Chronological Trends in Progression-Free, Overall, and Post-Progression Survival in First-Line Therapy for Advanced NSCLC

Peer-reviewed author version

Rutkowski, Jacek; Saad, Everardo D.; BURZYKOWSKI, Tomasz; BUYSE, Marc & Jassem, Jacek (2019) Chronological Trends in Progression-Free, Overall, and Post-Progression Survival in First-Line Therapy for Advanced NSCLC. In: JOURNAL OF THORACIC ONCOLOGY, 14(9), p. 1619-1627.

DOI: 10.1016/j.jtho.2019.05.030
Handle: http://hdl.handle.net/1942/29995

Chronological Trends in Progression-free, Overall and Post-progression Survival in First-line Therapy for Advanced Non-small-cell Lung Cancer

Jacek Rutkowski, Everardo D. Saad, Tomasz Burzykowski, Marc Buyse, Jacek Jassem

PII: S1556-0864(19)30415-0

DOI: https://doi.org/10.1016/j.jtho.2019.05.030

Reference: JTHO 1425

To appear in: *Journal of Thoracic Oncology*

Received Date: 23 January 2019

Revised Date: 14 May 2019

Accepted Date: 16 May 2019

Please cite this article as: Rutkowski J, Saad ED, Burzykowski T, Buyse M, Jassem J, Chronological Trends in Progression-free, Overall and Post-progression Survival in First-line Therapy for Advanced Non-small-cell Lung Cancer, *Journal of Thoracic Oncology* (2019), doi: https://doi.org/10.1016/j.jtho.2019.05.030.

Chronological Trends in Progression-free, Overall and Post-progression Survival in First-line Therapy for Advanced Non-small-cell Lung Cancer

Jacek Rutkowski,[1*] Everardo D. Saad,[2,3*] Tomasz Burzykowski,[3,4] Marc Buyse,[4,5] and Jacek Jassem[1]

[1]Medical University of Gdańsk, Gdańsk, Poland, [2]Dendrix Research, Sao Paulo, Brazil; [3]International Drug Development Institute, Louvain-la-Neuve, Belgium, [4]Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University, Diepenbeek, Belgium, and [5]International Drug Development Institute, San Francisco, CA

*Everardo D. Saad and Jacek Rutkowski contributed equally to this work.

Address for correspondence

Jacek Jassem, MD, PhD

Department of Oncology and Radiotherapy

Medical University of Gdansk

Dębinki 7 St., 80-211 Gdansk, Poland

Tel. +48 58 3492270, e-mail: jjassem@gumed.edu.pl

Founding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosure: Dr. Saad reports other from International Drug Development Institute, outside the submitted work. The remaining authors declare no conflict of interest.

ABSTRACT

BACKGROUND: There is a debate about the merits of progression-free survival (PFS) versus overall survival (OS) as primary end points in non-small-cell lung cancer (NSCLC). It has been postulated that post-progression therapy may influence OS in both arms. To investigate this issue, we analyzed chronological trends in PFS and OS in advanced NSCLC using restricted mean survival times (RMSTs).

METHODS: We digitized survival curves from first-line phase III trials published between 1998 and 2015 in 13 leading journals to compute RMSTs for PFS and OS at three truncation landmarks (5, 12, and 18 months).

RESULTS: Among the 161 trials identified, RMSTs could be computed for both endpoints in 102, 97 and 82 trials for the 5, 12 and 18 months truncation landmarks, respectively. Post-progression survival (PPS) in the control arm, quantified as mean OS minus mean PFS truncated at 18 months, was on average 3.3 months between 1998 and 2003, 4.4 months between 2004 and 2009, and 5.4 months between 2010 and 2015. This increase was due to increasing RMST for OS over time, with no increase in RMST for PFS. The average within-trial difference in RMSTs between experimental and control arm was close to 0 for OS and less than 1 month for PFS.

