Targeted Oncology (2025) 20:45–56
https://doi.org/10.1007/s11523-024-01108-2

SYSTEMATIC REVIEW

Delayed Separation of Kaplan–Meier Curves is Commonly Observed in Studies of Advanced/Metastatic Solid Tumors Treated with Anti-PD-(L)1 Therapy: Systematic Review and Meta-Analysis

Do-Youn Oh[1] · Nana Rokutanda[2] · Magdalena Żotkiewicz[3] · Philip He[4] · Jennifer Stocks[2] · Melissa L. Johnson[5]

Accepted: 7 October 2024 / Published online: 10 November 2024

Abstract

Background Immune checkpoint inhibitor (ICI) Kaplan–Meier (KM) curves often show delayed survival benefit followed by long-term survival in a subgroup of patients. Such outcomes can violate the proportional hazards assumption, leading to a loss of statistical power.

Objective We aimed to determine common trends in delayed separation to inform future ICI clinical trials.

Patients and Methods A literature search was performed using Trialtrove® to identify phase III trials of antiprogrammed cell death (ligand)-1 (anti-PD-[L]1) agents in locally advanced/metastatic solid tumors published between January 2010 and September 2021. The frequency of delayed separation of overall survival (OS) and progression-free survival (PFS) KM curves, correlation between duration of delayed separation in OS/PFS KM curves, and correlation between duration of delayed separation in OS/PFS KM curves with corresponding hazard ratios (HRs) were assessed in all-comer and PD-L1 enriched populations.

Results Eighty-five studies with OS/PFS KM curves were identified. Most studies showed delayed separation of OS (> 67.9%) and PFS (> 54.5%) KM curves. The correlation between the duration of delayed separation in OS/PFS KM curves was strongest in the PD-L1 enriched population (adjusted $R^2 = 0.66$). No correlation was seen between the duration of delayed separation of OS KM curves and OS HR. A modest correlation was seen between the duration of delayed separation of PFS KM curves and PFS HR in all-comer and PD-L1 enriched populations (adjusted $R^2 = 0.24$ and 0.31, respectively).

Conclusions Delayed separation of KM curves was common in clinical trials of anti-PD-(L)1 agents. Understanding delayed separation is key to clinical study designs and assessing outcomes with ICIs.

Key Points

Meta-analyses of anti-PD-(L)1 agent trials showed delayed separation of KM curves.

Combining anti-PD-(L)1 agents with chemotherapy may reduce the delayed separation.

Delayed separation may be less likely in patient populations with high PD-L1 expression.

With delayed separation, a positive study outcome may be seen with longer follow-up.

Delayed separation should be considered in immune checkpoint inhibitor trials.

Extended author information available on the last page of the article

1 Introduction

Immune checkpoint inhibitors (ICIs) have transformed the treatment landscape across multiple tumor types [1–5] and many agents, including anti-programmed cell death (ligand)-1 (anti-PD-[L]1) inhibitors, and have demonstrated durable clinical activity both as monotherapy and combination therapy [4–8].

Previous studies have shown that ICIs exhibit different patterns of clinical benefit compared with other/non-ICI therapies, making survival analysis challenging [9]. Several factors may contribute to the challenges in demonstrating survival benefit, including the underlying mechanism of action of ICIs and the underlying tumor response [10]. Clinical trials comparing ICIs with other/non-ICI therapies have shown both conventional and non-conventional response patterns [10]. Conventional response patterns can include

nonresponse, or response with a reduction in tumor burden and/or stable disease, whereas non-conventional response patterns can include a delayed survival effect, followed by long-term survival in a subgroup of patients [10].

The nonconventional response patterns of ICIs can manifest in Kaplan–Meier (KM) curves when comparing ICIs with other/non-ICI therapies. For example, KM curves for ICIs and other/non-ICI therapies often overlap or crossover during early follow-up due to a delayed survival effect with ICIs, and a clear separation between the two KM curves only becomes apparent several months after starting ICI treatment [11]. Despite showing delayed separation, ICIs can still provide durable responses and long-term survival benefits [11, 12]. The presence of long-term survivors can generally be characterized by the appearance of a flattened, long tail (plateau) in the survival curve of the ICI treatment arm [10, 12]. This phase can occur many months after initial treatment and can persist long after treatment has ceased [10, 12].

In order for new cancer treatments to be approved by regulatory agencies, they need to demonstrate statistically significant improvements in clinically meaningful endpoints of their clinical trials, which may include overall survival (OS) and/or progression-free survival (PFS) [13]. Median OS/PFS and hazard ratios (HRs) are typically used to estimate the treatment effect for survival endpoints in oncology clinical trials [10]. The HR is usually calculated from a Cox proportional hazards model, which assumes that the ratio of two hazard functions is constant over time. However, when a delayed separation, early crossover, or plateau of KM curves occurs, the proportional hazards assumption may be violated [10, 14]. This can result in a loss of statistical power and the inability to demonstrate a significant difference in survival between treatment arms [10].

We performed a systematic review and meta-analysis of phase III randomized clinical trials of anti-PD-(L)1 agents in patients with locally advanced or metastatic solid tumors. This study aimed to determine common trends in delayed separation to inform future ICI clinical trial design and analysis.

2 Methods

This meta-analysis was reported in line with the preferred reporting items for systematic reviews and meta-analysis (PRISMA) and the PRISMA extension statement for network meta-analysis [15]. The review was not registered, and a protocol was not prepared.

2.1 Data Sources and Searches

An electronic literature search was conducted using the Trialtrove® database (Trialtrove® Citeline, 2023) to identify studies published between January 2010 and September 2021. The data cut-off of January 2010 was selected to capture published studies for phase III clinical trials prior to the approval of anti-PD-(L)1 agents in the USA [16]. Search terms included oncology, phase III, PD-(L)1 antagonist, ICIs, randomized, and controlled. The search strategy is detailed in the Data Supplement (Appendix 1).

2.2 Study Selection

Only phase III randomized controlled trials of anti-PD-(L)1 agents in patients with locally advanced or metastatic solid tumors were included. Eligible publications included full papers, congress abstracts, posters, and oral presentations. Publications not in the English language and those in the neoadjuvant/adjuvant setting were excluded.

2.3 Data Extraction

The main text and supplementary materials for each study were evaluated for data extraction. Jennifer Stocks and Nana Rokutanda extracted and summarized the following data using a standardized form for each study: year of publication, study ID, tumor type, total number of randomized patients, study population, line of therapy, treatment regimen, OS or PFS HR, and associated KM curves. Only the latest results for each study and study cohorts (i.e., no substudies) were extracted. Primary publications and secondary publications reporting long-term or extended follow-up could be included. For studies with multiple experimental treatment arms, each experimental arm was considered as an individual result and listed separately.

2.4 Quality Assessment

After data extraction, three independent reviewers (Jennifer Stocks, Magdalena Żotkiewicz, and Nana Rokutanda) reviewed the full text of relevant publications for final inclusion. If there were disagreements, a fourth reviewer (Philip He) was included. Disagreements were resolved by consensus.

