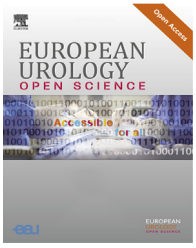
available at www.sciencedirect.com
journal homepage: www.eu-openscience.europeanurology.com



European Association of Urology

Bladder Cancer

Surrogate Endpoints as Predictors of Overall Survival in Metastatic Urothelial Cancer: A Trial-level Analysis

Fady Ghali [a,*], *Yibai Zhao* [b], *Devin Patel* [c], *Teresa Jewell* [d], *Evan Y. Yu* [e], *Petros Grivas* [e], *R. Bruce Montgomery* [e], *John L. Gore* [a], *Ruth B. Etzioni* [b], *Jonathan L. Wright* [a]

[a] *Department of Urology, University of Washington School of Medicine, Seattle, WA, USA;* [b] *Biostatistics Program, Fred Hutch Cancer Center, Seattle, WA, USA;* [c] *The Urology Clinic of Colorado, Denver, CO, USA;* [d] *Library Services, University of Washington School of Medicine, Seattle, WA, USA;* [e] *Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA*

Article info

Article history:
Accepted November 4, 2022

Associate Editor:
Guillaume Ploussard

Keywords:
Surrogate endpoints
Bladder cancer
Clinical trials

Abstract

Background: Surrogate endpoints (SEs), such as progression-free survival (PFS) and objective response rate (ORR), are frequently used in clinical trials. The relationship between SEs and overall survival (OS) has not been well described in metastatic urothelial cancer (MUC).
Objective: We evaluated trial-level data to assess the relationship between SEs and OS. We hypothesize a moderate surrogacy relationship between both PFS and ORR with OS.
Design, setting, and participants: We systematically reviewed phase 2/3 trials in MUC with two or more treatment arms, and report PFS and/or ORR, and OS.
Outcome measurements and statistical analysis: Linear regression was performed, and the coefficient of determination (R^2) and surrogate threshold effect (STE) estimate were determined between PFS/ORR and OS.
Results and limitations: Of 3791 search results, 59 trials and 62 comparisons met the inclusion criteria. Of the 53 trials that reported PFS, 31 (58%) reported proportional hazard regression for PFS and OS. Linear regression across trials demonstrated an R^2 of 0.60 between hazard ratio (HR) for PFS (HR^{PFS}) and HR for OS (HR^{OS}), and an STE of 0.41. Linear regression of ΔPFS (median PFS in months of the treatment arm – that of the control arm) and ΔOS demonstrated an R^2 of 0.12 and an STE of 14.1 mo. Thirty trials reported ORRs. Linear regression for ORR^{ratio} and HR^{OS} among all trials found an R^2 of 0.08; an STE of 95% was not reached at any value and ΔORR and HR^{OS} similarly demonstrated a poor correlation with an R^2 value of 0.03.
Conclusions: PFS provides only a moderate level of surrogacy for OS; An HR^{PFS} of $\leq$0.41 provides 95% confidence of OS improvement. ORR is weakly correlated with OS and should be de-emphasized in MUC clinical trials. When PFS is discussed, proportional hazard regression should be reported.
Patient summary: We examined the relationship between surrogate endpoints, common outcomes in clinical trials, with survival in urothelial cancer trials. Progression-free survival is moderately correlated, while objective response rate

* Corresponding author. Department of Urology, University of Washington School of Medicine, 318 10th Avenue E, Unit B7, Seattle, WA 98102, USA. Tel. +1 626 329 9705.
E-mail address: fghali@uw.edu (F. Ghali).

https://doi.org/10.1016/j.euros.2022.11.003

had a poor correlation with survival and should be de-emphasized as a primary endpoint.