CONCLUSIONS: There is a progressive increase in PPS in NSCLC trials, likely from salvage therapy. These results question both PFS and OS as sensitive end points in first-line trials, but suggest that the outlook for patients is improving regardless of within-trial gains.

INTRODUCTION

There has been significant progress in the treatment of non-small cell lung cancer (NSCLC) from the identification and targeting of key molecular alterations—such as epidermal growth factor receptor (EGFR) activating mutations, and ALK-EML4 and ROS1 alterations [1-5]—as well as from the use of immunotherapy.[6-9] In parallel to these advances, there has been a debate about the adequacy of overall survival (OS) versus tumor-based end points —such as progression-free survival (PFS)—as the most suitable and reliable primary end points for phase III trials in various tumor types, including NSCLC.[10] One of the issues raised in this debate is the possibility that OS may be considerably influenced by treatments administered after disease progression, and cross-over treatments in particular, which may dilute the treatment effect.[10, 11] Corroborating this concern, Hotta *et al* [12] and Hayashi *et al* [13] have shown that both median OS and post-progression survival (PPS) have increased along the years, with absent or less pronounced simultaneous increases in median PFS in first-line trials in advanced NSCLC. However, assessment of the relationship between trends in PFS and OS over time are best made using mean survival times rather than the medians, which do not have additive properties.[14] In order to assess that relationship, and in the attempt to quantify the magnitude of increase in PPS over the years, we investigated trends in OS and PFS results during 18 years of clinical research in NSCLC by using restricted mean survival times (RMSTs).

MATERIAL AND METHODS

Search strategy

We used PubMed, the medical subject headings ‘lung neoplasms’ and ‘drug therapy’, and the filter option ‘randomized controlled trials’ to search for clinical trials on

systemic anticancer therapies published between January 1, 1998, and December 31, 2015 in 13 leading journals in the field of clinical oncology or lung cancer specifically (*Annals of Oncology*, *British Journal of Cancer*, *Cancer*, *Chest*, *Clinical Cancer Research*, *European Journal of Cancer*, *Journal of Clinical Oncology*, *Journal of the National Cancer Institute*, *Journal of Thoracic Oncology*, *Lancet Oncology*, *Lung Cancer*, *The Lancet* and *The New England Journal of Medicine*). We only included phase III trials whose main results had been published in the study period. We only included first-line treatment trials or those testing a maintenance strategy after the first line. We excluded trials on supportive therapy alone, papers reporting combined analyses of two or more separate trials already included in the study, randomized phase II trials, and companion studies on correlative biology or prognostic factors. Finally, we excluded randomized trials for which no phase was explicitly stated if they had fewer than 100 evaluable patients per arm. With a cut-off date of June 2016, we tried to obtain published OS results in later publications for the trials that fell within the study period and for which OS results were immature when the main paper was published.

Data collection and definitions

For each selected trial, we retrieved the main paper reporting the efficacy results and abstracted the general characteristics of the trial and relevant data on OS and PFS, time to tumor progression (TTP) or time to treatment failure (TTF). Despite their slightly different meanings, we treated PFS, TTP and TTF together due to their frequently overlapping definitions by investigators.[15] When not explicitly stated, we considered the primary end point as the metric used for sample-size calculation or the end point first cited in the ‘Methods’ or ‘Results’ section of papers. We defined

trials with molecular selection of patients as those for which such selection was an inclusion criterion.