2.5 Data Synthesis and Statistical Analysis

For analysis, extracted studies were categorized into two subgroups: the all-comer population, which included all patients enrolled in a study regardless of PD-L1 expression,

and the PD-L1 enriched population, which included all patients enrolled in a study with high PD-L1 expression, as per the original study protocol. If a single high PD-L1 expression cut-off was used within a study, all patients with PD-L1 expression above this cut-off were included in the PD-L1 enriched population, regardless of sample size. If multiple PD-L1 expression cut-offs were used within a study, the highest cut-off was selected, provided that the sample size included 100 or more patients per treatment arm. If the sample size included fewer than 100 patients per treatment arm, the next highest cut-off was selected.

Studies were further categorized by tumor type and treatment regimen within each subgroup. Tumor type categories included lung, gastrointestinal (GI), or other (genitourinary, head and neck cancer, melanoma, women's cancers, or other solid tumors). Treatment categories included anti-PD-(L)1 monotherapy versus placebo/best supportive care (BSC)/observation, anti-PD-(L)1 monotherapy versus chemotherapy (CT)/other standard of care (SoC), and anti-PD-(L)1 plus CT versus CT/other SoC.

For many studies under consideration with an HR cut-off > 0.80 but ≤ 0.85, there was an apparent trending treatment effect between KM curves (i.e., trend for improved OS or PFS when comparing the anti-PD-[L]1 treatment arm with the comparator arm). However, when an HR cut-off ≤ 0.80 was applied, the number of studies available for analysis was limited. Thus, an HR cut-off ≤ 0.85 was selected to evaluate the trending treatment effect and delayed separation. Studies with available OS and PFS HRs were first evaluated to determine if there was a trending treatment effect for the anti-PD-(L)1 treatment arm compared with the comparator arm. This assessment was performed separately in both the all-comer and PD-L1 enriched populations. Studies could be included in the all-comer and PD-L1 enriched populations if they showed a trending treatment effect for OS but not PFS and vice versa. Study comparisons with a trending treatment effect for OS and PFS were then manually assessed to determine if a delayed separation or early crossing of KM curves occurred. Delayed separation was defined as a 2-month or longer delay in KM curve separation from the start of the study when comparing the anti-PD-(L)1 treatment with the comparator arm. This definition was based on the Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 guidelines, which recommend that scheduled assessments for tumor assessments be performed every 6–8 weeks (< 2 months) following treatment initiation [17]. An early crossing of KM curves was defined as a worsening of KM curves at the start of the study, followed by improvement (separation), when comparing the anti-PD-(L)1 treatment arm with the comparator arm.

In study comparisons with a trending treatment effect and delayed separation of OS and PFS KM curves, the duration of delayed separation was estimated using the least squares mean method. The survival time point before the latest separation of OS and PFS KM curves for the anti-PD-(L)1 treatment arm compared with the comparator arm was used as the basis for this calculation. The duration of delayed separation was quantified using a precision level of 0.25 months. The duration of delayed separation of OS and PFS KM curves in the all-comer population versus the PD-L1 enriched population was compared by the mean difference of duration of delay. For studies without delayed separation, the duration of delay was assigned as zero. Wilcoxon signed rank tests were performed to compare the duration of OS and PFS delay in studies that showed a delayed separation of both OS and PFS KM curves.

The correlation between the duration of delayed separation in OS and PFS, and corresponding HRs, was assessed by multivariate linear regression analysis weighted by sample size. The model included the duration of delayed separation in OS as a dependent variable and the duration of delayed separation in PFS as an explanatory variable, adjusted for tumor type and treatment regimen. The coefficient of determination (adjusted R^2) was used to determine whether the duration of delayed separation in PFS could predict the duration of delayed separation in OS, and whether the delayed separation could predict the HR.

Data analysis was performed using R (version 4.1.2) and Statistical Analysis Software (SAS, version 9.4, SAS Inc.).

3 Results

3.1 Systematic Review and Characteristics

The electronic search yielded 179 studies; 175 were screened for inclusion (Fig. 1). In total, 90 studies were excluded during the screening process (Fig. 1). Studies with cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) regimens, targeted therapy, and those where the comparator arm included immuno-oncology or non-checkpoint inhibitor immunotherapy agents, were excluded due to the small number of studies available (Fig. 1). After screening, 85 unique studies had available OS or PFS KM curves, including 137 unique comparisons between the anti-PD-(L)1 treatment arm and the comparator arm (Fig. 1). A summary of studies included in the meta-analysis is shown in the Data Supplement (Appendix 2).

3.2 Determination of a Trending Treatment Effect

Most study comparisons showed a trending treatment effect (HR ≤ 0.85) for OS and PFS in the all-comer (67.1% and 63.0%, respectively) and PD-L1 enriched (80.0% and 71.0%, respectively) populations (Table 1). Anti-PD-(L)1 monotherapy versus placebo/BSC/observation was the most