1. Introduction

The selection of proper endpoints for clinical trials is imperative to the accurate interpretation of trial results, and to achieving the goal of novel therapies to prolong and/or improve the quality of patients' lives [1,2]. In oncology trials, overall survival (OS) is the gold-standard clinical endpoint. Surrogate endpoints (SEs), conversely, are measurable outcomes that are not intrinsically beneficial for patients, but are known or thought to predict a meaningful clinical benefit outcome, such as OS [3–5]. SEs are utilized because they shorten clinical trial times, and often sample size, resulting in decreased cost and quicker regulatory review with possible expedited access of novel therapies to patients [4,6–8].

Recently, SE utilization in therapy approval has increased, often without demonstrating an OS benefit [3,9–12]. This trend is in response to a 1992 policy shift by the Food and Drug Administration (FDA) allowing for the approval of certain therapies based on a demonstration of an SE benefit in a single phase 2 trial, rather than the prior requirement to demonstrate an OS benefit in phase 3 trials [3,5,13]. Careful attention to this trend is warranted as the relationship between SEs and OS may not be established fully in the context of each malignancy type, and thus an unverified assumption about clinical benefit undergirds a significant proportion of novel therapeutics [4,14].

Urothelial cancer (UC) is a frequent and aggressive malignancy, with 83 730 new cases and an estimated number of 17 200 deaths in 2021 [15]. Although the relationship between SEs and OS has been explored in multiple other malignancies evaluating trial-level data, metastatic UC (mUC) trials have not been examined. As the relationship between SEs and OS can also be a function of the biology of the specific cancer, as well as the class of therapy being evaluated (among several other potential confounders), endpoint surrogacy must be evaluated within each cancer therapy setting [3,16,17]; our analysis is focused on mUC. To address this gap in knowledge, we reviewed clinical trials in mUC to explore the relationship between commonly used surrogates: progression-free survival (PFS) and objective response rate (ORR), with OS. We hypothesize a moderate surrogacy relationship (R^2 of ~0.5–0.7 [18]) between PFS and ORR with OS.

2. Patients and methods

2.1. Database search

Search strings for PubMed and EMBASE (Elsevier) with the assistance of an information specialist used controlled vocabulary and free text terms for (1) UC, (2) advanced or metastatic stage, and (3) clinical trials with two or more arms. Databases were searched in August 2021. Search results were deduplicated in EndNoteX9 [19] and exported to Excel for title, abstract, and full-text reviews. Studies were included if these investigated mUC, had multiple arms, were randomized clinical trials, were not surgical or radiation trials, and reported one SE and OS. Meeting abstracts, nonrandomized clinical trials, prospective cohort, reviews, and retrospective studies were excluded. Studies that included both upper-tract UC and bladder cancer were included in the analysis, but studies that investigated only patients with upper tract tumors were excluded. Trials that did not report both OS, and PFS and/or ORR were excluded.

2.2. Selection strategy

Search results underwent a two-pass review for inclusion: a focused review of publication titles and abstracts was performed for initial screening. A secondary review of the text of articles and data abstraction was then carried out.

2.3. Data abstraction

Pertinent data were extracted from each manuscript. These included the first author, publication year, participant number, percentage (%) of male participants, crossover allowance in the study design, intervention/drug type, median follow-up, OS, PFS, and ORR for each arm, and hazard ratio (HR) for OS (HR^{OS}) and PFS (HR^{PFS}).

2.4. Data analysis

Median values and interquartile ranges (IQRs) for the number of patients, publication year, percent male participants, ΔPFS in months (median PFS in months of the treatment arm minus that of the control arm), HR^{PFS}, ORR^{ratio} (calculated as $ORR^{treatment}/ORR^{control}$), ΔORR (median ORR [%] of the treatment arm minus that of the control arm), HR^{OS}, and ΔOS (median OS [mo] of the treatment arm minus that of the control arm) were determined. Prespecified subgroups were described separately, including trials without crossover, trials with an immune checkpoint inhibitor (ICI) agent as a treatment-arm intervention, non-ICI trials, and trials with follow-up of longer than 24 mo.