Statistical analysis

The (unrestricted) mean survival (PFS or OS) of patients under a certain treatment can be computed as the area under the Kaplan-Meier survival curve corresponding to that treatment when the last patient under observation suffers the event of interest. When the last patient is censored, the (unrestricted) mean survival cannot be computed. In that case, the most frequently employed alternative is to compute the RMST, which is the mean survival in the time interval between time 0 to a truncation landmark, *t*. For the comparison of the RMST in two arms of a given trial, *t* is usually chosen as the shortest time for which survival data are available for both arms. Since our goal was to estimate PPS on the basis of (restricted) mean PFS and OS for a large number of trials, we defined various arbitrary truncation landmarks to be used across trials. For each trial, we identified the published Kaplan-Meier curves for PFS (or TTP/TTF when this was the case) and OS, and digitized such curves using CurveSnap.[16] We used the same density of points (10/month) for all curves, thus ensuring the same precision. Each curve was digitized in triplicate, once by each different individual, and one of the authors (JR) selected manually the digitization that appeared to better represent the curve shape. Using the digitized curves, we computed RMSTs automatically, using software code created for the study purpose. We chose three truncation landmarks (5, 12 and 18 months) based on the number of curves that would provide sufficient information until such landmarks. We computed RMST from the available PFS and OS curves from each trial and the average of these RMSTs for each year of publication. The RMSTs thus computed indicate the

average (progression-free or overall) survival of patients in the control arm over the period comprised by *t*. For the description of chronological trends, we used the control arm in each trial for simplicity, under the assumption that improvements obtained from clinical trials are reflected in the evolution of the control arms of future trials. However, to provide a more complete view of chronological trends, we also performed some analyses using the experimental arms of each trial. This was straightforward in two-arm trials; for trials with more than two arms, we took the average RMST (for PFS and OS) of the experimental arms. We quantified the magnitude of mean PFS and OS gain as the difference in RMSTs between the control and experimental arms. For the illustrative analysis involving median PFS and OS times, we used the same set of trials for which RMSTs could be computed for *t*=12 months, the landmark closer to the expected median OS in contemporary first-line trials. In order to account for the potential influence of cross-over on OS, we abstracted data from each trial about the presence of such cross-over in the control and one experimental arm. We then conducted sensitivity analyses of trends in PFS and OS over time according to the use of cross-over.

RESULTS

General characteristics of the phase III trials

The search yielded 161 phase III trials that were eligible for analysis. There were 144 trials testing a first-line therapy, and 17 in which the randomized intervention was maintenance or treatment extension after first line. A total of 73,850 patients were randomized to 353 treatment arms, with a median of 173 patients per arm. The yearly mean number of patients per trial increased until 2009, when it started to plateau or even decrease (Figure 1). This decrease was in part due to 11 trials with molecularly-

selected patients, all of which published after 2006: the median number of patients per arm in these trials was 115, compared with 181 in trials with no molecular selection. However, the decrease after 2009 may also have been due to increasing use of PFS as primary end point, as this usually requires fewer patients (Figure 2).

The number of arms per trial was two in 137 cases, three in 17, and four in seven trials (three of which with a factorial 2 x 2 design). Nineteen trials had a placebo-, observation- or best supportive care-alone arm. Three trials had two co-primary end points. In trials with a single primary end point, this was OS in 103 (64% of the total) trials; PFS (N=26), TTP (N=1) or TTF (N=1) in 28 (17.4%) cases (henceforward called PFS); response rate or clinical benefit rate in 16 (10.0%); quality of life or toxicity in 10 trials (6.2%); and PFS without grade 4 toxicity in one (0.6%). Over the 18 years spanned by the study, the use of PFS as primary end point progressively increased, with an accompanying decrease in the use of response or clinical benefit rates (Figure 2).

Time trends in mean and median OS and PFS times

For *t*=5 months, computation of RMSTs was possible for 145 trials in the case of OS, 113 trials in the case of PFS, and 102 trials for both of these end points. For *t*=12 months, the RMST could be computed in 140 trials for OS and 107 trials for PFS; computation of both RMSTs for the same trial was possible in 97 trials. Finally, for *t*=18 months, the corresponding numbers of trials for which the RMSTs could be computed were 122 for OS, 89 for PFS, and 82 for both end points. Table 1 shows a descriptive comparison between trials for which RMSTs could or could not be computed for PFS and OS truncated at 18 months. Trials for which RMSTs could be

computed were larger and published more recently, whereas the distribution of primary end points was similar. The trials for which RMSTs could be computed were also less likely to report gain in OS and more likely to report gain in PFS, but this may be related to the fact that these trials were published more recently (and hence used PFS as primary end point more frequently). Table 2 shows the number of trials for each 2-year period of interest for which both of these means could be computed.

