


Exploration of modified progression-free survival as a novel surrogate endpoint for overall survival in immuno-oncology trials

Zi-Xian Wang,^{1,2} Hao-Xiang Wu,^{2,3} Li Xie,⁴ Wu-Hao Lin,^{1,2} Fei Liang,⁵ Jin Li,⁶ Zhi-Min Yang,⁷ Rui-Hua Xu ^{1,2}

To cite: Wang Z-X, Wu H-X, Xie L, *et al.* Exploration of modified progression-free survival as a novel surrogate endpoint for overall survival in immuno-oncology trials. *Journal for ImmunoTherapy of Cancer* 2021;**9**:e002114. doi:10.1136/jitc-2020-002114

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jitc-2020-002114>).

Z-XW, H-XW, LX and W-HL are joint first authors.

JL, Z-MY and R-HX are joint senior authors.

Accepted 21 February 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Rui-Hua Xu;
xurh@systucc.org.cn

ABSTRACT

Background Progression-free survival (PFS) exhibits suboptimal performance as the surrogate endpoint for overall survival (OS) in trials studying immune checkpoint inhibitors (ICIs). Here we propose a novel surrogate endpoint, modified PFS (mPFS), which omits the events of disease progression (but not deaths) within 3 months after randomization.

Methods PubMed, EMBASE, and the Cochrane Central Register of Controlled Trials were searched for randomized trials studying ICIs in advanced solid tumors with available PFS and OS data up to May 2020. Individual patient-level data (IPD) for PFS and OS were reconstructed for eligible trials. A simulation-based algorithm was used to match the reconstructed, disconnected PFS and OS IPD, and 1000 independent simulated datasets of matched PFS-OS IPD were generated for each trial. mPFS durations and statuses were then measured for each of the matched PFS-OS IPD. Trial-level correlation between Cox HRs for PFS or mPFS and HRs for OS was assessed using Pearson correlation coefficient (r_p) weighted by trial size; patient-level correlation between PFS or mPFS and OS was assessed using Spearman's rank correlation coefficient (r_s). Findings were further validated using the original IPD from two randomized ICI trials.

Results Fifty-seven ICI trials totaling 29,429 participants were included. PFS HR showed moderate correlation with OS HR ($r_p=0.60$), and PFS was moderately correlated with OS at the patient level (median $r_s=0.66$; range, 0.65–0.68 among the 1000 simulations). In contrast, mPFS HR achieved stronger correlation with OS HR (median $r_p=0.81$; range, 0.77–0.84), and mPFS was more strongly correlated with OS at the patient level (median $r_s=0.79$; range, 0.78–0.80). The superiority of mPFS over PFS remained consistent in subgroup analyses by cancer type, therapeutic regimen, and treatment setting. In both trials with the original IPD where experimental treatment significantly improved OS, mPFS successfully captured such clinical benefits whereas PFS did not.

Conclusions mPFS outperformed PFS as the surrogate endpoint for OS in ICI trials. mPFS is worthy of further investigation as a secondary endpoint in future ICI trials.

INTRODUCTION

Over the past decades, there have been significant advances in immunotherapy to treat

various cancer types, among which immune checkpoint inhibitors (ICIs) are the most promising.^{1,2} ICIs employ unique mechanisms to activate or rehabilitate self-immunity toward tumors, which may result in delayed clinical effects and long-term survival benefits.^{3,4} If these characteristics of ICIs are not accounted for in the trial design stage, loss of statistical power and substantially prolonged follow-up may occur.^{5–7} This represents a marked challenge in the development and approval of effective ICIs, especially for trials in which overall survival (OS) is the primary endpoint.

For clinical trials of advanced solid tumors, RECIST-based progression-free survival (PFS) and objective response rate (ORR) are widely applied as surrogate endpoints to detect early signals of drug activity. However, meta-analyses of trials involving ICIs showed that the treatment effects, as assessed using PFS and ORR, correlated moderately to poorly with OS,^{8,9} questioning the continued use of PFS and ORR as surrogate endpoints in ICI trials. Some investigators suggest using immune-related tumor response evaluation criteria to refine the definition of tumor response and progression.^{10–12} However, as these criteria lack consensus and are still not widely used in ICI trials, there is a lack of sufficient data to evaluate the reliability of immune-related ORR and PFS as surrogate endpoints for ICIs.

The suboptimal correlation between PFS and OS in ICI trials may be attributed to disease progression (PD) followed by either tumor shrinkage (pseudoprogression),^{13–15} or a long post-PD survival, both of which suggest delayed effects of ICIs. We define such PD events as “low-quality PD events” here, but they cannot be accurately identified due to the deficiency in robust biomarkers and in the understanding of underlying molecular