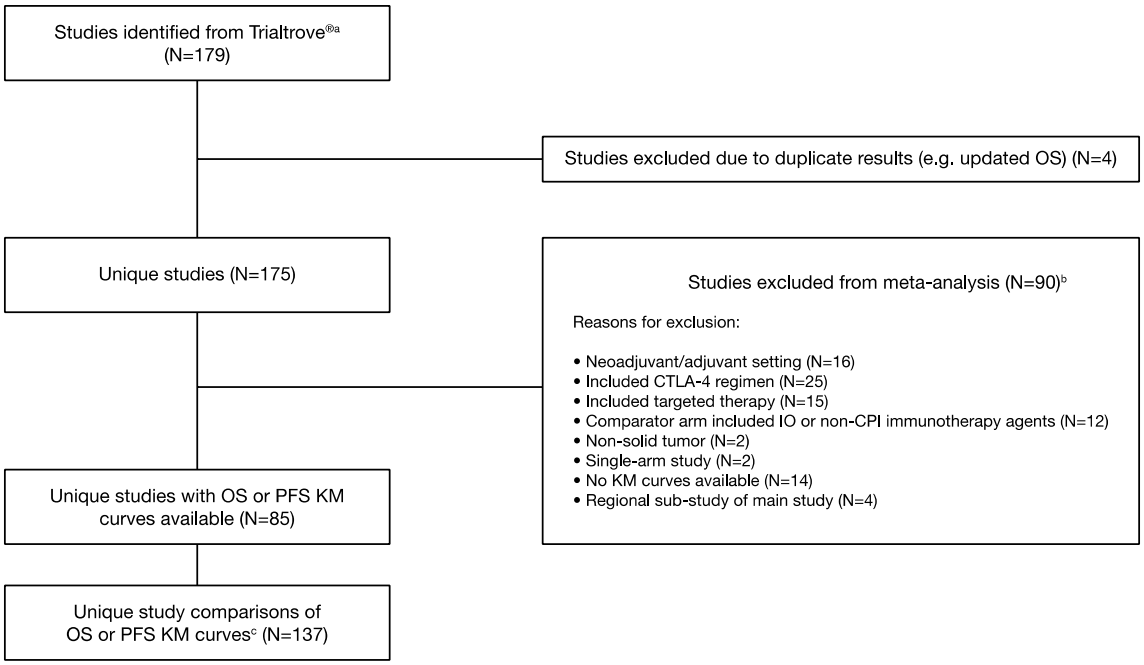

Fig. 1 Flowchart of the study selection procedure. [a]Search terms and conditions included oncology, phase III, PD-L1 antagonist, immune checkpoint inhibitor, randomized, and controlled. For a complete list, please see the Data Supplement (Appendix 1). [b]Some studies were excluded for more than one exclusion criteria. [c]Some studies have both all-comer and PD-L1 enriched populations, or multiple experimental arms versus the same control (e.g., anti-PD-[L]1 plus CT versus CT/other SoC and anti-PD-[L]1 monotherapy versus CT/other SoC). "Comparison" is the statistical analysis performed to compare OS or PFS for individuals treated with one experimental regimen versus control in the all-comer or PD-L1 enriched populations. *CPI*, checkpoint inhibitor; *CT*, chemotherapy; *IO*, immuno-oncology; *KM*, Kaplan–Meier; *OS*, overall survival; *PD-(L)1*, programmed cell death (ligand)-1; *PFS*, progression-free survival; *SoC*, standard of care

common treatment regimen associated with a trending treatment effect for OS (100.0%) and PFS (100.0%) in the all-comer population (Table 1). The most common treatment regimen associated with a trending treatment effect in the PD-L1 enriched population was anti-PD-(L)1 plus CT versus CT/other SoC (OS 91.7%; PFS 100.0%) (Table 1). In the all-comer population, the most common cancer type associated with a trending treatment effect was GI cancer (OS 81.3%) or lung cancer (PFS 87.1%) (Table 1). In the PD-L1 enriched population, lung cancer was the most common cancer type associated with a trending treatment effect (OS 92.3%; PFS 85.7%) (Table 1).

3.3 Delayed Separation of KM Curves

A delayed separation (at least 2 months) of OS and PFS KM curves occurred in the majority of study comparisons with a trending treatment effect (HR $\leq$ 0.85) in the all-comer (73.6% and 58.7%, respectively) and PD-L1 enriched (67.9% and 54.5%, respectively) populations (Table 1). Anti-PD-(L)1 monotherapy versus CT/other SoC was the most common treatment regimen associated with a delayed separation of KM curves (all-comer, OS 80.0%, PFS 85.7%; PD-L1 enriched, OS 70.6%, PFS 77.8%) (Table 1). The most common cancer type associated with a delayed separation of KM curves was "other" cancers (genitourinary, head and neck cancer, melanoma, women's cancers, or other solid tumors) (all-comer, OS 93.3%, PFS 90.0%; PD-L1 enriched, OS 85.7%, PFS 80.0%) (Table 1).

3.4 Early Crossing of KM Curves

In studies with a trending treatment effect (HR $\leq$ 0.85), an early crossing of OS and PFS KM curves occurred in 13.2% and 6.5% of studies in the all-comer population and 21.4% and 9.1% of studies in the PD-L1 enriched population, respectively (Table 1). Anti-PD-(L)1 monotherapy versus CT/other SoC was the most common treatment regimen associated with an early crossing of KM curves (all-comer, OS 30.0%, PFS 28.6%; PD-L1 enriched, OS 35.3%, PFS 22.2%) (Table 1). No specific tumor type was associated with an early crossing of KM curves (Table 1).

3.5 Duration of Delayed Separation of KM Curves

In studies with a trending treatment effect (HR $\leq$ 0.85), the least squares mean duration of delayed separation of KM curves was longest with anti-PD-(L)1 plus CT versus CT/

Table 1 Summary of studies included in the meta-analysis with trending treatment effect (HR ≤ 0.85) and delayed separation (at least 2 months) or an early crossing of KM curves

Trending treatment effect, *n*/*N* (%)	Overall survival		Progression-free survival	
	All-comer (*N* = 79)	PD-L1 enriched (*N* = 35)	All-comer (*N* = 73)	PD-L1 enriched (*N* = 31)
Overall	53/79 (67.1)	28/35 (80.0)	46/73 (63.0)	22/31 (71.0)
Treatment regimen				
Anti-PD-(L)1 monotherapy versus CT/other SoC	20/38 (52.6)	17/23 (73.9)	7/30 (23.3)	9/18 (50.0)
Anti-PD-(L)1 monotherapy versus placebo/BSC/observation	7/7 (100.0)	0	7/7 (100.0)	0
Anti-PD-(L)1 + CT versus CT/other SoC	26/34 (76.5)	11/12 (91.7)	32/36 (88.9)	13/13 (100.0)
Tumor type				
Lung	25/31 (81.0)	12/13 (92.3)	27/31 (87.1)	12/14 (85.7)
Gastrointestinal	13/16 (81.3)	9/10 (90.0)	9/14 (64.3)	5/7 (71.4)
Other[a]	15/32 (46.9)	7/12 (58.3)	10/28 (35.7)	5/10 (50.0)
Delayed separation of KM curves in studies with a trending treatment effect, *n*/*N* (%)	All-comer (*N* = 53)	PD-L1 enriched (*N* = 28)	All-comer (*N* = 46)	PD-L1 enriched (*N* = 22)
Overall	39/53 (73.6)	19/28 (67.9)	27/46 (58.7)	12/22 (54.5)
Treatment regimen				
Anti-PD-(L)1 monotherapy versus CT/other SoC	16/20 (80.0)	12/17 (70.6)	6/7 (85.7)	7/9 (77.8)
Anti-PD-(L)1 monotherapy versus placebo/BSC/observation	4/7 (57.1)	0	4/7 (57.1)	0
Anti-PD-(L)1 + CT versus CT/other SoC	19/26 (73.1)	7/11 (63.6)	17/32 (53.1)	5/13 (38.5)
Tumor type				
Lung	15/25 (60.0)	8/12 (66.7)	13/27 (48.1)	6/12 (50.0)
Gastrointestinal	10/13 (76.9)	5/9 (55.6)	5/9 (55.6)	2/5 (40.0)
Other[a]	14/15 (93.3)	6/7 (85.7)	9/10 (90.0)	4/5 (80.0)
Early crossing of KM curves in studies with a trending treatment effect, *n*/*N* (%)	All-comer (*N* = 53)	PD-L1 enriched (*N* = 28)	All-comer (*N* = 46)	PD-L1 enriched (*N* = 22)
Overall	7/53 (13.2)	6/28 (21.4)	3/46 (6.5)	2/22 (9.1)
Treatment regimen				
Anti-PD-(L)1 monotherapy versus CT/other SoC	6/20 (30.0)	6/17 (35.3)	2/7 (28.6)	2/9 (22.2)
Anti-PD-(L)1 monotherapy versus placebo/BSC/observation	0/7 (0.0)	0	0/7 (0.0)	0
Anti-PD-(L)1 + CT versus CT/other SoC	1/26 (3.8)	0/11 (0.0)	1/32 (3.1)	0/13 (0.0)
Tumor type				
Lung	2/25 (8.0)	3/12 (25.0)	2/27 (7.4)	2/12 (16.7)
Gastrointestinal	2/13 (15.4)	1/9 (11.1)	1/9 (11.1)	0/5 (0.0)
Other[a]	3/15 (20.0)	2/7 (28.6)	0/10 (0.0)	0/5 (0.0)