Figure 3 displays the evolution over time of PFS and OS in control arms, considering the yearly average RMSTs with the 5, 12 and 18 months truncation landmarks; for illustration, the average median times for PFS and OS are also shown. Not surprisingly because of the short truncation landmark, there is little separation in the curves that depict the yearly average mean PFS and OS for t=5 months (Panel A). Such separation only becomes evident with the truncation landmarks of 12 (Panel B) and 18 (Panel C) months. Moreover, a chronological trend for an increase in mean OS with no accompanying increase in mean PFS becomes apparent with t=12 months, and more pronounced with t=18 months. When median times are considered, a trend for increasing yearly average median OS with no accompanying increase in PFS is also evident (Panel D). In this case, however, the degree of separation may be misleading, because it does not directly indicate PPS.

Table 3 shows average RMSTs for PFS and OS in control arms for each 6-year period spanned by the study, and according to the truncation landmarks of 5, 12 and 18 months. If the difference between RMSTs for OS and PFS over 18 months is used to quantify PPS, the average duration of PPS was 8.9-5.6=3.3 months in trials published between 1998 and 2003, 10.0-5.6=4.4 months in those published between

2004 and 2009, and 11.4-6.0=5.4 months in those published between 2010 and 2015 (see also Figure 3, Panel C).

Magnitude of mean PFS and OS gains over time

Considering the trials for which RMSTs for both PFS and OS could be computed, the average difference in RMSTs between experimental and control arms per 6-year period and overall can be seen in Table 4. On average across trials, and regardless of period or truncation landmark, there was essentially no gain in mean OS, whereas the average gain in mean PFS was always less than 1 month. For t=5 months, the differences in RMSTs between experimental and control arms within trials ranged from a loss of 1 month to a gain of 2 months for PFS, and from a loss of 0.4 month to a gain of 1.2 months for OS. For t=12 months, these differences ranged from a loss of 1.8 months to a gain of 3.3 months for PFS, and from a loss of 1.2 months to a gain of 1.9 months for OS. For t=18 months, these differences ranged from a loss of 1.9 months to a gain of 5.1 months for PFS, and from a loss of 1.6 months to a maximum gain of 2.7 months for OS. When only trials with a significant gain in OS are considered among those with sufficient data for analysis with t=18 months, the average gain in OS was 1.2 month; in contrast, this average was 0 among trials with no significant gain in OS.

The potential influence of cross-over

Cross-over was a feature of trial design or recommended by protocol in only six of the 161 trials overall. All these trials had two arms and had the first-line therapy as their main focus; in three of six cases, cross-over was recommended for both trial arms, whereas in the other three cross-over was from the control to the experimental

arm (N=2) or vice-versa. When reported (N=4), the percentage of patients from the control arm crossing-over to the same or similar agent as the experimental arm ranged from 70.0% to 94.6%. For the remaining 155 trials, any post-progression therapy was mentioned in 94 trials, 81 of which providing the percentages of patients receiving such therapy. The median percentage of post-progression therapy was 42% in both control and experimental arms, with a clear increase in this percentage over time (data not shown). Of the 94 trials in which post-progression therapy was mentioned, cross-over could be characterized in 36 cases by looking at the percentages of patients receiving the same or a similar agent as first-line therapy in the post-progression setting. The median percentage of patients crossing-over was 18.5% in the control arms and 19.1% in the experimental arms. Of the total of 42 trials with cross-over (whether or not by design), 10 reported a significant gain in OS (23.8%); among 119 trials without cross-over (or no mention thereof), a significant gain in OS was reported in 27 (22.7%).