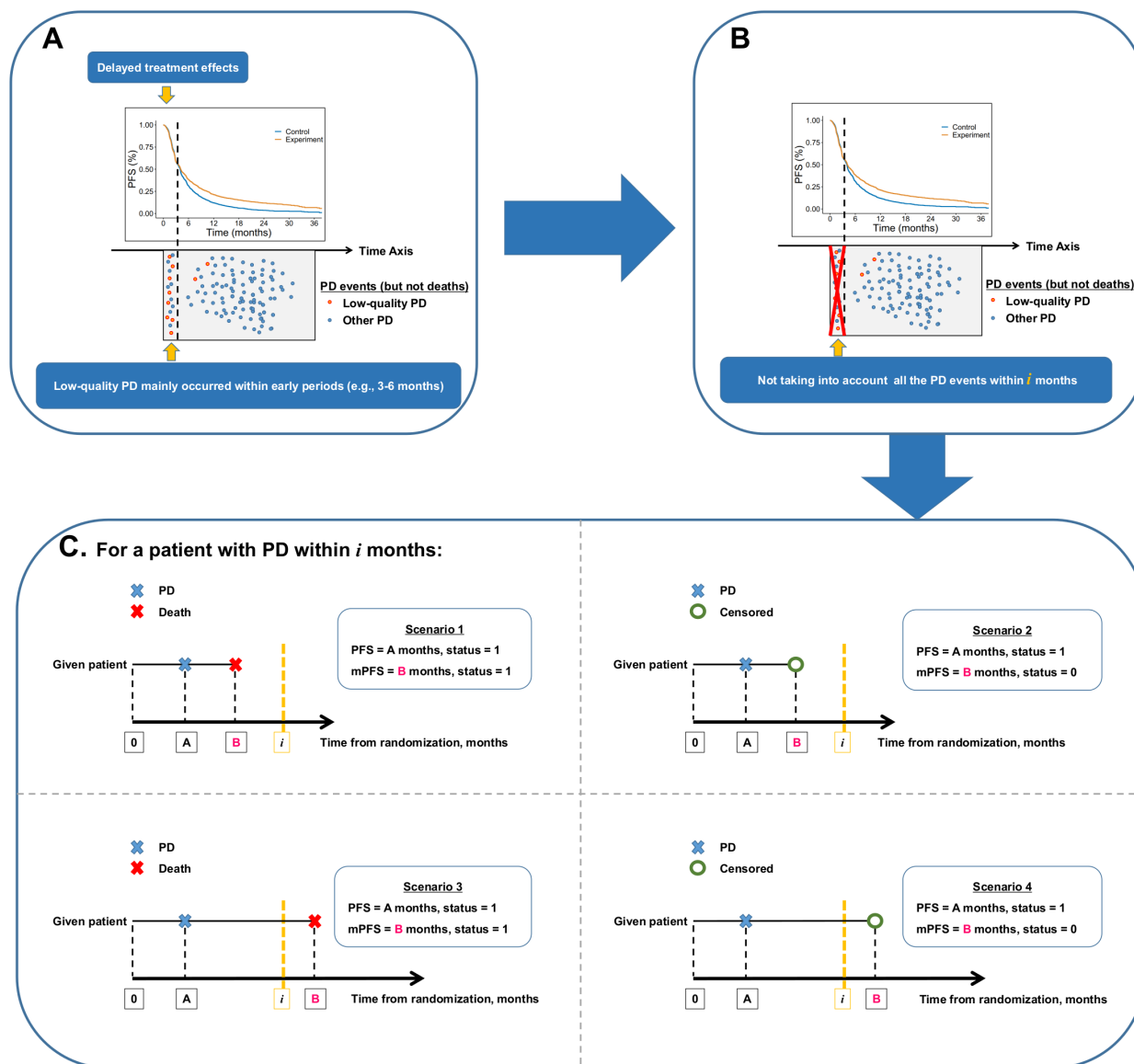


Figure 1 Rationale for and measurement of modified progression-free survival (mPFS). (A) With the presence of delayed effects of checkpoint inhibitors, low-quality disease progression (PD) events are more likely to occur in the early period (eg, first 3–6 months) after randomization. (B) By not taking into account the PD events (but not deaths) before a certain timepoint (month i), we may remove the majority of low-quality PD events. (C) For a patient with PD within i months, this PD event was omitted while the follow-up continued until death; hence, the PFS duration for this patient equaled the time to PD, whereas the mPFS duration equaled the time to death or last follow-up. Exemplar scenario 1: PD and death within i months; scenario 2: PD within i months and censored within i months at the time of last follow-up; scenario 3: PD within i months and death beyond i months; scenario 4: PD within i months and censored beyond i months at the time of last follow-up. A status of 1 indicates occurrence of a PFS or mPFS event.

mechanisms for these clinical phenomena. Nevertheless, in view of the unique patterns of survival curves under delayed effects of ICIs (eg, delayed separation of curves after 3–6 months from randomization),¹⁶ we speculate that “low-quality PD events” are more likely to occur in the early period (eg, first 3–6 months) after randomization (figure 1A). Thus, by omitting these early PD events (but not deaths) before appropriate timepoints, we may remove the majority of “low-quality PD events” (figure 1B) and improve the agreement with OS.

In this study of reconstructed and original individual patient-level data (IPD) from ICI trials, we propose

and validate a novel surrogate endpoint, modified PFS (mPFS), which applies the same definition of traditional PFS except it purposefully omits the PD events within 3 months after randomization.

METHODS

Selection of randomized clinical trials

We searched PubMed, EMBASE, and the Cochrane Central Register of Controlled Trials for published trial reports from January 1, 2000 to May 31, 2020. We combined both MeSH and free-text words to identify

relevant studies. The detailed search strategy is available in online supplemental methods. Randomized trials investigating ICIs for advanced solid cancers with available Kaplan-Meier curves for OS and PFS were all potentially eligible. We excluded single-arm phase I or phase II trials, dose-finding trials, and neoadjuvant or adjuvant setting trials. We also excluded trials reporting PFS data solely based on immune-related criteria for assessing tumor response. For eligible studies, we further reviewed the full text and supplemental materials of relevant publications to look for open access IPD.

Two authors (Z-XW and H-XW) screened the trials independently for eligibility and extracted the data from each included trial. Any discrepancies were resolved by consensus.

Reconstruction of IPD and definition of mPFS

We reconstructed de-identified IPD for PFS and OS based on digitized survival curve data (online supplemental figure S1A).¹⁷ We used DigitizeIt software V.2.2 (<http://www.digitizeit.de/>) to measure the time and survival probability coordinates on the Kaplan-Meier curves. The number of patients at risk and the total number of events were also extracted. The data were then input into an algorithm on the basis of iterative numerical methods to solve the inverted Kaplan-Meier equations.¹⁷ We then applied the Cox proportional hazards model to the reconstructed IPD to evaluate HRs for OS (HR_{OS}) and PFS (HR_{PFS}). As shown in figure S2, HR_{OS} and HR_{PFS} obtained from the reconstructed IPD had excellent agreement with HR_{OS} and HR_{PFS} obtained from the original IPD.

