A Proposal for Progression-Free Survival Assessment in Patients with Early Progressive Cancer

TAKANORI TANASE[1,2], CHIKUMA HAMADA[2], TAKAYUKI YOSHINO[3] and ATSUSHI OHTSU[4]

[1]*Department of Data Science, Taiho Pharmaceutical Co., Ltd, Tokyo, Japan;*
[2]*Department of Management Science, Graduate School of Engineering,*
Tokyo University of Science, Tokyo, Japan;
[3]*Department of Gastroenterology and Gastrointestinal Oncology,*
National Cancer Center Hospital East, Kashiwa, Japan;
[4]*National Cancer Center Hospital East, Kashiwa, Japan*

Abstract. *Background/Aim: Progression-free survival (PFS), which is evaluated in oncology clinical trials, is determined based on tumor progression evaluated according to an assessment schedule. There is possibly a bias in median PFS and hazard ratio (HR) for PFS depending on the assessment schedule referring to randomized controlled trials (RCTs) in patients with metastatic colorectal cancer. Materials and Methods: We re-analyzed the PFS in the FTD/TPI phase 2 trial by changing the assessment schedule. To assess biases in median PFS and HR for PFS resulting from different assessment schedules, we performed a computational simulation. Results: The reanalysis of FTD/TPI phase 2 trial and the simulation results showed that there were biases in median PFS and HR for PFS. Conclusion: In RCTs for early progressive cancer, median PFS is dependent on the assessment schedule; however, HR for PFS can be assessed without clinically-meaningful differences between assessment schedules, regardless of biomarker assumptions.*

A randomized controlled trial (RCT) has been established to assess the therapeutic effect of drugs. Overall survival, which is defined as the time from randomization to death by any cause, is used as a true endpoint in oncology clinical trials. In nearly a decade, the use of progression-free survival (PFS), which is defined as the time from randomization to tumor progression or death by any cause, as a surrogate end-point has become of increasing importance. RCTs to be conducted for new drug application are designed to meet the requirements of the regulatory authorities. The Food and Drug Administration (1) and the European Medicines Agency (2) have released documents on the definition and the basic nature of PFS. The antitumor effect of drugs has been assessed according to the Response Evaluation Criteria in Solid Tumors (RECIST) guideline (version 1.1) (3).

Oncology clinical trials aimed at the development of new drugs are often conducted for later-line treatment of patients with early progressive cancer. We reviewed the RCTs in patients with metastatic colorectal cancer who were refractory to standard chemotherapy and questioned the relationship between median PFS and tumor assessment schedule (4-8). The shapes of the PFS curves were very similar in the four trials (4-7) in which the initial tumor assessment was performed at 8 weeks after study treatment, with an overlap ranging from 50-100% in PFS between the treatment groups and with similar median PFS between the treatment groups. In contrast, in the trial reported by Yoshino *et al.* (8) in which the initial tumor assessment was performed at 4 weeks, all the PFS curves were different between the treatment groups from 4 weeks. In general, partial overlap or intersection in the Kaplan–Meier curve suggests the presence of a biomarker. The *RAS* (*KRAS* or *NRAS*) mutation has been established as a biomarker for assessing the response to cetuximab and panitumumab treatment for patients with metastatic colorectal cancer (9-11); however, currently, there is no established biomarker for assessing the response to regorafenib and trifluridine and tipiracil hydrochloride (FTD/TPI) treatment. We assumed that no difference in median PFS exists despite reduced risks for tumor progression or death because of the presence of biomarkers as well as the difference in tumor assessment schedules.

Correspondence to: Takanori Tanase, Department of Data Science, Taiho Pharmaceutical Co., Ltd, 1-2-4, Uchikanda, Chiyoda, Tokyo 101-0047, Japan. Tel: +81 332934397, Fax: +81 332938588, e-mail: t-tanase@taiho.co.jp