We assessed the evolution over time of yearly average RMST for PFS and OS in control arms, considering the same 82 trials with sufficient data for the 18-month truncation landmark, according to the presence of cross-over. We observed similar trends for an increasing average OS and stable average PFS both in trials with no reported cross-over (N=55) and in those with cross-over (whether or not by design; N=27; data not shown). Among trials with cross-over, average PFS and OS according to period were 5.7 months and 9.4 months for 1998 to 2003 (PPS=3.7 months); 5.3 months and 10.9 months for 2004 to 2009 (PPS=5.6 months); and 5.9 months and 12.6 months for 2010 to 2015 (PPS=6.7 months). Corresponding averages for trials with no cross-over were 5.8 months and 9.0 months for 1998 to 2003 (PPS=3,2

months); 5.6 months and for 9.8 months 2004 to 2009 (PPS=4.2 months); and 6.1 months and 10.6 months for 2010 to 2015 (PPS=4.5 months).

Finally, we investigated trends in the average difference in RMSTs (with the 18-month truncation landmark) between experimental and control arms within trials according to the presence of cross-over. For trials with cross-over (N=27), the average difference in RMSTs for PFS and OS were 0.3 month and 0.3 month for 1998 to 2003; 1.0 month and 0.4 month for 2004 to 2009; and 1.5 month and 0 month for 2010 to 2015. Corresponding differences for trials with no cross-over were 0.5 month and 0 month for 1998 to 2003; 0.5 month and 0.2 month for 2004 to 2009; and 0.5 month and 0.2 month for 2010 to 2015. When the overall 18-year period is considered, the average difference in RMSTs for PFS and OS were 1.1 month and 0.2 month, respectively, for trials with cross-over, and 0.5 month and 0.2 month, respectively, for those without cross-over.

DISCUSSION

Our results disclose a progressive increase in mean OS with no obvious increase in mean PFS following first-line therapy for advanced NSCLC over the years. Although less appropriate from a methodological point of view, the same findings hold true when these chronological trends are assessed using medians. Moreover, the comparison of gains in mean PFS and OS times within trials show no substantial gain on average. Therefore, our results collectively suggest that the outlook for patients has clearly improved over the 18-year period of analysis, but not necessarily due to improvements within trials. Rather, our data suggest that the increase in OS over

time is more likely due to post-progression therapy than to the treatments tested within the trials.

In contrast with two previous smaller reviews that analyzed PFS and OS using medians,[12, 13] we ascertained trends in both end points using the RMST, a more appropriate measure in the attempt to investigate differences in duration of OS and PFS (*i.e*., the duration of PPS), [14] In a previous study, Trinquart *et al,*[17] computed RMSTs for the primary end point in 54 trials in several cancer types and therapeutic settings, in an attempt to compare treatment effects estimated by the hazard ratio (HR) and by differences and ratios of RMSTs. These authors pointed out that the HR may overestimate the treatment effect as measured by the ratio of RMST. Several other authors pointed to advantages of using RMSTs, including in settings where hazards are non-proportional.[14, 18, 19]

The extent to which our findings and conclusions are misleading due to the use of RMSTs is unknown. One of the alternatives—the use of medians—is seen by many as inadequate towards to goal of investigating PPS.[14] Another alternative—the computation of unrestricted means using observed data or parametric assumptions is either not feasible or subject to the choice of models. These considerations notwithstanding, confirmation of our findings would have important implications for trial design in NSCLC, especially with regard to the choice of the primary end point. There is an ongoing debate about which is the most adequate primary end point in several tumor types, including NSCLC. We have argued that OS is generally inadequate in settings for which there is salvage therapy,[20] which is increasingly the case in NSCLC. The current results seem to confirm this contention, because there

has been essentially no gain in mean OS within trials, with a maximum gain of 2.7 months. On the other hand, we and others have argued in favor of PFS as primary end point,[10, 11] but the current results also question whether PFS is a sensitive end point in this setting, because no substantial gain in mean PFS within trials could be noted on average; in this case, however, the maximum mean gain in PFS was 5.1 months (for *t*=18 months). Once again, the extent to which the doubts about the sensitivity of PFS and OS in our analysis are due to the use of RMSTs—as opposed to conventional metrics—remains unknown. As shown in Table 1, significant gains in PFS (as ascertained by conventional methods) were reported in 37% of trials, whereas only 23% reported a significant gain in OS.