Six studies had multiple experimental arms versus the same control and are counted multiple times in relevant categories

BSC, best supportive care; *CT*, chemotherapy; *HR*, hazard ratio; *KM*, Kaplan–Meier; *n*, number of evaluable studies; *N*, total number of studies; *PD-(L)1*, programmed cell death (ligand)-1; *SoC*, standard of care

[a]Other includes genitourinary, melanoma, head and neck, women's cancer, or other solid tumors

other SoC (OS 4.4 months) or anti-PD-(L)1 monotherapy versus CT/other SoC (PFS 3.3 months) in the all-comer population (Fig. 2A, C). In the PD-L1 enriched population, the least squares mean duration of delayed separation of KM curves was longest with anti-PD-(L)1 monotherapy versus CT/other SoC (OS 4.5 months; PFS 4.9 months) (Fig. 2B, D). In both the all-comer and PD-L1 enriched populations, the least squares mean duration of delayed

separation of KM curves was longest in "other" cancers (genitourinary, head and neck cancer, melanoma, women's cancers, or other solid tumors) (all-comer, OS 5.4 months, PFS 3.7 months; PD-L1 enriched, OS 5.2 months, PFS 4.8 months) (Fig. 2A–D).

3.6 Duration of Delayed Separation of KM Curves in the All-Comer Population Compared With the PD-L1 Enriched Population

Overall, the mean (95% CI) duration of delayed separation of OS and PFS KM curves was numerically longer in the all-comer population compared with the PD-L1 enriched population (OS 0.94 [− 0.22–2.11] months; PFS 0.71 [0.14–1.28] months) based on the paired comparison (Table 2).

3.6.1 Correlation Between the Duration of Delayed Separation of OS and PFS KM Curves

The correlation between the duration of delayed separation of OS and PFS KM curves was stronger in the PD-L1 enriched population (adjusted $R^2 = 0.66$) versus the all-comer population (adjusted $R^2 = 0.22$) (Fig. 3A, B). Findings were generally consistent when the correlation between the duration of delayed separation of OS and PFS KM curves was assessed by tumor type (lung cancer and GI cancer; Supplementary Fig. 1).

3.6.2 Correlation Between the Duration of Delayed Separation of OS and PFS KM Curves and Corresponding HRs

In the all-comer and PD-L1 enriched populations, no correlation was seen between the OS HR and the duration of

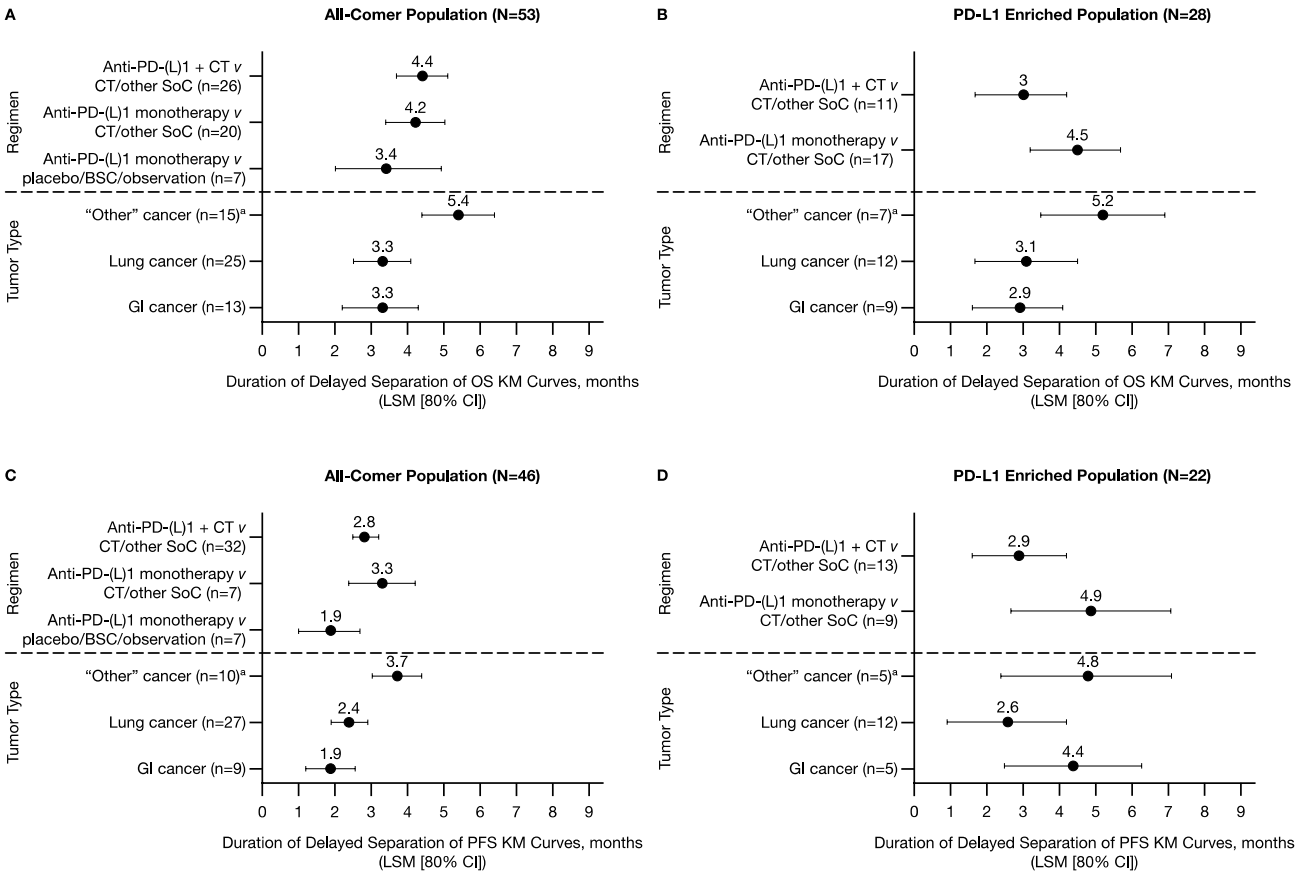

Fig. 2 Duration of delayed separation of OS and PFS KM curves by treatment regimen and tumor type in **A**, **C** the all-comer population and **B**, **D** the PD-L1 enriched population. [a]Includes genitourinary, head and neck cancer, melanoma, women's cancers, or other solid tumors. The analysis only included studies with a trending treatment effect (HR ≤ 0.85). One study can be used multiple times for each unique comparison. Forest plots show the least squares mean of duration of delay, by regimen and tumor types, using a weighted regression model based on each study's sample size. *BSC*, best supportive care; *CT*, chemotherapy; *GI*, gastrointestinal; *HR*, hazard ratio; *KM*, Kaplan–Meier; *LSM*, least squares mean; *OS*, overall survival; *PD-(L)1*, programmed cell death (ligand)-1; *PFS*, progression-free survival; *SoC*, standard of care