To identify the optimal cut-off timepoint to define mPFS, we modified the measurement of PFS by omitting the PD events (but not deaths) within i ($i=2, 3, 4, 5, 6$) months from randomization. Figure 1C illustrates the difference in measuring traditional PFS and mPFS durations and statuses. For a patient with PD within i months, this PD event was omitted while the follow-up continued until death; hence, the traditional PFS duration for this patient equaled the time to PD, whereas the mPFS duration equaled the time to death or the last follow-up. For a patient with PD after i months, traditional PFS and mPFS durations and statuses are identical.

The reconstructed PFS and OS IPD were de-identified and disconnected with each other; therefore, matched PFS-OS IPD were required to evaluate the mPFS duration and status, as well as the HR for mPFS (HR_{mPFS}). We applied a simulation-based algorithm to match the reconstructed PFS IPD to the OS IPD (online supplemental methods and figure S1C). The algorithm generated datasets of matched PFS-OS IPD that must fulfill the following rules: (1) For a given patient, the PFS duration should be no longer than the OS duration; and (2) the patients with events in the OS IPD dataset should be a subgroup of the patients with events in the PFS IPD dataset. Considering that these requirements are insufficient for capturing the original, matched PFS-OS IPD, for each trial we performed 1000 simulations and generated 1000

qualified datasets of matched PFS-OS IPD. The reason for considering 1000 simulations as adequate is detailed in Statistical analysis section.

Statistical analysis

For multi-arm trials, treatment effects from pairwise comparisons were pooled according to the Cochrane Collaboration's recommendation to form a single effect. The validity of PFS and mPFS as surrogate endpoints for OS was assessed at both the trial level and patient level.¹⁸ For trial-level correlation, we calculated Pearson correlation coefficient (r_p) on the basis of the natural log-transformed HR_{PFS} or HR_{mPFS} and HR_{OS} , weighted by trial sample size. We also calculated the surrogate threshold effect (STE), defined as the minimum treatment effect on PFS or mPFS necessary to predict a non-zero effect on OS.¹⁹ A STE greater than 1.00 indicates a trend of underestimating the treatment effect on OS by PFS or mPFS, and vice versa. At the patient level, the bivariate Copula distribution of PFS or mPFS and OS was estimated using Plackett's Copula model.²⁰ The strength of the association between the PFS or mPFS and OS was then quantified using Spearman's rank correlation coefficient (r_s) estimated by the bivariate Copula distribution.²⁰

For mPFS defined by a given cut-off timepoint, the 1000 independent datasets of matched PFS-OS IPD gave rise to one HR_{OS} and 1000 independent HR_{mPFS} for each trial. We randomly divided these datasets into 1000 groups, each group including one dataset for each trial. Within each group, we then calculated the weighted r_p for trial-level correlation between HR_{mPFS} and HR_{OS} , as well as r_s for patient-level correlation between mPFS and OS, thus resulting in 1000 independent r_p , STE, and r_s . For PFS, only one r_p and STE was obtained for trial-level correlation between HR_{PFS} and HR_{OS} , whereas 1000 independent r_s were obtained for patient-level correlation between PFS and OS. As shown in online supplemental figure S3, the distribution of r_p and r_s were stabilized after cumulating 1000 matched PFS-OS datasets for each trial, suggesting that 1000 simulations were adequate to inform the comparative surrogacy of mPFS versus PFS.

For the original IPD from ICI trials, we applied the Cox models to measure HR_{OS} , HR_{PFS} , and HR_{mPFS} . We also evaluated patient-level correlation between PFS or mPFS and OS using the Copula-based r_s .

A two-sided p value <0.05 or one-sided p value <0.025 was considered statistically significant. All statistical analyses were performed using R software V.3.6.0 (<http://www.r-project.org>).

RESULTS

Characteristics of the eligible trials

A total of 57 trials met the selection criteria and were included in this study (online supplemental figure S4). The trial characteristics are summarized in online supplemental table S1. A total of 29,429 patients were enrolled in these trials. Ten different tumor types were examined,

predominantly non-small cell lung cancer (NSCLC, 21 trials) and melanoma (10 trials). Twenty-five trials were in the first-line setting. Forty-eight trials studied anti-PD-1/PD-L1 containing regimens and 14 studied anti-CTLA4 containing regimens (anti-PD-1/PD-L1 plus anti-CTLA4, 5 trials). Twenty-seven trials studied ICI monotherapy and 12 investigated an ICI plus chemotherapy.

Performance of PFS as a surrogate endpoint in ICI trials

As shown in **figure 2A**, HR_{PFS} exhibited moderate trial-level correlation with HR_{OS} ($r_p=0.60$ (95% CI 0.39 to 0.74)). HR_{PFS} showed a pattern of underestimating treatment effects, particularly when HR_{OS} was between 0.75 and 1, which resulted in a STE of 1.21 (**figure 2A**). For patient-level correlation between PFS and OS, the median r_s was 0.66 (95% CI 0.65 to 0.67; range, 0.65 (95% CI 0.64 to 0.66) to 0.68 (95% CI 0.67 to 0.69) among the 1000 simulations).