Interestingly, the current results show a progressive increase in OS despite increasing use of PFS as primary end point (Figure 2) and no increased frequency of OS gain usually ascertained by the logrank test (data not shown). Moreover, the progressive increase in OS reported herein confirms previous analyses based on a smaller number of trials and on real-life data.[13, 21, 22] Finally, these results are in line with the various recent instances in which a very active agent (as judged by improvements in response rate or PFS) did not lead to improved OS, most likely because of cross-over after disease progression.[3, 4, 23]

Our study is subject to several limitations, the chief of which is the inability to compute RMSTs for all eligible trials. Several authors have called for more extended publication of results from phase III trials using RMSTs and other measures of treatment effect that may complement the HR.[14, 17, 18] It is hoped that increased publication of such results will allow for more complete analyses in the future. A

second limitation is the potential for publication bias. Our analysis was restricted to articles published in selected medical journals indexed in PubMed and within a limited time period. This might have led to identification of phase III trials that are systematically different from trials that have not been retrieved in our search. However, we have consciously made this restriction, as the journals included in the analysis report the vast majority of practice-influencing phase III trials. It is possible, therefore, that we have analyzed studies with gains in OS preferentially. Another limitation of our study relates to the inability to control for several variables that may confound the associations and trends observed. For example, many recent trials have included molecularly-selected patient populations with an improved prognosis; the influence of these or even more subtle differences in patient profile on the current results is uncertain. Such differences could possibly relate to tumor histology, performance status, and other prognostic or predictive factors. Likewise, improved trial methodology over the years may have influenced our results. Similarly, we cannot ascertain the extent to which stage migration, other differences in patient selection, or improvements in supportive care have influenced the apparent increase in OS over time. Finally, incomplete reporting of information on cross-over and our use of aggregated data—which precludes separate analyses of results among patients with and without cross-over—prevent us from reliably ascertaining the influence of cross-over on the trends for increasing OS without an accompanying trend for increasing PFS. Nevertheless, our sensitivity analyses suggest that the influence of cross-over is either not apparent or cannot be properly quantified, given the similarity of results for (1) evolution over time of PFS and OS in control arms and for (2) trends in average difference in RMSTs between experimental and control arms within trials, when trials with and without cross-over are compared.

Our results may not be relevant for immunotherapy trials, only two of which were represented in our sample.[24, 25] In some of the randomized trials of antibodies against the programmed death (ligand) 1 pathway, a significantly increased OS was not always accompanied by an increased PFS.[6, 9, 26] Since most immunotherapy trials were published after our analysis period, our results apply essentially to chemotherapy and targeted therapy. Despite these limitations, the current study is the first to assess RMSTs in a large number of trials within one specific setting.

Conclusions

There is a progressive increase in PPS in NSCLC trials, but this appears to be more likely a result of post-trial use of salvage therapy. These results question both PFS and OS as sensitive end points in first-line trials, but suggest that the outlook for patients is improving regardless of within-trial gains. It is hoped that population enrichment and the use of targeted agents and immunotherapy with markedly improved efficacy will allow for treatment effects that are larger than the ones currently observed. If so, OS will continuously improve for patients inside and outside of clinical trials.

ACKNOWLEDGEMENTS

The authors are grateful to Cezary Komarzeniec for his help in the preparation of the script for automated computation of restricted mean survival times, and to Kamila Bemben, Kamil Buczkowski and Zofia Alchimowicz for their help in digitizing the curves.

REFERENCES

1. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009;361:947-957.