Table 2 Paired analysis of duration of delayed separation of OS and PFS KM curves in the all-comer versus PD-L1 enriched populations

Delay in all-comer population minus delay in PD-L1 enriched population	Number of studies	Difference in duration of delay in all-comer versus PD-L1 enriched population (min–max), months	Difference in mean duration of delay in all-comer versus PD-L1 enriched population (95% CI), months	Wilcoxon signed-rank test *P* value
OS	18	− 3.50–6.50	0.94 (− 0.22–2.11)	0.0884
PFS	12	− 0.50–2.25	0.71 (0.14–1.28)	0.0195

The analysis only included studies with trending treatment effect (HR ≤ 0.85). The OS analysis included 18 unique studies and 19 unique comparisons, and the PFS analysis included 12 unique studies and 12 unique comparisons

HR, hazard ratio; *KM*, Kaplan–Meier; *OS*, overall survival; *PD-L1*, programmed cell death ligand-1; *PFS*, progression-free survival

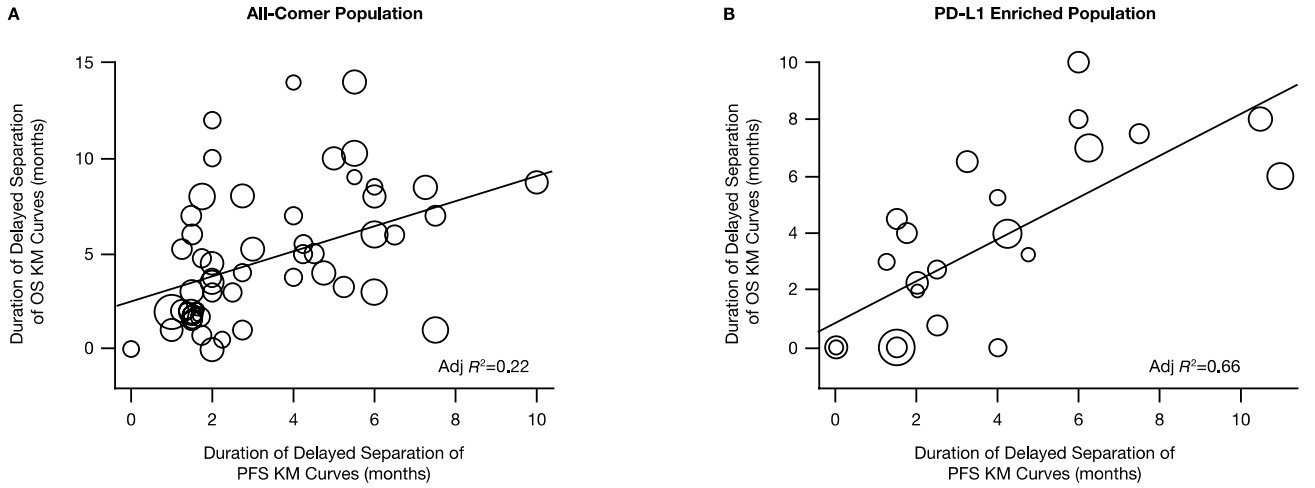

Fig. 3 Correlation between the duration of delayed separation of OS and PFS KM curves in **A** the all-comer population and **B** the PD-L1-enriched population. The area of each circle is proportional to each study's sample size. The analysis only included studies with a trending treatment effect (HR ≤ 0.85). One study can be used multiple times for each unique comparison. *Adj*, adjusted; *HR*, hazard ratio; *KM*, Kaplan–Meier; *OS*, overall survival; *PD-L1*, programmed cell death ligand-1; *PFS*, progression-free survival; R^2, regression coefficient

delayed separation of OS KM curves (adjusted $R^2 = -0.03$ and adjusted $R^2 = 0.18$, respectively) (Fig. 4A, B). Findings were generally consistent when the correlation between the duration of delayed separation of OS KM curves and corresponding HRs was assessed by tumor type (lung cancer and GI cancer; Supplementary Figs. 2 and 3, respectively). A modest correlation was seen between the PFS HR and the duration of delayed separation of PFS KM curves in the all-comer and PD-L1 enriched populations (adjusted $R^2 = 0.24$ and adjusted $R^2 = 0.31$, respectively) (Fig. 4C, D). The correlation between the PFS HR and the duration of delayed separation of PFS KM curves in the all-comer and PD-L1 enriched populations (lung cancer and GI cancer; Supplementary Figs. 2 and 3, respectively) appeared to be driven by studies that included patients with lung cancer only.

4 Discussion

In this systematic review and meta-analysis, we found that delayed separation of OS and PFS KM curves occurred in most studies with a trending treatment effect (HR ≤ 0.85) when comparing anti-PD-(L)1 agents with their respective control.

In both the all-comer and PD-L1 enriched populations, delayed separation of OS and PFS KM curves was most common in studies that compared anti-PD-(L)1 monotherapy versus CT/other SoC and least common in studies that compared anti-PD-(L)1 plus CT versus CT/other SoC. These findings suggest that combination treatment with anti-PD-(L)1 agents plus CT may help to drive an earlier separation of OS and/or PFS KM curves.

A delayed separation of KM curves was more common in studies that enrolled all-comer versus PD-L1 enriched populations. When comparing the duration of delayed separation