Table I. *Simulation results.*

Scenario	True HR for PFS	True median PFS (weeks)*	Simulation result[†]		
			Median PFS (weeks)*	HR for PFS (Bias[‡])	Power
1	0.5	6 *vs.* 3	9.08 *vs.* 8.08	0.566 (13%)	68.04%
			9.07 *vs.* 6.15	0.543 (9%)	73.11%
			8.22 *vs.* 4.67	0.530 (6%)	76.91%
2	0.5	8 *vs.* 4	12.06 *vs.* 8.15	0.543 (9%)	73.11%
			11.43 *vs.* 6.47	0.529 (6%)	77.06%
			10.29 *vs.* 6.07	0.524 (5%)	78.20%
3	0.5	12 *vs.* 6	16.42 *vs.* 9.11	0.526 (5%)	77.81%
			15.21 *vs.* 9.07	0.520 (4%)	79.14%
			15.80 *vs.* 8.23	0.519 (4%)	79.14%
4	0.67	6 *vs.* 4	8.47 *vs.* 8.15	0.701 (5%)	71.17%
			9.08 *vs.* 6.26	0.689 (3%)	75.03%
			8.11 *vs.* 6.07	0.683 (2%)	76.68%
5	0.67	8 *vs.* 5.33	12.08 *vs.* 8.27	0.689 (3%)	75.03%
			11.78 *vs.* 7.50	0.681 (2%)	77.20%
			10.07 *vs.* 7.82	0.679 (2%)	78.16%
6	0.67	12 *vs.* 8	16.17 *vs.* 12.08	0.679 (2%)	77.85%
			15.07 *vs.* 11.77	0.676 (1%)	78.90%
			15.96 *vs.* 10.06	0.676 (1%)	78.63%
7	0.75	6 *vs.* 4.5	8.34 *vs.* 8.18	0.774 (3%)	71.64%
			9.05 *vs.* 6.34	0.765 (2%)	75.21%
			8.06 *vs.* 7.27	0.762 (2%)	76.64%
80	0.75	8 *vs.* 6	12.05 *vs.* 8.35	0.765 (2%)	75.21%
			11.89 *vs.* 9.08	0.760 (1%)	76.99%
			10.04 *vs.* 8.06	0.759 (1%)	77.65%
9	0.75	12 *vs.* 9	16.06 *vs.* 14.73	0.759 (1%)	77.49%
			15.03 *vs.* 12.07	0.757 (1%)	78.66%
			15.95 *vs.* 11.40	0.757 (1%)	78.35%
10	BM+: 0.5	BM+: 6 *vs.* 3	8.17 *vs.* 8.08	0.725 (6%)	69.14%
	BM-: 1	BM-: 3 *vs.* 3	6.30 *vs.* 6.15	0.710 (4%)	75.34%
	Overall:0.685	Overall: 4.17 *vs.* 3	6.48 *vs.* 4.37	0.702 (3%)	77.69%
11	BM+: 0.5	BM+: 8 *vs.* 4,	8.31 *vs.* 8.15	0.710 (4%)	75.34%
	BM-: 1	BM-: 4 *vs.* 4,	7.96 *vs.* 6.26	0.701 (2%)	78.18%
	Overall:0.685	Overall: 5.55 *vs.* 4	7.93 *vs.* 6.07	0.698 (2%)	79.03%
12	BM+: 0.5	BM+: 12 *vs.* 6,	12.97 *vs.* 8.43	0.699 (2%)	79.04%
	BM-: 1	BM-: 6 *vs.* 6,	11.92 *vs.* 9.07	0.694 (1%)	80.42%
	Overall:0.685	Overall: 8.33 *vs.* 6	10.52 *vs.* 8.09	0.694 (1%)	80.37%
13	BM+: 0.67	BM+: 6 *vs.* 4	8.21 *vs.* 8.14	0.825 (3%)	72.00%
	BM-: 1	BM-: 4 *vs.* 4	6.40 *vs.* 6.26	0.817 (2%)	75.73%
	Overall:0.804	Overall: 4.87 *vs.* 4	7.76 *vs.* 6.06	0.814 (1%)	76.92%
14	BM+: 0.67	BM+: 8 *vs.* 5.33	8.41 *vs.* 8.26	0.817 (2%)	75.73%
	BM-: 1	BM-: 5.33 *vs.* 5.33	10.82 *vs.* 6.80	0.813 (1%)	77.32%
	Overall:0.804	Overall: 6.49 *vs.* 5.33	8.17 *vs.* 7.90	0.812 (1%)	78.03%
15	BM+: 0.67	BM+: 12 *vs.* 8	15.73 *vs.* 12.07	0.811 (1%)	78.23%
	BM-: 1	BM-: 8 *vs.* 8	12.18 *vs.* 11.90	0.809 (1%)	79.27%
	Overall:0.804	Overall: 9.74 *vs.* 8	11.97 *vs.* 10.04	0.810 (1%)	79.12%
16	BM+: 0.5	BM+: 6 *vs.* 3	8.23 *vs.* 8.08	0.664 (9%)	67.25%
	BM-: 0.75	BM-: 4 *vs.* 3	6.95 *vs.* 6.15	0.643 (6%)	73.79%
	Overall:0.607	Overall: 4.87 *vs.* 3	7.36 *vs.* 4.47	0.631 (4%)	77.39%
17	BM+: 0.5	BM+: 8 *vs.* 4	9.18 *vs.* 8.15	0.643 (6%)	73.79%
	BM-: 0.75	BM-: 5.33 *vs.* 4	9.99 *vs.* 6.30	0.631 (4%)	77.27%
	Overall:0.607	Overall: 6.49 *vs.* 4	8.50 *vs.* 6.05	0.626 (3%)	78.61%
18	BM+: 0.5	BM+: 12 *vs.* 6	14.80 *vs.* 8.65	0.627 (3%)	78.46%
	BM-: 0.75	BM-: 8 *vs.* 6	12.71 *vs.* 9.04	0.621 (2%)	79.80%
	Overall:0.607	Overall: 9.74 *vs.* 6	12.32 *vs.* 8.15	0.621 (2%)	7976%
19	BM+: 0.67	BM+: 6 *vs.* 4	8.30 *vs.* 8.15	0.738 (4%)	70.45%
	BM-: 0.75	BM-: 5.33 *vs.* 4	8.10 *vs.* 6.26	0.727 (3%)	74.55%
	Overall:0.706	Overall: 5.65 *vs.* 4	7.98 *vs.* 6.08	0.721 (2%)	76.68%