Performance of mPFS as a surrogate endpoint in ICI trials

Figure 2B illustrates the trial-level correlation between HR_{mPFS} and HR_{OS} , and patient-level correlation between mPFS and OS when the cut-off timepoint to define mPFS was changed from months 2 to 6. The correlation between HR_{mPFS} and HR_{OS} dramatically improved when the cut-off was changed from months 2 to 3, and reached a stabilized plateau at the later cut-off timepoints (**figure 2B**, left panel); this pattern remained consistent after stratified by cancer type, therapeutic regimen, and treatment setting (online supplemental figure S5). Notably, when the cut-off was at month 3 or later timepoints, the correlation between HR_{mPFS} and HR_{OS} was markedly stronger than that between HR_{PFS} and HR_{OS} . The median r_p was 0.81 (95% CI 0.69 to 0.90; range, 0.77 (95% CI 0.58 to 0.87) to 0.84 (95% CI 0.75 to 0.91)) when the cut-off was at month 3; the median STE was 1.01 (range, 0.99–1.04) at this cut-off timepoint.

At the patient level, the correlation between mPFS and OS was stronger than that between PFS and OS at all examined cut-off timepoints (**figure 2B**, right panel). The median r_s was 0.79 (95% CI 0.78 to 0.80; range, 0.78 (95% CI 0.77 to 0.79) to 0.80 (95% CI 0.80 to 0.81)) when the cut-off was at month 3, although r_s increased proportionally at the later cut-off timepoints. However, the number of mPFS events diminished continuously when the cut-off was elevated from months 3 to 6 (online supplemental figure S6). When the cut-off was at month 3, the maximum of the relative reduction in the number of events (ie, the difference in mPFS and PFS event numbers divided by PFS event number) varied from 1.7% to 33.5% by trial. More importantly, the number of mPFS (cut-off at month 3) events was consistently greater than that of OS events for all trials (online supplemental figure S6).

By weighing the improvement in the trial-level and patient-level correlation between mPFS and OS against the reduction in the number of events when the cut-off was elevated from months 3 to 6, we eventually selected month 3 as the cut-off to define mPFS. For conciseness, “mPFS” is still used to indicate this concept in the following text. **Figure 2C** shows a representative example of the excellent agreement

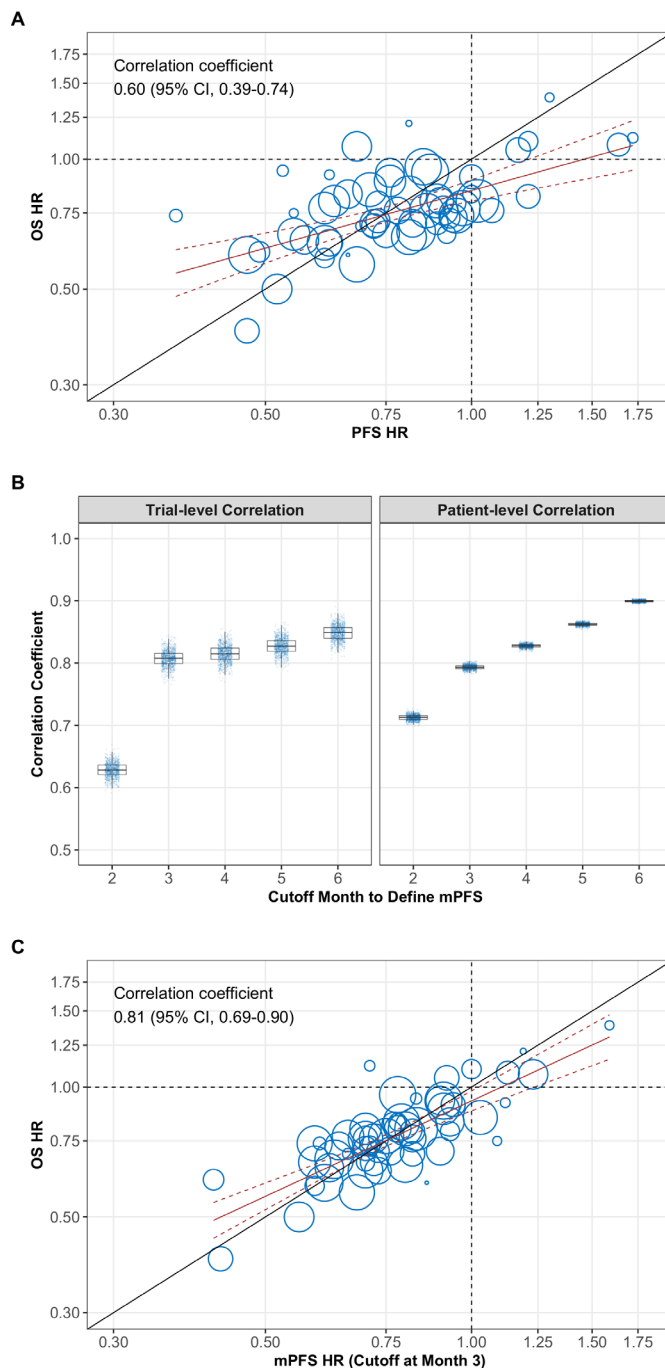


Figure 2 Performance of progression-free survival (PFS) and modified PFS (mPFS) as surrogate endpoints for overall survival (OS) in immuno-oncology trials. (A) Trial-level correlation between HRs for PFS and OS. Size of circles is proportional to trial sample size. (B) Trial-level correlation between HRs for mPFS and OS, as well as patient-level correlation between mPFS and OS by cut-off timepoint for defining mPFS among the reconstructed, matched PFS-OS datasets. (C) Trial-level correlation between HRs for mPFS (cut-off at month 3) and OS in an exemplar simulation where the correlation coefficient equals its median. Size of circles is proportional to trial sample size.

between HR_{mPFS} and HR_{OS} when r_p was equal to its median. In order to investigate whether the decrease in the number of events would lead to the decrease in the statistical power