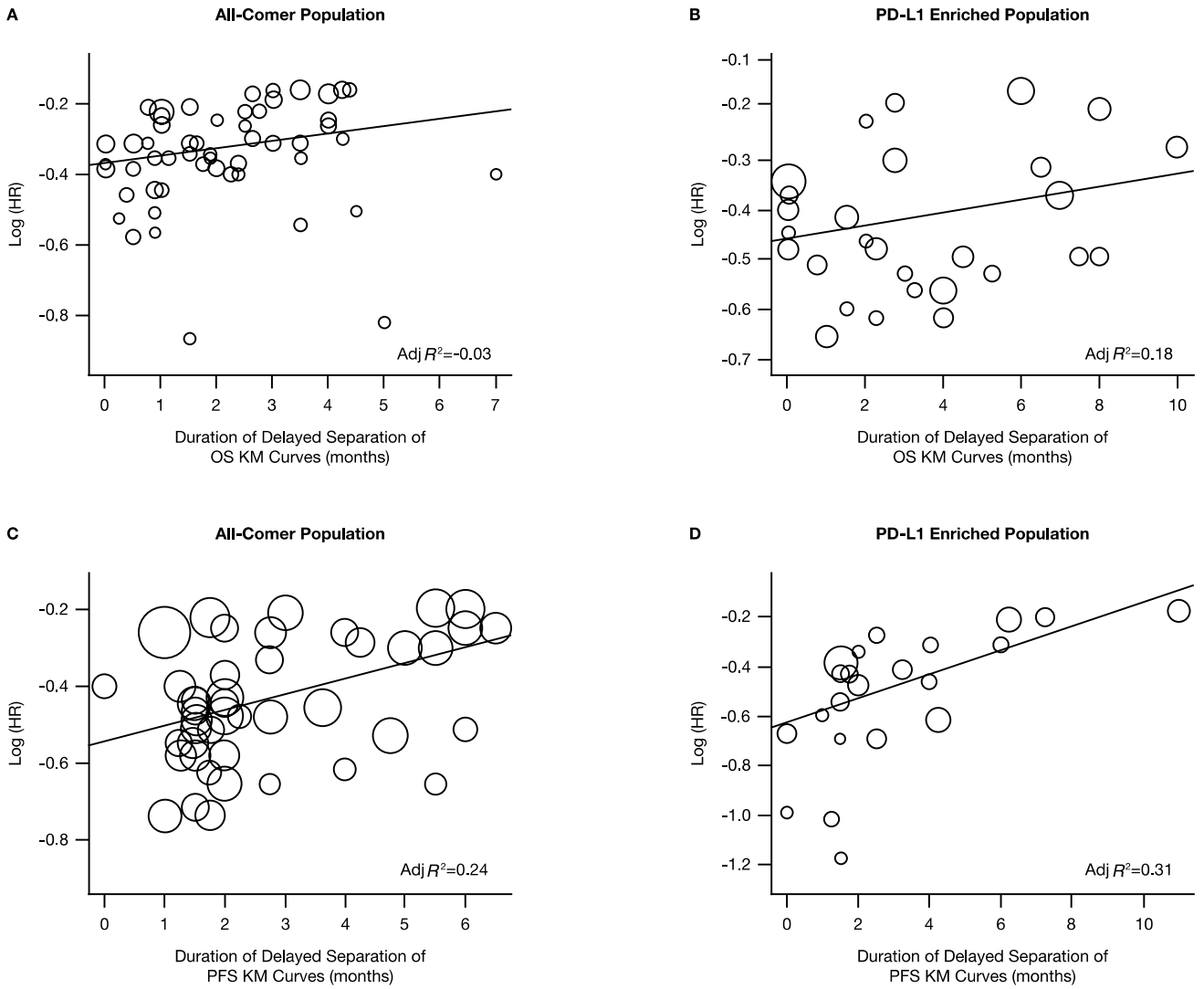

Fig. 4 Correlation between the duration of delayed separation of OS/PFS KM curves and corresponding HRs in **A**, **C** the all-comer population and **B**, **D** the PD-L1 enriched population. The area of each circle is proportional to each study's sample size. The analysis only included studies with trending treatment effect (HR ≤ 0.85). One study can be used multiple times for each unique comparison. *Adj*, adjusted; *HR*, hazard ratio; *KM*, Kaplan–Meier; *OS*, overall survival; *PD-L1*, programmed cell death ligand-1; *PFS*, progression-free survival; R^2, regression coefficient

in the all-comer versus PD-L1 enriched population, the mean duration of delayed separation was numerically shorter in the PD-L1 enriched population. In addition, a correlation between the delayed separation of OS and PFS KM curves was found in the PD-L1 enriched population but not in the all-comer population. Findings were generally consistent when the correlation between the duration of delayed separation of OS and PFS KM curves and corresponding HRs was assessed by tumor type (lung cancer and GI cancer). This observation suggests that in certain indications, patient populations with high PD-L1 expression may be less likely to show a delayed separation effect and may be more likely to have a positive study outcome. This observation may be due, in part, to PD-L1 expression enriching the efficacy of immunotherapy in certain indications.

While there was a modest correlation between the duration of delayed separation of PFS KM curves and PFS HR, no apparent correlation was seen for OS. The modest correlation of PFS KM curves and HR appeared to be driven by studies that included patients with lung cancer but not GI cancer. While limited by the small number of evaluable studies, the lack of correlation in studies that included patients with GI cancer may be due to the relatively large HR. These findings are consistent with the understanding that immunotherapy may have a modest impact on treating advanced GI cancer [18]. Regardless, findings imply that in the presence of delayed separation, a positive study outcome may still be possible, provided that there is sufficient

follow-up, highlighting the importance of adequate maturity and follow-up when designing ICI clinical trials.

It is well known that ICIs often exhibit a delayed separation of KM curves [10]. However, there has been no comprehensive analysis of the features of delayed separation of KM curves across anti-PD-(L)1 agents, to our knowledge. The results from our meta-analysis highlight the frequent occurrence of delayed separation with anti-PD-(L)1 agents compared with other/non-ICI therapies and suggest that many previous phase III clinical trials of anti-PD-(L)1 agents may not have considered potential delayed separation effects in their study design or statistical analyses. The results from our study are valuable as they shed light on the frequency of delayed separation with anti-PD-(L)1 agents and stress the need to consider patient and regimen selection, optimal endpoints, and appropriate statistical testing for the design of future ICI clinical trials.

When designing ICI clinical trials, treatment assumptions should include the possible delayed separation effects. For example, researchers should increase sample size and follow-up duration for event-driven trials, as the power loss may be significant in situations where there is a long duration of delayed separation. If the study has an interim analysis, it should be performed at an appropriate time point when the treatment effect is expected to be demonstrated.

The observed lack of a significant correlation between the delay in the separation of KM curves and HR in this meta-analysis suggests that the relevance of conventional study endpoints of median survival and HRs may need to be re-evaluated for studies with ICIs [1, 10, 19]. Considering this, alternative metrics that better suit the mechanism of action of ICIs should be explored, in addition to median survival and HRs, when designing clinical trials to capture the prolonged activity of therapy and characterize patients' treatment benefits. Long-term measures may include landmark survival analysis at clinically meaningful time points, restricted mean survival time, depth and durability of response, long-term survival, and treatment-free survival [20–22]. Of note, restricted mean survival time may be well suited to quantify the clinical benefit in situations where there is significant long-term survival in the experimental versus the control arm [23]. In support of this concept, the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS) was updated in 2017 to incorporate possible long-term survival with immunotherapy. For therapies that are not likely to be curative with a primary endpoint of OS, if median OS with the standard treatment is $\leq$ 12 months, an increase in 2-year survival $\geq$ 10% was added as one of the criteria for Grade 4 for a substantial magnitude of clinical benefit [24].

Alternative statistical models that account for delayed separation of the survival curve or long-term survival should also be considered in order to improve the planning and analysis of ICI clinical trials [25]. Diagnostics of non-proportional hazards for KM curves could be considered, such as log–log survival plots and hazard rate curves. In addition, piecewise exponential models could be used to describe different HRs before and after a specified timing of delayed onset of survival benefit [26], or mixed cure rate models could be used when the scientific rationale for the presence of long-term survivors is strong [27]. Weighted log-rank tests could also be used as an alternative to the traditional log-rank test to increase the power of the trial when non proportional hazards are expected [28, 29]. Of note, a recent systematic review and meta-analysis of 63 immuno-oncology clinical trials compared the log-rank test with the MaxCombo test (uses a combination of Fleming–Harrington weighted log-rank test and adaptively selects the best test statistic based on underlying data) [29]. Among 150 comparisons, the log-rank test was shown to be less powerful than the MaxCombo test and could not detect clinically meaningful benefits in 15 comparisons [29]. For example, in two case studies of clinical trials, MYSTIC [30] and Impower110 [31], which are also included in this meta-analysis and have nonproportional hazards due to a crossing of OS KM curves, the MaxCombo tests were able to detect a significantly greater improvement in OS compared with the control arm, whereas the log-rank test was not [29].