Table I. *Continued*

Table I. *Continued*

Scenario	True HR for PFS	True median PFS (weeks)*	Simulation result[†]		
			Median PFS (weeks)*	HR for PFS (Bias[‡])	Power
20	BM+: 0.67	BM+: 8 *vs.* 5.33	10.73 *vs.* 8.26	0.727 (3%)	74.55%
	BM-: 0.75	BM-: 7.11 *vs.* 5.33	11.65 *vs.* 7.32	0.720 (2%)	77.26%
	Overall:0.706	Overall: 7.54 *vs.* 5.33	9.38 *vs.* 7.86	0.718 (2%)	77.92%
21	BM+: 0.67	BM+: 12 *vs.* 8	15.97 *vs.* 12.11	0.718 (2%)	78.03%
	BM-: 0.75	BM-: 10.67 *vs.* 8	14.01 *vs.* 11.83	0.715 (1%)	78.75%
	Overall: 0.706	Overall: 11.31 *vs.* 8	14.53 *vs.* 10.07	0.715 (1%)	79.19%

HR: Hazard ratio; PFS: progression-free survival; BM+: biomarker-positive population; BM-: biomarker-negative population; *Experimental group *vs.* control group; [†]Tumor assessment schedule; upper: every 8 weeks, middle: every 6 weeks, and lower: week 4, 8, 12, and every 8 weeks thereafter. [‡]100 × (estimated HR for PFS – true HR for PFS)/true HR for PFS. The median PFS, the HR for PFS, and the power were calculated for overall population in the scenarios of biomarker-positive/-negative assumption.

In actual RCTs, tumor assessment is generally scheduled every 6 or 8 weeks according to the RECIST guideline. Regarding the bias resulting from different assessment intervals, Freidlin *et al.* (12) described that the potential event–time bias is caused by a higher number of unscheduled tumor assessments in the control group than that in the experimental group for unblended RCTs. Panageas *et al.* (13) reported the results of simulation of median PFS and bias from the Kaplan–Meier method with various tumor assessment schedules, and it is known that median PFS is not measured accurately depending on the tumor assessment schedule. Carroll (14) proposed the formula to calculate a HR for PFS under the assumption of exponential distribution for event time and tumor assessment at regular intervals. However, whether a median PFS and a hazard ratio (HR) for PFS depends on the tumor assessment schedule remains to be evaluated in RCTs for patients with early progressive cancer. In this study, we performed a simulation using several tumor assessment schedules and assessed the effect on median PFS and HR for PFS to investigate the effect of the tumor assessment schedules on PFS by assuming a biomarker, which has an interaction effect on the study treatment.