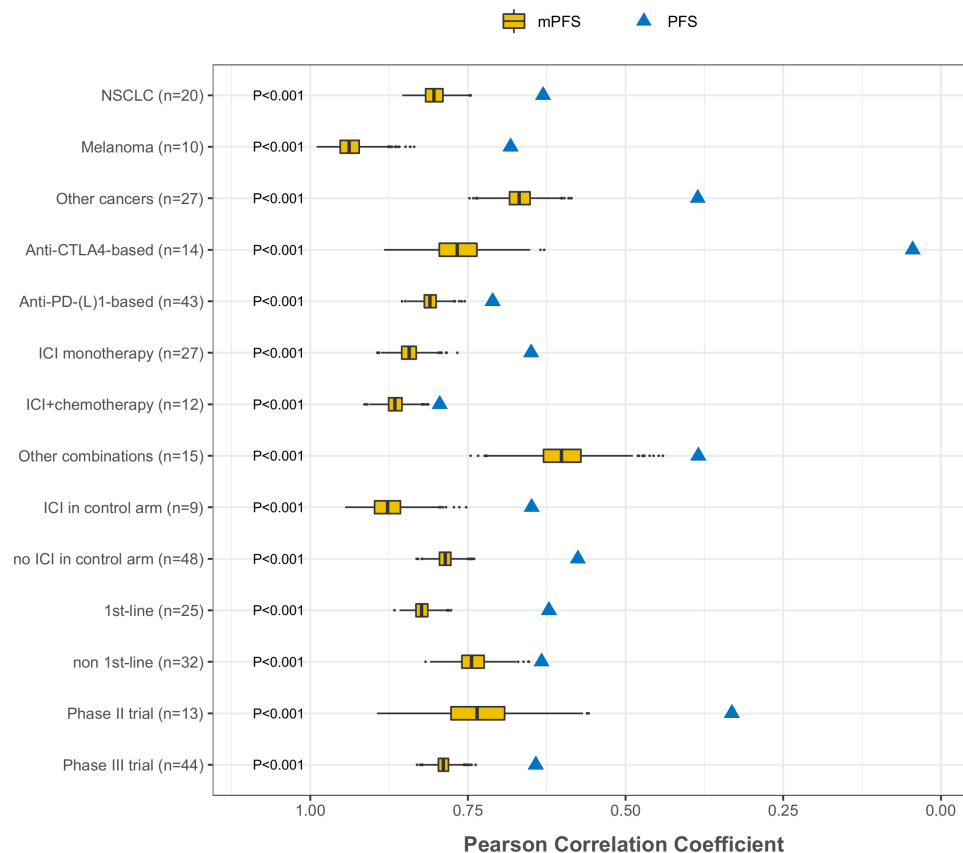


Figure 3 Correlation between HRs for progression-free survival (PFS) and overall survival vs correlation between HRs for modified PFS (mPFS) and overall survival by cancer type, therapeutic regimen, and treatment setting. The x-axis shows the trial-level correlation between PFS HRs and OS HRs as well as between mPFS HRs and OS HRs. Anti-PD-1/PD-L1 plus anti-CTLA4 trials (n=5) were only included in the subgroup analysis of anti-CTLA4 trials. One-sided Wilcoxon signed-rank test p values are shown for examined subgroups.

with mPFS, we further analyzed the log-rank test Z statistic of mPFS versus PFS in 32 trials where the experimental treatment significantly improved OS. As shown in online supplemental figure S7, mPFS showed a greater Z statistic than PFS in 19 (59.4%) of these trials.

Subgroup analysis

As shown in figure 3, the superiority of mPFS over PFS was consistent in subgroup analyses by cancer type, therapeutic regimen, and treatment setting. Notably, the correlation between PFS HR and OS HR was found to be much worse in trials involving anti-CTLA4 therapy than in those involving anti-PD-(L)1 therapy, with PFS HR either underestimating or overestimating treatment effects in the former subgroup (online supplemental figure S8). Five of the included trials involved anti-CTLA4 and anti-PD-(L)1 combination therapy. In comparison with PFS HR, mPFS HR was closer to OS HR in four of these trials (online supplemental figure S9).

Validation of the utility of mPFS using original IPD from ICI trials

Two randomized ICI trials for advanced NSCLC (POPLAR²¹ and OAK²²) provided open access original IPD.²³ For the POPLAR study, there were more OS and PFS events in the open access original IPD than in the initial publication,²¹ owing to a longer follow-up duration (median, 21.8 vs 14.8

months). For the OAK study, the numbers of OS and PFS events in the open access original IPD were identical to those reported in the initial publication.²²

In both the POPLAR and OAK studies, atezolizumab significantly improved OS compared with docetaxel (unstratified HR 0.68 (95% CI 0.51 to 0.89), two-sided log-rank test p=0.006, and HR 0.73 (95% CI 0.62 to 0.86), p<0.001, respectively; figure 4A,B). However, PFS did not differ between treatment arms in the POPLAR (HR 0.91 (95% CI 0.71 to 1.17), p=0.471) and OAK (HR 0.94 (95% CI 0.81 to 1.08), p=0.381) studies (figure 4C,D). In contrast, we observed a significant difference in mPFS in favor of atezolizumab in both the POPLAR (HR 0.74 (95% CI 0.57 to 0.96), p=0.021) and OAK (HR 0.64 (95% CI 0.55 to 0.84), p<0.001) studies (figure 4E,F).

For patient-level correlation between mPFS and OS, the r_s was 0.80 (95% CI 0.74 to 0.86) in the POPLAR study and 0.80 (95% CI, 0.76 to 0.83) in the OAK study, greater than the r_s between PFS and OS (0.70 (95% CI 0.62 to 0.77) in POPLAR and 0.64 (95% CI 0.59 to 0.69) in OAK).