Lastly, physicians should recognize that ICIs have different patterns of treatment effects when compared with CT and targeted therapy [32]. Moreover, physicians should understand that treatment effects with ICIs can be delayed and can even occur after apparent disease progression on the initial post-baseline scan [33]. Thus, in addition to treating with ICIs until disease control and/or clinical benefit, physicians should consider continuing treatment beyond initial disease progression in patients who do not experience severe toxicities [32]. The concept of "treatment beyond progression" has been adopted in many trials evaluating ICIs [32].

The main limitations of this study include the definition of the PD-L1 enriched population, which was based on individual study reports and impacted by the use of different assays and cut-off criteria applied across studies, and the inclusion criteria for the delayed separation analysis, which only included studies with available KM curves and a trending treatment effect (HR $\leq$ 0.85); this may have contributed to bias in the analyses. Furthermore, due to the small sample size of individual cancer types, several cancer types were grouped for analysis and listed as "other" cancers. Interpretation of results in this subgroup may be limited as response characteristics and clinical settings may differ depending on the type of cancer. Lastly, the correlation analysis between the duration of delayed separation of OS/PFS KM curves and corresponding HRs in the all-comer and PD-L1 enriched populations may have been limited by OS and PFS parameters not being entirely separable, as PFS may include a component of OS [34].

5 Conclusion

In summary, delayed separation of KM curves is common in clinical trials with anti-PD-(L)1 therapy. Understanding delayed separation is key to future study designs and assessing outcomes with ICIs. The findings from this meta-analysis highlight the importance of patient and regimen selection, optimal endpoints, and appropriate statistical testing in ICI clinical trials.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11523-024-01108-2.

Acknowledgments Medical writing support, under the guidance of the authors, was provided by Eilidh McLachlan, PhD, on behalf of CMC Connect, a division of IPG Health Medical Communications, and was funded by AstraZeneca, in accordance with Good Publication Practice (GPP 2022) guidelines (Ann Intern Med 2022;175(9):1298-1304).

Declarations

Funding Open Access funding enabled and organized by Seoul National University Hospital.

Conflict of Interest D.-Y.O. has received research funding from Array, AstraZeneca, BeiGene, Eli Lilly, Handok, MSD, Novartis, and Servier, and consulting or advisory fees from Arcus Biosciences, ASLAN, AstraZeneca, Basilea, Bayer, BeiGene, Bristol Myers Squibb/Celgene, Genentech/Roche, Halozyme, IQVIA, Merck Serono, Novartis, Taiho, Turning Point, Yuhan, and Zymeworks. N.R. is an employee of and holds stock in AstraZeneca. M.Ż. is an employee of AstraZeneca. P.H. is an employee of and holds stock in Daiichi Sankyo. J.S. is an employee of and holds stock in AstraZeneca. M.L.J. has received research funding (to institution) from AbbVie, Acerta, Adaptimmune, Amgen, Apexigen, Arcus Biosciences, Array BioPharma, Artios Pharma, AstraZeneca, Atreca, BeiGene, BerGenBio, BioAlta, Black Diamond, Boehringer Ingelheim, Bristol Myers Squibb, Calithera Biosciences, Carisma Therapeutics, Checkpoint Therapeutics, City of Hope National Medical Center, Corvus Pharmaceuticals, Curis, CytomX, Daiichi Sankyo, Dracen Pharmaceuticals, Dynavax, Elicio Therapeutics, Eli Lilly, EMD Serono, EQRx, Erasca, Exelixis, Fate Therapeutics, Genentech/Roche, Genmab, Genocea Biosciences, GlaxoSmithKline, Gritstone Oncology, Guardant Health, Harpoon, Helsinn Healthcare SA, Hengrui Therapeutics, Hutchison MediPharma, IDEAYA Biosciences, IGM Biosciences, Immunitas Therapeutics, Immunocore, Incyte, Janssen, Jounce Therapeutics, Kadmon Pharmaceuticals, Kartos Therapeutics, Loxo Oncology, Lycera, Memorial Sloan Kettering, Merck, Merus, Mirati Therapeutics, Mythic Therapeutics, NeoImmune Tech, Neovia Oncology, Novartis, Numab Therapeutics, Nuvalent, OncoMed Pharmaceuticals, Palleon Pharmaceuticals, Pfizer, PMV Pharmaceuticals, Rain Therapeutics, RasCal Therapeutics, Regeneron Pharmaceuticals, Relay Therapeutics, Revolution Medicines, Ribon Therapeutics, Rubius Therapeutics, Sanofi, Seven and Eight Biopharmaceuticals/Birdie Biopharmaceuticals, Shattuck Labs, Silicon Therapeutics, Stem CentRx, Syndax Pharmaceuticals, Takeda Pharmaceuticals, Tarveda, TCR2 Therapeutics, Tempest Therapeutics, Tizona Therapeutics, TMUNITY Therapeutics, Turning Point Therapeutics, University of Michigan, Vyriad, WindMIL Therapeutics, and Y-mAbs Therapeutics, and received consulting or advisory fees (to institution) from AbbVie, Amgen, Arcus Biosciences, Arrivent, Astellas, AstraZeneca, Axelia Oncology, Black Diamond, Calithera Biosciences, Daiichi Sankyo, EcoR1, Genentech/Roche, Genmab, Genocea Biosciences, GlaxoSmithKline, Gritstone Oncology, Ideaya Biosciences, Immunocore, iTeos, Janssen, Jazz Pharmaceuticals, Merck, Mirati Therapeutics, Molecular Axiom, Novartis, Oncorus, Pyramid Biosciences, Regeneron Pharmaceuticals, Revolution Medicines, Sanofi-Aventis, SeaGen, Synthekine, Takeda Pharmaceuticals, Turning Point Therapeutics, and VBL Therapeutics.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Availability of Data and Material The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code Availability Not applicable.

Author Contributions N.R., J.S., M.Ż., and P.H. contributed to the study design and contributed to the acquisition and analysis of data. D.-Y.O. and N.R. supervised the study. All authors contributed to the interpretation of data, critically revised the manuscript, and provided final approval.