Materials and Methods

We re-analyzed the results of PFS in the randomized phase 2 trial of FTD/TPI (8) as an experimental treatment (clinical trial registration ID: JapicCTI-090880). Although this trial had been originally scheduled to perform tumor assessment at 4, 8, 12, and every 8 weeks thereafter, we prepared a dataset assuming that the assessments were performed every 8 weeks. The median PFS was estimated using the Kaplan–Meier method, and the HR for PFS was estimated using the Cox proportional hazard model. Statistical analysis was performed using SAS 9.4.

Statistical simulations for PFS were performed using three tumor assessment schedules based on the results of RCTs in patients with metastatic colorectal cancer refractory to standard chemotherapy (4-8). The RCTs in simulation were designed as two-group parallel controlled trials. It was assumed that the time of tumor progression followed the exponential distribution and the Weibull distribution. Tumor assessment was scheduled according to the following three types of interval: every 8 weeks (Sc8), every 6 weeks (Sc6), or at 4, 8, 12, and every 8 weeks thereafter (Sc4). Tumor progression was completed during the period of 7 days before and after the standard visit date with normal distribution. The same scenarios were included in the assumption of the interaction between study treatment and biomarker-positive/-negative (biomarker-positive patients have a good response to experimental treatment). The number of patients was set to meet the power of 80% and the two-sided significance level of 5% for each scenario. The median PFS, HR for PFS, and the power were estimated by the Monte Carlo simulation using R 3.2.3. The number of iterations was 10,000 for each scenario.

Results

Based on the analysis result of PFS in the modified tumor assessment schedule in FTD/TPI phase 2 trial, the median PFS was 2.1 months (95% confidence interval=2.0-3.7) in the FTD/TPI group and 1.9 months (95%CI=1.9-1.9) in the placebo group, and the HR for PFS was 0.48 (95%CI=0.33-0.70; logrank $p<0.0001$). The PFS curve is shown in Figure 1.

The simulation results assuming the exponential distribution are shown in Table I. Median PFS in the control group was dependent on the initial time of tumor assessment in the scenarios 1, 10, and 16, and these scenarios reproduced the trial results for patients with metastatic colorectal cancer (4-8). The biases in HR for PFS emerged, and the power decreased to less than 75% in scenarios 1, 2, 4, 7, 10, 13, 16, 17, 19, and 20 for Sc8 or Sc6. The decrease of power to less than 75% was not observed in the scenarios for Sc4. The HR for PFS was not affected in any tumor assessment schedule when the true median PFS in the
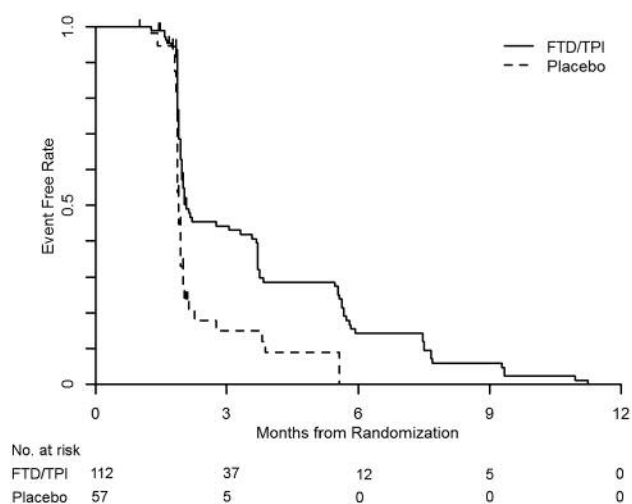
Figure 1. *PFS curve reanalyzed after tumor assessment schedule was changed to every 8 weeks in the FTD/TPI phase 2 trial. FTD/TPI: Trifluridine and tipiracil hydrochloride.*

experimental group was 12 weeks. The simulation results assuming the Weibull distribution were not shown because these results and discussion were almost similar to those assuming the exponential distribution.