2. Kwak EL, Bang YJ, Camidge DR, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 2010;363:1693-1703.

3. Rosell R, Carcereny E, Gervais R, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* 2012;13:239-246.

4. Shaw AT, Kim DW, Nakagawa K, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* 2013;368:2385-2394.

5. Sequist LV, Yang JC, Yamamoto N, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013;31:3327-3334.

6. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373:1627-1639.

7. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373:123-135.

8. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* 2016.

9. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* 2017;389:255-265.

10. Soria JC, Massard C, Le Chevalier T. Should progression-free survival be the primary measure of efficacy for advanced NSCLC therapy? *Ann Oncol* 2010;21:2324-2332.

11. Saad ED, Katz A, Hoff PM, et al. Progression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature. *Ann Oncol* 2010;21:7-12.

12. Hotta K, Kiura K, Fujiwara Y, et al. Role of survival post-progression in phase III trials of systemic chemotherapy in advanced non-small-cell lung cancer: a systematic review. *PLoS One* 2011;6:e26646.

13. Hayashi H, Okamoto I, Morita S, et al. Postprogression survival for first-line chemotherapy of patients with advanced non-small-cell lung cancer. *Ann Oncol* 2012;23:1537-1541.

14. A'Hern RP. Restricted Mean Survival Time: An Obligatory End Point for Time-to-Event Analysis in Cancer Trials? *J Clin Oncol* 2016;34:3474-3476.

15. Saad ED, Katz A. Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. *Ann Oncol* 2009;20:460-464.

16. CurveSnap. Available at https://curvesnap.en.softonic.com/.

17. Trinquart L, Jacot J, Conner SC, et al. Comparison of Treatment Effects Measured by the Hazard Ratio and by the Ratio of Restricted Mean Survival Times in Oncology Randomized Controlled Trials. *J Clin Oncol* 2016;34:1813-1819.

18. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC medical research methodology* 2013;13:152.

19. Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol* 2014;32:2380-2385.

421 20. Saad ED, Buyse M. Statistical controversies in clinical research: end points other than
422 overall survival are vital for regulatory approval of anticancer agents. *Ann Oncol*
423 2016;27:373-378.
424 21. Morgensztern D, Waqar S, Subramanian J, et al. Improving survival for stage IV non-
425 small cell lung cancer: a surveillance, epidemiology, and end results survey from 1990 to
426 2005. *J Thorac Oncol* 2009;4:1524-1529.
427 22. Breathnach OS, Freidlin B, Conley B, et al. Twenty-two years of phase III trials for
428 patients with advanced non-small-cell lung cancer: sobering results. *J Clin Oncol*
429 2001;19:1734-1742.
430 23. Fukuoka M, Wu YL, Thongprasert S, et al. Biomarker analyses and final overall
431 survival results from a phase III, randomized, open-label, first-line study of gefitinib versus
432 carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer
433 in Asia (IPASS). *J Clin Oncol* 2011;29:2866-2874.
434 24. Alfonso S, Valdes-Zayas A, Santiesteban ER, et al. A randomized, multicenter,
435 placebo-controlled clinical trial of racotumomab-alum vaccine as switch maintenance therapy
436 in advanced non-small cell lung cancer patients. *Clin Cancer Res* 2014;20:3660-3671.
437 25. Giaccone G, Bazhenova LA, Nemunaitis J, et al. A phase III study of
438 belagenpumatucel-L, an allogeneic tumour cell vaccine, as maintenance therapy for non-small
439 cell lung cancer. *Eur J Cancer* 2015;51:2321-2329.
440 26. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously
441 treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a
442 randomised controlled trial. *Lancet* 2016;387:1540-1550.
443

Table 1. Characteristics of trials with and without available RMSTs truncated at 18 months.