References

1. Michielin O, Lalani A-K, Robert C, Sharma P, Peters S. Defining unique clinical hallmarks for immune checkpoint inhibitor-based therapies. J Immunother Cancer. 2022;10:e003024.
2. Jacob JB, Jacob MK, Parajuli P. Review of immune checkpoint inhibitors in immuno-oncology. Adv Pharmacol. 2021;91:111–39.
3. Zhou F, Zhou C-C. Immunotherapy in non-small cell lung cancer: advancements and challenges. Chin Med J (Engl). 2021;134:1135–7.
4. Abou-Alfa GK, Lau G, Kudo M, Chan SL, Kelley RK, Furuse J, et al. Tremelimumab plus durvalumab in unresectable hepatocellular carcinoma. NEJM Evid. 2022;1:EVIDoa2100070.
5. Oh D-Y, He AR, Qin S, Chen L-T, Okusaka T, Vogel A, et al. Durvalumab plus gemcitabine and cisplatin in advanced biliary tract cancer. NEJM Evid. 2022;1:EVIDoa2200015.
6. Galsky MD, Arija JA, Bamias A, Davis ID, De Santis M, Kikuchi E, et al. Atezolizumab with or without chemotherapy in metastatic urothelial cancer (IMvigor130): a multicentre, randomised, placebo-controlled phase 3 trial. Lancet. 2020;395:1547–57.
7. Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal

junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. Lancet. 2021;398:27–40.
8. Gandhi L, Rodriguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med. 2018;378:2078–92.
9. Monnickendam G, Zhu M, McKendrick J, Su Y. Measuring survival benefit in health technology assessment in the presence of nonproportional hazards. Value Health. 2019;22:431–8.
10. Quinn C, Garrison LP, Pownell AK, Atkins MB, de Pouvourville G, Harrington K, et al. Current challenges for assessing the long-term clinical benefit of cancer immunotherapy: a multi-stakeholder perspective. J Immunother Cancer. 2020;8:e000648.
11. Horn L, Spigel DR, Vokes EE, Holgado E, Ready N, Steins M, et al. Nivolumab versus docetaxel in previously treated patients with advanced non-small-cell lung cancer: two-year outcomes from two randomized, open-label, phase III trials (CheckMate 017 and CheckMate 057). J Clin Oncol. 2017;35:3924–33.
12. Burtness B, Harrington KJ, Greil R, Soulières D, Tahara M, de Castro G Jr., et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. Lancet. 2019;394:1915–28.
13. US Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics. Guidance for industry. 2018. https://www.fda.gov/media/71195/download. Accessed September 29, 2022.
14. Lin EP, Hsu CY, Berry L, Bunn P, Shyr Y. Analysis of cancer survival associated with immune checkpoint inhibitors after statistical adjustment: a systematic review and meta-analyses. JAMA Netw Open. 2022;5:e2227211.
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
16. Gong J, Chehrazi-Raffle A, Reddi S, Salgia R. Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: a comprehensive review of registration trials and future considerations. J Immunother Cancer. 2018;6:8.
17. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228–47.
18. Hecht JR, Mitchell J, Morelli MP, Anandappa G, Yang JC. Next-generation approaches to immuno-oncology in GI cancers. Am Soc Clin Oncol Educ Book. 2023;43:e389072.
19. de Miguel M, Calvo E. Clinical challenges of immune checkpoint inhibitors. Cancer Cell. 2020;38:326–33.
20. Chen TT. Milestone survival: a potential intermediate endpoint for immune checkpoint inhibitors. J Natl Cancer Inst. 2015;107:dvj156.
21. Federico Paly V, Kurt M, Zhang L, Butler MO, Michielin O, Amadi A, et al. Heterogeneity in survival with immune checkpoint inhibitors and its implications for survival extrapolations: a case study in advanced melanoma. MDM Policy Pract. 2022;7:23814683221089659.
22. Delgado A, Guddati AK. Clinical endpoints in oncology—a primer. Am J Cancer Res. 2021;11:1121–31.
23. Wang Z-X, Wu H-X, Xie L, Wang Y-N, Yang L-P, He M-M, et al. Correlation of milestone restricted mean survival time ratio with overall survival hazard ratio in randomized clinical trials of immune checkpoint inhibitors: a systematic review and meta-analysis. JAMA Netw Open. 2019;2:e193433.
24. Cherny NI, Dafni U, Bogaerts J, Latino NJ, Pentheroudakis G, Douillard JY, et al. ESMO-magnitude of clinical benefit scale version 1.1. Ann Oncol. 2017;28:2340–66.
25. Anagnostou V, Yarchoan M, Hansen AR, Wang H, Verde F, Sharon E, et al. Immuno-oncology trial endpoints: capturing clinically meaningful activity. Clin Cancer Res. 2017;23:4959–69.
26. Schoenfeld DA, Finkelstein DM. Assessing survival benefit when treatment delays disease progression. Clin Trials. 2016;13:352–7.
27. Othus M, Barlogie B, Leblanc ML, Crowley JJ. Cure models as a useful statistical tool for analyzing survival. Clin Cancer Res. 2012;18:3731–6.
28. Su Z, Zhu M. Is it time for the weighted log-rank test to play a more important role in confirmatory trials? Contemp Clin Trials Commun. 2018;10:A1–2.
29. Mukhopadhyay P, Ye J, Anderson KM, Roychoudhury S, Rubin EH, Halabi S, et al. Log-rank test vs maxcombo and difference in restricted mean survival time tests for comparing survival under nonproportional hazards in immuno-oncology trials: a systematic review and meta-analysis. JAMA Oncol. 2022;8:1294–300.
30. Rizvi NA, Cho BC, Reinmuth N, Lee KH, Luft A, Ahn M-J, et al. Durvalumab with or without tremelimumab vs standard chemotherapy in first-line treatment of metastatic non-small cell lung cancer: the MYSTIC phase 3 randomized clinical trial. JAMA Oncol. 2020;6:661–74.
31. Herbst RS, Giaccone G, de Marinis F, Reinmuth N, Vergnenegre A, Barrios CH, et al. Atezolizumab for first-line treatment of PD-L1-selected patients with NSCLC. N Engl J Med. 2020;383:1328–39.
32. Borcoman E, Kanjanapan Y, Champiat S, Kato S, Servois V, Kurzrock R, et al. Novel patterns of response under immunotherapy. Ann Oncol. 2019;30:385–96.
33. Kreamer KM. Immune checkpoint blockade: a new paradigm in treating advanced cancer. J Adv Pract Oncol. 2014;5:418–31.
34. Belin L, Tan A, De Rycke Y, Dechartres A. Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review. Br J Cancer. 2020;122:1707–14.

Authors and Affiliations

Do-Youn Oh[1] · Nana Rokutanda[2] · Magdalena Żotkiewicz[3] · Philip He[4] · Jennifer Stocks[2] · Melissa L. Johnson[5]

✉ Do-Youn Oh
ohdoyoun@snu.ac.kr

Nana Rokutanda
nana.rokutanda@astrazeneca.com

Magdalena Żotkiewicz
magdalena.zotkiewicz@astrazeneca.com

Philip He
phe@dsi.com

Jennifer Stocks
jennifer.stocks@astrazeneca.com

Melissa L. Johnson
mjohnson@tnonc.com

[1] Division of Medical Oncology, Department of Internal Medicine, Cancer Research Institute, Seoul National University College of Medicine, Seoul National University Hospital, Seoul 03080, South Korea

[2] AstraZeneca, Gaithersburg, MD 20878, USA

[3] AstraZeneca, Warsaw, Poland

[4] Biostatistics and Data Management, Daiichi Sankyo, Basking Ridge, NJ 07920, USA

[5] Sarah Cannon Research Institute-Tennessee Oncology, Nashville, TN 37203, USA