Discussion

We re-analyzed the PFS in the modified tumor assessment schedule to every 8 weeks using FTD/TPI phase 2 trial data. There were few differences in the median PFS between the treatment groups (2.1 *vs*. 1.9 months from the reanalysis; 2.0 *vs*. 1.0 months from the original), which were consistent with the results of other studies (4-7). Regardless of biomarker-positive/-negative patients, the median PFS would be easily affected by the initial time of tumor assessment. The HR for PFS was estimated conservatively than the original result (reanalysis HR=0.48; original HR=0.41); however, the differences would not be clinically meaningful. These results suggest that median PFS and HR for PFS would be affected statistically by the tumor assessment schedule in practical study for patients with early progressive cancer. Median PFS should not be considered as "the median of the time of disease progression or death" but as "the time when 50% or more patients have progressed disease or died" if median PFS is close to the initial time of tumor assessment. It might be, therefore, useful for summarizing PFS to calculate restricted mean survival time or the PFS rate (the number of patients with PFS divided by the number of all patients) at the initial time of tumor assessment as an alternative to median PFS.

The simulation results reproduced the reanalysis results of PFS in the FTD/TPI phase 2 trial. The median PFS was affected by the initial time of tumor assessment when the true median PFS was less than or equal to the initial time of tumor assessment. Statistical biases affected to the power (<75%: more than 5% reduction from 80%) were also observed in HR for PFS when the true median PFS in a group was less than or equal to the initial time of tumor assessment and the true median PFS in a group was less than 12 weeks; however, the differences in HR for PFS were not clinically meaningful between the tumor assessment schedules (for example, 0.566 for Sc8, 0.543 for Sc6, and 0.530 for Sc4 in scenario 1). From the aspect of statistical power, these biases in HR for PFS would be few in RCTs assessing the tumor progression according to schedule of every 8 or 6 weeks when the true median PFS in the experimental group is greater than or equal to 12 weeks. The simulation results indicated that the PFS curve depends on the tumor assessment schedule if there is no potential biomarker that has an interaction effect on the study treatment. We cannot speculate whether there are potential biomarkers for experimental treatment from the PFS curve. Potential biomarkers should be considered based on clinical hypotheses in reference to subgroup analysis, *p*-value for interaction, or other exploratory analyses (15).

We conducted a systematic review of articles published on PubMed between Jan 1, 2007 and Dec 31, 2016. The reports selected for review were limited to the RCTs using either placebo alone or best supportive care alone as the control group. We discovered possible biases in the median PFS and the HR for PFS for 38 RCTs, due to a median PFS that was less than or equal to the initial time of tumor assessment for following cancers: adrenocortical cancer, colorectal cancer, gall bladder cancer, gastric cancer, gastrointestinal stromal tumor, hepatocellular cancer, head and neck cancer, malignant pleural mesothelioma, non–small–cell lung cancer, oesophageal cancer, pancreatic cancer, pancreatic neuroendocrine tumor, prostate cancer, renal cell cancer, soft tissue sarcoma, and urothelial cancer.

From another point of view, the initial tumor assessment at 4 weeks would affect the best overall response (BOR). For example, it was assumed that the criterion of stable disease (SD) as the BOR was defined to be at least 6 weeks. Patients who are evaluated as having non-progressive disease (non-PD) at 4 weeks and whose second assessment is unavailable for any reason are determined to be not evaluable (NE) as the BOR because they do not meet the criterion of SD. This is also caused by the difference between independent review and investigator review. If the initial assessment is performed at 4 weeks and the patient is determined to be PD by an investigator at 4 weeks, followed by treatment discontinuation, the subsequent imaging results cannot be evaluated. If the patient is considered to have a non-PD in

an independent review, the patient is determined to be NE because the patient does not meet the criterion of SD. This would result in an increase in the number of patients who are NE in independent reviews. The FTD/TPI phase 2 trial showed that the proportion of patients determined to be NE in the investigator review was 0.9% (1/112) in the FTD/TPI group and no patients in the placebo group; in the independent review, it was 8.9% (10/112) in the FTD/TPI group and 12.3% (7/57) in the placebo group. The disease control rate (DCR), that was defined as the proportion of patients with BOR of complete response, partial response, or SD, in the independent review was less than that in the investigator review because more patients were determined to be NE in the independent review. DCR in the independent review was 43.8% (49/112) in the FTD /TPI group and 10.5% (6/57) in the placebo group, and DCR in the investigator review was 54.5% (61/112) in the FTD /TPI group and 14.0% (8/57) in the placebo group.

