point in cancer clinical trials. *Cancer J* 15:401–405
- Hodi FS, O'Day SJ, McDermott DF et al (2010) Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 363:711–723
- Ioannidis JP (1998) Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA J Am Med Assoc* 279:281–286
- Kantoff PW, Higano CS, Shore ND et al (2010) Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *N Engl J Med* 363:411–422
- Kay A, Higgins J, Day AG, Meyer RM, Booth CM (2012) Randomized controlled trials in the era of molecular oncology: methodology, biomarkers, and end points. *Ann Oncol* 23:1646–1651
- Kyzas PA, Loizou KT, Ioannidis JPA (2005) Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* 97:1043–1055
- Lebwohl D, Kay A, Berg W, Baladi JF, Zheng J (2009) Progression-free survival: gaining on overall survival as a gold standard and accelerating drug development. *Cancer J* 15:386–394
- Mellqvist U-H, Gimsing P, Hjertner O et al (2013) Bortezomib consolidation after autologous stem cell transplantation in multiple myeloma: a Nordic Myeloma Study Group randomized phase 3 trial. *Blood* 121:4647–4654
- Miles DW, Chan A, Dirix LY et al (2010) Phase III study of bevacizumab plus docetaxel compared with placebo plus docetaxel for the first-line treatment of human epidermal growth factor receptor 2—negative metastatic breast cancer. *J Clin Oncol* 28:3239–3247
- Miller K, Wang M, Gralow J et al (2007) Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med* 357:2666–2676
- Motzer RJ, Escudier B, Tomczak P et al (2013) Axitinib versus sorafenib as second-line treatment for advanced renal cell carcinoma: overall survival analysis and updated results from a randomised phase 3 trial. *Lancet Oncol* 14:552–562
- Ocaña A, Amir E, Vera F, Eisenhauer EA, Tannock IF (2011) Addition of bevacizumab to chemotherapy for treatment of solid tumors: similar results but different conclusions. *J Clin Oncol* 29:254–256
- O'Shaughnessy J, Gray RJ et al (2010) A meta-analysis of overall survival data from three randomized trials of bevacizumab (BV) and first-line chemotherapy as treatment for patients with metastatic breast cancer (MBC). *J Clin Oncol* 28(suppl):115s abstr 1005
- Prasad V, Cifu A (2011) Medical reversal: why we must raise the bar before adopting new technologies. *Yale J Biol Med* 84:471–478
- Prasad V, Cifu A, Ioannidis JPA (2012) Reversals of established medical practices: evidence to abandon ship. *JAMA J Am Med Assoc* 307:37–38
- Rini BI, Escudier B, Tomczak P et al (2011) Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet* 378:1931–1939

- Robert NJ, Dieras V, Glaspy J et al (2011) RIBBON-1: randomized, double-blind, placebo-controlled, phase III trial of chemotherapy with or without bevacizumab for first-line treatment of human epidermal growth factor receptor 2-negative, locally recurrent or metastatic breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol* 29:1252–1260
- Sridhara R, Johnson JR, Justice R, Keegan P, Chakravarty A, Pazdur R (2010) Review of oncology and hematology drug product approvals at the US food and drug administration between July 2005 and December 2007. *J Natl Cancer Inst* 102:230–243
- Svensson S, Menkes DB, Lexchin J (2013) Surrogate outcomes in clinical trials: a cautionary tale. *JAMA Int Med* 173:611–612
- Wilkerson J, Fojo T (2009) Progression-free survival is simply a measure of a drug's effect while administered and is not a surrogate for overall survival. *Cancer J* 15:379–385