DISCUSSION

In this large-scale analysis of reconstructed IPD from up-to-date ICI trials, we demonstrated that the proposed

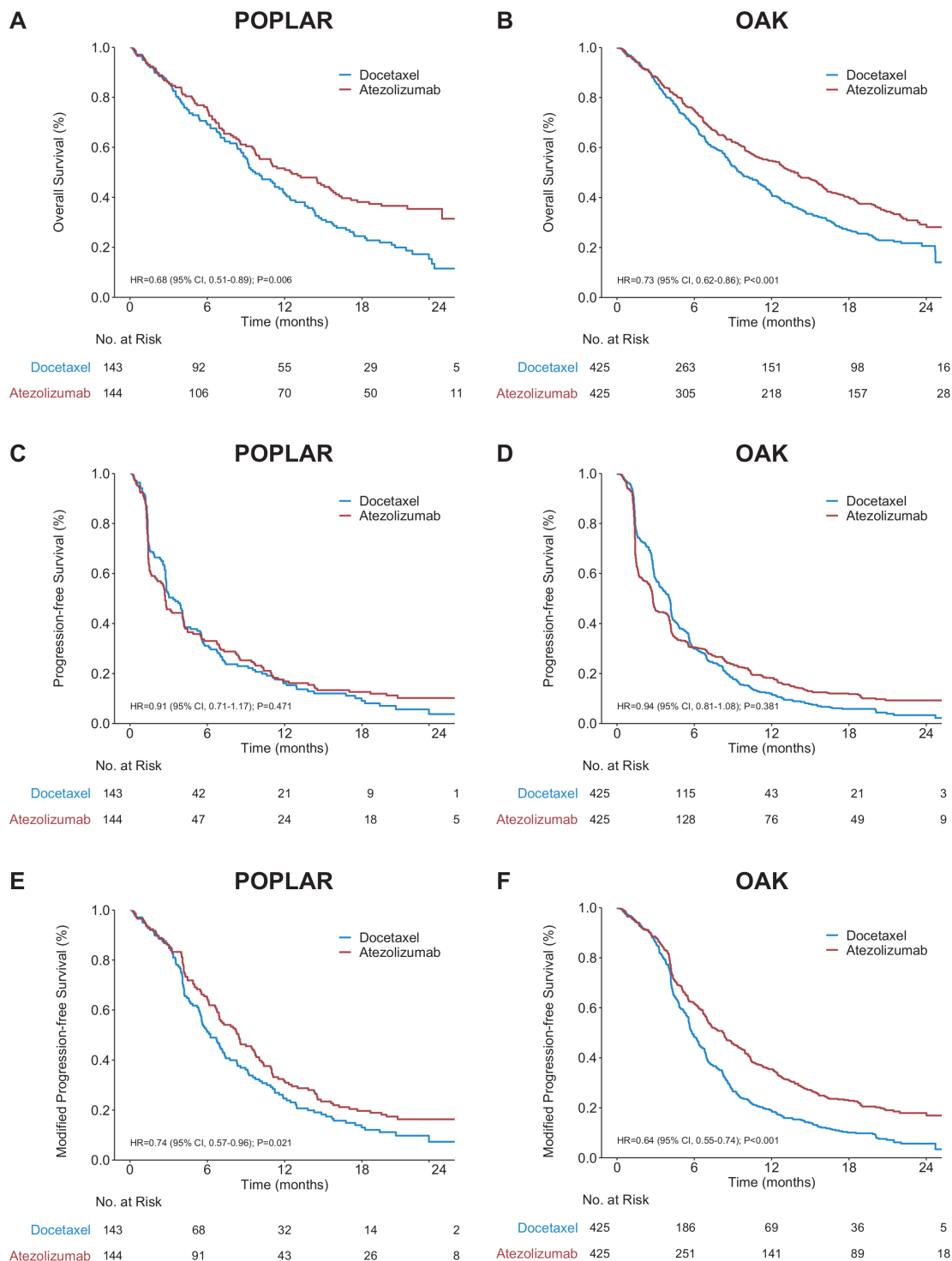


Figure 4 Kaplan-Meier curves for overall survival (A, B), progress-free survival (C, D), and modified progress-free survival (E, F) in the POPLAR and OAK studies. P values from two-sided, unstratified log-rank tests are shown.

mPFS achieved stronger correlation with OS than PFS at both the trial level and patient level. The superiority of mPFS was further verified using the original IPD from two NSCLC ICI trials. Together, these findings indicate that mPFS outperforms PFS as the surrogate endpoint for OS in ICI trials.

The RECIST-based PFS has been commonly used as the surrogate endpoint for OS to evaluate novel anti-cancer therapies in clinical trials, as well as in the accelerated

approval process of new therapies.²⁴ However, for ICI trials, the treatment effects on PFS exhibited moderate-to-poor correlation with OS.^{8,9} Specifically, PFS can fail to capture the OS benefits of ICIs, as seen in the POPLAR and OAK studies.^{21, 22} In such cases, effective therapies may not have been approved if PFS was adopted as the primary endpoint. Notably, with simple modification to the definition of PFS events, the proposed mPFS gained an improved ability to predict OS benefits while retaining

the advantages of PFS, as it accounts for a greater number of events than OS and is unlikely to be affected by subsequent therapies after progression. As such, using mPFS as the surrogate endpoint for OS may shorten trial durations and accelerate the discovery of efficacious ICI-based therapies.

The improved correlation between mPFS and OS may be explained by the removal of “low-quality PD events” within 3 months from randomization. We observed a dramatic increase in r_p between mPFS and OS when the cut-off was changed from months 2 to 3, followed by a stabilized plateau thereafter despite a continuous reduction in the number of events, which is consistent with prior observations that pseudoprogression generally occurred within 3 months after randomization,^{25 26} and supports our speculation that “low-quality PD events” (either pseudoprogression or PD followed by a long post-PD survival) mainly occur in the first 3 months after randomization. It is true that the biological rationale for omitting all the PD events within 3 months is not definite, as these PD events may not necessarily be low-quality PD events. However, mPFS can still be readily interpreted in a biologically reasonable manner, which is an important attribute for a surrogate endpoint. As mPFS required only minor modifications to PFS, the number of the omitted PDs relative to the PFS event number was only minimal to moderate across all trials. Thereby, mPFS effects can be interpreted similarly to PFS effects, informing the efficacy of the experimental treatment in decreasing the risk of death or disease progression.