In conclusion, we demonstrated the characteristic of PFS for patients with early progressive cancer. The median PFS would be dependent on the tumor assessment schedule, and the statistical power would be decreased when the median PFS is less than or equal to the initial time of tumor assessment. The effects on the HR for PFS would not be clinically meaningful; however, a warning is necessary in the decrease of statistical power.

Conflicts of Interest

Acknowledgements

References

1 Food and Drug Administration: Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics, 2007.

2 European Medicines Agency: Guideline on the Evaluation of Anticancer Medicinal Products in Man, 2013.

3 Eisenhauer EA, Therasse P, Bogaerts J Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D and Verweij J: New Response Evaluation Criteria in Solid Tumors: Revised RECIST Guideline (version 1.1). Eur J Cancer *45*: 228-247, 2009.

4 Jonker DJ, O'Callaghan CJ, Karapetis CS, Zalcberg JR, Tu D, Au H-J, Berry SR, Krahn M, Price T, Simes RJ, Tebbutt NC, Van Hazel G, Wierzbicki R, Langer C and Moore MJ: Cetuximab for the treatment of colorectal cancer. N Engl J Med *357*: 2040-2048, 2007.

5 Van Cutsem E, Peeters M, Siena S, Humblet Y, Hendlisz A, Neyns B, Canon J-L, Van Laethem J-L, Maurel J, Richardson G, Wolf M and Amado RG: Open-label, randomized, phase 3 clinical trial of panitumumab plus best supportive care versus best supportive care in patients with chemotherapy-refractory metastatic colorectal cancer. J Clin Oncol *25*: 1658-1664, 2007.

6 Grothey A, Van Cutsem E, Sobrero A, Siena S, Falcone A, Ychou M, Humblet Y, Bouché O, Mineur L, Barone C, Adenis A, Tabernero J, Yoshino T, Lenz HJ, Goldberg RM, Sargent DJ, Cihon F, Cupit L, Wagner A and Laurent D: Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. Lancet *381*: 303-312, 2013.

7 Mayer RJ, Van Cutsem E, Falcone A, Yoshino T, Garcia-Carbonero R, Mizunuma N, Yamazaki K, Shimada Y, Tabernero J, Komatsu Y, Sobrero A, Boucher E, Peeters M, Tran B, Lenz H-J, Zaniboni A, Hochster H, Cleary JM, Prenen H, Benedetti F, Mizuguchi H, Makris L, Ito M and Ohtsu A: Randomized Trial of TAS-102 for Refractory Metastatic Colorectal Cancer. N Engl J Med *372*: 1909-1919, 2015.

8 Yoshino T, Mizunuma N, Yamazaki K, Nishina T, Komatsu Y, Baba H, Tsuji A, Yamaguchi K, Muro K, Sugimoto N, Tsuji Y, Moriwaki T, Esaki T, Hamada C, Tanase T and Ohtsu A: TAS-102 monotherapy for pretreated metastatic colorectal cancer: a double-blind, randomised, placebo-controlled phase 2 trial. Lancet Oncol *13*: 993-1001, 2012.

9 National Comprehensive Cancer Network: Clinical Practice Guidelines in Oncology, Colon Cancer Version 2.2017.

10 National Comprehensive Cancer Network: Clinical Practice Guidelines in Oncology, Rectal Cancer Version 2.2017.

11 Taniguchi H, Yamazaki K, Yoshino T, Muro K, Yatabe Y, Watanabe T, Ebi H, Ochiai A, Baba E and Tsuchihara K: Japanese Society of Medical Oncology: Japanese Society of Medical Oncology Clinical Guidelines: RAS (KRAS/NRAS) mutation testing in colorectal cancer patients. Cancer Sci *106*: 324-327, 2015.

12 Freidlin B, Korn EL, Hunsberger S, Gray R, Saxman S and Zujewski JA: Proposal for the Use of Progression-Free Survival in Unblinded Randomized Trials. J Clin Oncol *25*: 2122-2126, 2007.

13 Panageas KS, Ben-Porat L, Dickler MN, Chapman PB and Schrag D: When You Look Matters: The Effect of Assessment Schedule on Progression-Free Survival. J Natl Cancer Inst *99*: 428-432, 2007.

14 Carroll KJ: Analysis of progression-free survival in oncology trials: Some common statistical issues. Pharm Stat *6*: 99-113, 2007.

15 Lipkovich I, Dmirienko A and D'Agonstino RB: Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. Stat Med *36*: 136-196, 2017.