Characteristic	Trials with available RMSTs (N=82)	Trials without available RMSTs (N=79)
Year or publication		
1998 to 2003	19 (23%)	28 (35%)
2004 to 2009	27 (33%)	32 (41%)
2010 to 2015	36 (44%)	19 (24%)
Patients per arm, median (IQR)	183 (138; 304)	154 (99; 222)
Primary end point*		
Overall survival	55 (69%)	48 (61%)
Progression-free survival	14 (18%)	14 (18%)
Other	11 (14%)	16 (20%)
Reported gain in overall survival**	16 (20%)	21 (27%)
Reported gain in progression-free survival**	33 (40%)	20 (32%)

Percentages are rounded to integers. IQR, interquartile range; RMST, restricted mean survival time.

*Trials with co-primary endpoints not shown.**Denominator is the number of trials with information on gains, which we deemed to be present if there was a statistically significant difference in any pairwise comparison (or factor comparison in factorial trials) for PFS or OS following the trialists' definitions, regardless of which of these was the primary end point and of the number of arms in a given trial, and regardless of whether the difference favored the experimental or the control arm.

Table 2. Availability of data on mean PFS and OS for each 2-year period.

Year of publication	Number of trials	Number of trials with data for analysis according to truncation time (*t*)		
		For *t*=5 months	For *t*=12 months	For *t*=18 months
1998-99	12	6 (50%)	5 (42%)	4 (33%)
2000-01	15	6 (40%)	4 (27%)	3 (20%)
2002-03	20	13 (65%)	13 (65%)	12 (60%)
2004-5	22	16 (73%)	16 (73%)	12 (55%)
2006-07	17	9 (53%)	8 (47%)	6 (35%)
2008-09	20	10 (50%)	10 (50%)	9 (45%)
2010-11	19	14 (74%)	14 (74%)	13 (68%)
2012-13	21	17 (81%)	16 (76%)	15 (71%)
2014-15	15	11 (73%)	11 (73%)	8 (53%)
Total	161	102 (63%)	97 (60%)	82 (51%)

Table 3. Average RMSTs (in months) for each 6-year period of interest, control arms.

Years of publication	Average according to end point and truncation time (*t*)					
	For *t*=5 months (N=102)		For *t*=12 months (N=97)		For *t*=18 months (N=82)	
	PFS	OS	PFS	OS	PFS	OS
1998 to 2003	3.4	4.2	5.1	7.7	5.6	8.9
2004 to 2009	3.6	4.4	5.1	8.3	5.6	10.0
2010 to 2015	3.7	4.5	5.4	9.0	6.0	11.4

N indicates number of trials with available data.

Table 4. Average difference in RMSTs (in months) between experimental and control arms within trials.

Years of publication	Average difference according to end point and truncation time (*t*)					
	For *t*=5 months (N=102)		For *t*=12 months (N=97)		For *t*=18 months (N=82)	
	PFS	OS	PFS	OS	PFS	OS
1998 to 2003	0.3	0.1	0.3	0	0.3	0.1
2004 to 2009	0.1	0	0.5	0.1	0.7	0.2
2010 to 2015	0.2	0	0.7	0.1	0.9	0.1
Overall in 18 years	0.2	0	0.5	0.1	0.7	0.2

N indicates number of trials with available data.

FIGURE LEGENDS

Figure 1. Yearly mean number of patients per arm over time.

Figure 2. Frequency of use of overall survival (dark blue), progression-free survival (light blue), response or clinical benefit rates (dark green), quality of life or toxicity (light green), and other primary end points (red) in the three 6-year periods spanned by the study.

Figure 3. Time trends in restricted mean and in median progression-free and overall survival times. Panel A, yearly average of the restricted mean survival times (RMSTs) for progression-free (PFS) and overall survival (OS) in control arms, truncated at 5 months; Panel B, yearly average of the RMSTs for PFS and OS in control arms, truncated at 12 months; Panel C, yearly average of the RMSTs for PFS and OS in control arms, truncated at 18 months; Panel D, yearly average of the median PFS and OS in control arms.