As early PD events were replaced by subsequent deaths or censored at the last follow-up, the improved surrogacy of mPFS versus PFS could come at the cost of a prolonged follow-up duration. Despite this, we found that mPFS HR was more closely correlated with OS HR and less likely to underestimate treatment effects compared with PFS HR, and that mPFS showed a significantly greater Z statistic than PFS in the majority of trials. Therefore, mPFS may require a smaller number of events than PFS to achieve the same power, and hence may not necessarily require a longer follow-up duration than PFS in trials with the event-driven design.

The emerging immune-related criteria for assessing tumor response represent another way of optimizing the definition of PD, hence PFS. There have been several proposals for such criteria, from the Immune-Related Response Criteria (irRC 2009)¹⁰ to the immune-related RECIST (irRECIST 2013),¹¹ and the latest Immune RECIST (iRECIST 2017) guidelines.¹² Most of these criteria have centered on how to incorporate new lesions into the total tumor burden or how to define additional tumor response patterns that may occur after the initial tumor expansion. However, the lack of consensus limits the widespread use of these tools in ICI trials, making it difficult to assemble sufficient trial data to validate immune-related PFS or ORR as surrogate endpoints for ICIs. In this context, our proposed mPFS, which required only minor modification of PFS and was verified using

large-scale ICI trial data across multiple cancer types and treatment settings, may serve as a simple and practical endpoint in ICI trials.

It should be noted that mPFS was developed only for the design and data analysis of ICI trials. Therefore, using mPFS as the endpoint does not suggest that ICIs should always be continued beyond PD within 3 months; instead, a trial can employ mPFS as an endpoint to quantify the efficacy of ICIs, meanwhile complementing RECIST with the immune-related criteria to guide clinical decision-making on the (dis)continuation of these therapies for patients. The ultimate solution to precise identification of “low-quality PD” may rely on the breakthrough in predictive biomarkers, particularly liquid biopsies and plasma-based markers that are useful for on-treatment monitoring.^{27–29}

Subgroup analysis suggested much poorer surrogacy of PFS in trials investigating anti-CTLA4 therapy compared with those investigating anti-PD-(L)1 therapy, with PFS HR either underestimating or overestimating treatment effects in the former subgroup. The underestimation of treatment effects by PFS HR is possibly due to the delayed effects with ICIs, whereas the overestimation of treatment effects might be attributed to the higher toxicity of anti-CTLA4 therapy than that of anti-PD-(L)1 therapy. As a representative example, in the CA184-043 study where the PFS HR was smaller than OS HR (0.70 vs 0.85), the PFS curves overlapped in the first 3 months, whereas on-study deaths in the first 6 months were more common in the ipilimumab group than in the placebo group.³⁰ By omitting the early PD events, mPFS effectively mitigated the underestimation and overestimation of treatment effects in trials investigating anti-CTLA4 therapy.

A limitation of this study is that we analyzed reconstructed IPD rather than the original IPD. However, the methods we used for IPD reconstruction were verified with excellent accuracy and reproducibility in previous studies.^{31–33} In addition, our findings based on reconstructed IPD and simulations were only validated using original IPD from NSCLC trials, and hence further validation is required among other cancer types. Still, our subgroup analysis consistently showed the superior surrogacy of mPFS versus PFS across cancer type, suggesting that the delayed clinical effect of ICIs is a pan-cancer generic phenomenon and mPFS might be useful across multiple cancer types. Finally, as only a small number of the published ICI trials employed both the traditional PFS and immune-related PFS as trial endpoints, it is not feasible to evaluate the comparative surrogacy of mPFS versus immune-related PFS.

In summary, the proposed mPFS is superior to PFS as the surrogate endpoint for OS in ICI trials. mPFS is worthy of further investigation as a secondary endpoint in future ICI trials. We also call for efforts from investigators of published ICI trials to proactively validate mPFS using the original IPD. Once validated, mPFS could be used in future ICI trials to shorten trial durations, save time and

costs, and more importantly, accelerate the clinical utility of effective therapies to patients.

Author affiliations

¹Department of Medical Oncology, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, P. R. China

²Research Unit of Precision Diagnosis and Treatment for Gastrointestinal Cancer, Chinese Academy of Medical Sciences, Guangzhou, P. R. China

³Department of Clinical Research, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, P. R. China

⁴Clinical Research Center, Shanghai Jiao Tong University School of Medicine, Shanghai, P. R. China

⁵Department of Biostatistics, Zhongshan Hospital, Fudan University, Shanghai, P. R. China

⁶Department of Medical Oncology, Tongji University Shanghai East Hospital, Shanghai, P. R. China

⁷Medical Review Division 1, Center for Drug Evaluation, National Medical Products Administration, Beijing, P. R. China

Acknowledgements We thank Dr. Jiwei Zhao (Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, USA) for his suggestions for data analysis. We thank Dr. Melvin L.K. Chua (Division of Radiation Oncology and Medical Sciences, National Cancer Center Singapore, Singapore) for his suggestions for data interpretation. We thank Dr. Chong-Yang Duan (Department of Biostatistics, Southern Medical University, Guangzhou, China) for his suggestions for the revision of this work. We thank our native language editor, Mr. Christopher Lavender (Sun Yat-sen University Cancer Center, Guangzhou, P. R. China) for his assistance in editing this manuscript. Mr. Lavender is an employee of Sun Yat-sen University Cancer Center and did not receive compensation for his contribution.