The relationship between PFS and OS was evaluated in the following ways: linear regression between HR^{PFS} and HR^{OS}, and R^2 was computed. Linear regression between differences in median PFS (ΔPFS, in months) and ΔOS was also performed. We similarly evaluated the relationship between ORR and OS in two different ways: ORR^{ratio} and ΔOS, and ΔORR ($ORR^{treatment} - ORR^{control}$) and ΔOS were correlated by linear regression, and R^2 was computed. Additionally, the surrogate threshold effect (STE), the observed surrogate value that provides 95% confidence of an expected OS benefit, was calculated as follows: (1) linear regression is performed, (2) 95% predicted confidence interval (CI) bands are computed and graphed, and (3) if the dependent variable was HR^{OS}, then the STE was calculated by identifying the value at which the upper 95% predicted CI intercepts with the Y axis at $Y = 1$. When the dependent variable was ΔOS, then the STE was calculated by identifying the value at which lower 95% predicted CI intercepts the X axis ($Y = 0$) [17,20]. All analyses were performed using SPSS version 22.0 (IBM Corp., Armonk, NY, USA).

3. Results

3.1. Data collection

From the original search of two databases, 996 PubMed and 2795 Embase results were retrieved. After deduplication and manual screening, 3735 were excluded as those did not meet the criteria (Fig. 1); 59 trials and 62 comparisons were included in the analysis (Supplementary Table 1).

3.2. Trial description

The median trial sample size was 135 (IQR 85, 389), year of publication was 2016 (2007, 2020), HR^{PFS} was 0.86 (0.71, 1.03), ΔPFS was 0.2 (−1.55, 1.35) mo, ORR^{ratio} was 1.07 (0.76, 1.40), ΔORR was 3.0% (−10.0, 11.2%), HR^{OS} was 0.90 (0.80, 1.08), and ΔOS was 0.60 (−1.20, 2.58) mo. The median follow-up was 23.5 (14.9, 41.2) mo. Ten of 62 (16%) trials included crossover of the control arm to the treatment arm in the protocol, and 13 (21%) were evaluating immune checkpoint inhibition interventions (Table 1). Descriptive statistics of subgroups are shown in Table 1.

3.3. Correlation between SEs and OS

Of the 53 trials that reported PFS, 31 (58%) performed and reported proportional hazard regression for PFS and OS. Linear regression of all trials demonstrated an R^2 of 0.60 between HR^{PFS} and HR^{OS}, and the STE was calculated and found to be 0.41 (Fig. 2). Trials that did not allow crossover, ICI and non-ICI trials, trials with follow-up of >24 mo, and first-line and non–first-line trials were evaluated separately (Table 2).

Linear regression between ΔPFS and ΔOS demonstrated an R^2 of 0.12 and an STE of 14.1 mo (Fig. 2). Subgroups were analyzed with respect to ΔPFS and ΔOS (Table 2).

Linear regression for ORR^{ratio} and HR^{OS} including all trials demonstrated an R^2 of 0.08, and an STE of 95% was not