Contributors Z-XW, H-XW, LX, W-HL, Z-MY, and R-HX designed the study. R-HX, JL, and Z-MY supervised the study. Z-XW and H-XW contributed to the identification and selection of the trials, and the collection and checking of the data. Z-XW, LX, and W-HL performed the statistical analysis. All the authors interpreted the data. All the authors were involved in the drafting, review, and approval of the manuscript, and the decision to submit for publication.

Funding This work was supported by grants from the National Key R&D Program of China (2018YFC1313300).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethical approval was waived since we used only publicly available data and materials in this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. The data that support the findings of this study are available on request from the corresponding author.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Rui-Hua Xu <http://orcid.org/0000-0001-9771-8534>

REFERENCES

- 1 Khalil DN, Smith EL, Brentjens RJ, *et al*. The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. *Nat Rev Clin Oncol* 2016;13:273–90.
- 2 Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Science* 2018;359:1350–5.
- 3 Constantiniidou A, Alifieris C, Trafalis DT. Targeting programmed cell death-1 (PD-1) and ligand (PD-L1): a new era in cancer active immunotherapy. *Pharmacol Ther* 2019;194:84–106.
- 4 Rowshanravan B, Halliday N, Sansom DM. CTLA-4: a moving target in immunotherapy. *Blood* 2018;131:58–67.
- 5 Hegde PS, Chen DS. Top 10 challenges in cancer immunotherapy. *Immunity* 2020;52:17–35.
- 6 Mick R, Chen T-T. Statistical challenges in the design of late-stage cancer immunotherapy studies. *Cancer Immunol Res* 2015;3:1292–8.
- 7 Chen T-T. Statistical issues and challenges in immuno-oncology. *J Immunother Cancer* 2013;1:18.
- 8 Mushti SL, Mulkey F, Sridhara R. Evaluation of overall response rate and progression-free survival as potential surrogate endpoints for overall survival in immunotherapy trials. *Clin Cancer Res* 2018;24:2268–75.
- 9 Nie R-C, Chen F-P, Yuan S-Q, *et al*. Evaluation of objective response, disease control and progression-free survival as surrogate endpoints for overall survival in anti-programmed death-1 and anti-programmed death ligand 1 trials. *Eur J Cancer* 2019;106:1–11.
- 10 Wolchok JD, Hoos A, O'Day S, *et al*. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res* 2009;15:7412–20.
- 11 Nishino M, Giobbie-Hurder A, Gargano M, *et al*. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. *Clin Cancer Res* 2013;19:3936–43.
- 12 Seymour L, Bogaerts J, Perrone A, *et al*. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* 2017;18:e143–52.
- 13 Abdel-Rahman O. Nonconventional patterns of benefit of solid tumors treated with PD-(L)1 inhibitors: a systematic review. *Immunotherapy* 2017;9:995–1004.
- 14 Long GV, Weber JS, Larkin J, *et al*. Nivolumab for patients with advanced melanoma treated beyond progression: analysis of 2 phase 3 clinical trials. *JAMA Oncol* 2017;3:1511–9.
- 15 Robert C, Long GV, Brady B, *et al*. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* 2015;372:320–30.
- 16 Blumenthal GM, Zhang L, Zhang H, *et al*. Milestone analyses of immune checkpoint inhibitors, targeted therapy, and conventional therapy in metastatic non-small cell lung cancer trials: a meta-analysis. *JAMA Oncol* 2017;3:e171029.
- 17 Guyot P, Ades AE, Ouwers MJNM, *et al*. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol* 2012;12:9.
- 18 Xie W, Halabi S, Tierney JF, *et al*. A systematic review and recommendation for reporting of surrogate endpoint evaluation using meta-analyses. *JNCI Cancer Spectr* 2019;3:pkz002.
- 19 Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat* 2006;5:173–86.
- 20 Burzykowski T, Molenberghs G, Buyse M. The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. *J R Stat Soc Ser A Stat Soc* 2004;167:103–24.
- 21 Fehrenbacher L, Spira A, Ballinger M, *et al*. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* 2016;387:1837–46.
- 22 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;389:255–65.
- 23 Gandara DR, Paul SM, Kowanetz M, *et al*. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med* 2018;24:1441–8.
- 24 Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist* 2008;13 Suppl 2:19–21.
- 25 Katz SI, Hammer M, Bagley SJ, *et al*. Radiologic pseudoprogression during anti-PD-1 therapy for advanced non-small cell lung cancer. *J Thorac Oncol* 2018;13:978–86.
- 26 Necchi A, Joseph RW, Loriot Y, *et al*. Atezolizumab in platinum-treated locally advanced or metastatic urothelial carcinoma: post-progression outcomes from the phase II Ivigor210 study. *Ann Oncol* 2017;28:3044–50.

- 27 Bratman SV, Yang SYC, Iafolla MAJ, *et al.* Personalized circulating tumor DNA analysis as a predictive biomarker in solid tumor patients treated with pembrolizumab. *Nat Cancer* 2020;1:873–81.
- 28 Lee JH, Long GV, Menzies AM, *et al.* Association between circulating tumor DNA and pseudoprogression in patients with metastatic melanoma treated with anti-programmed cell death 1 antibodies. *JAMA Oncol* 2018;4:717–21.
- 29 Cabel L, Riva F, Servois V, *et al.* Circulating tumor DNA changes for early monitoring of anti-PD1 immunotherapy: a proof-of-concept study. *Ann Oncol* 2017;28:1996–2001.
- 30 Kwon ED, Drake CG, Scher HI, *et al.* Ipilimumab versus placebo after radiotherapy in patients with metastatic castration-resistant prostate cancer that had progressed after docetaxel chemotherapy (CA184-043): a multicentre, randomised, double-blind, phase 3 trial. *Lancet Oncol* 2014;15:700–12.
- 31 Royston P, Parmar MKB. 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 Med Res Methodol* 2013;13:152.
- 32 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;34:1813–9.
- 33 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;32:2380–5.