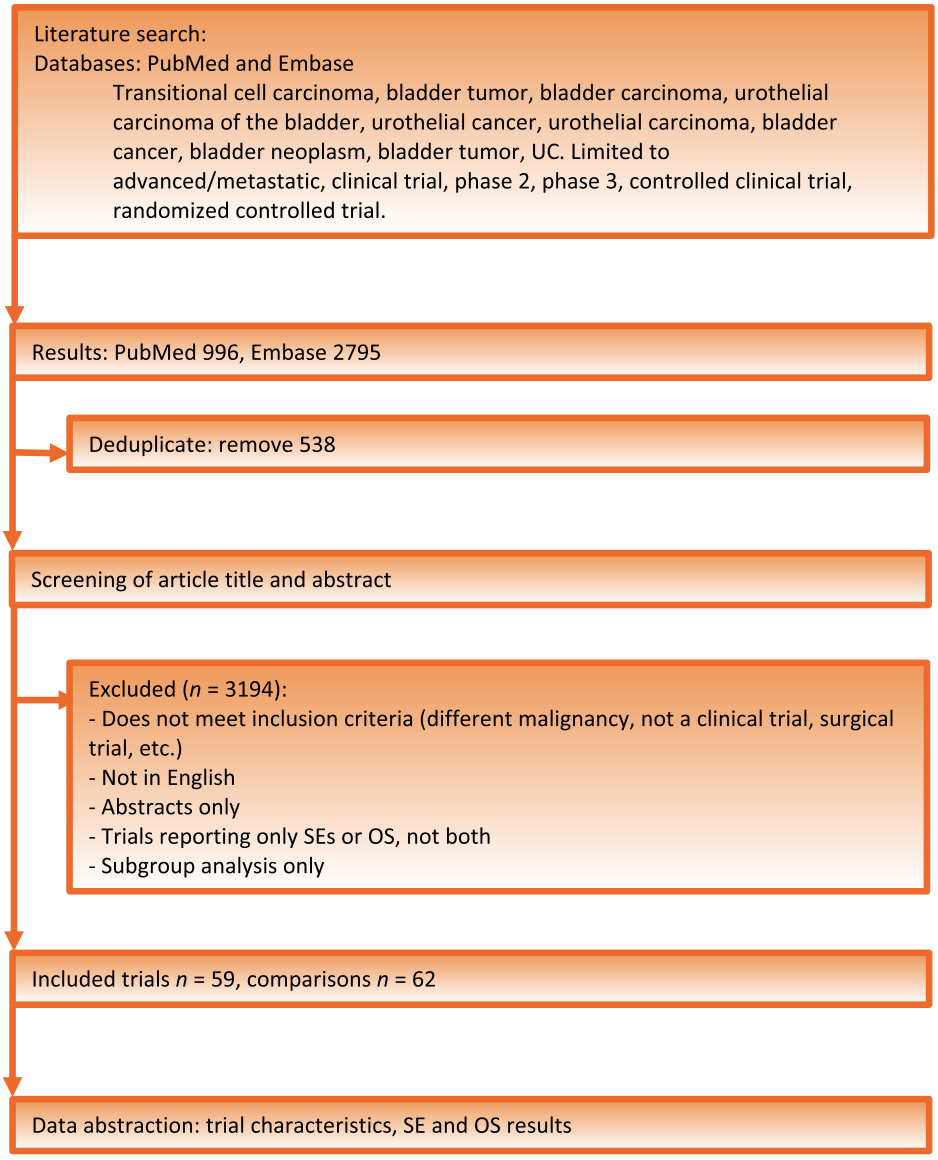


Fig. 1 – Systematic clinical trial review schema. OS = overall survival; SE = surrogate endpoint; UCC = urothelial carcinoma.

Table 1 – Descriptive analysis of trial comparisons of mUC

	All trials	No crossover	ICI	Non-ICI	Longer follow-up
Trial comparisons (%)	62 (100)	50 (81)	13 (21)	49 (79)	17 (27)
N	135 (85, 389)	121 (85, 370)	686 (176, 732)	110 (82, 237)	263 (110, 643)
% Male	76 (74, 80)	76 (74, 81)	75 (75,77)	77 (73, 81)	75 (74, 82)
Year of publication	2016 (2007, 2020)	2014 (2005, 2018)	2020 (2019, 2020)	2013 (2005, 2017)	2013 (2005, 2020)
HR^{PFS}	0.86 (0.71, 1.03)	0.85 (0.70, 1.06)	0.80, 0.64, 0.97)	0.87 (0.73, 1.07)	0.87 (0.75, 1.00)
ΔPFS (mo)	0.20 (−1.55, 1.35)	−0.10 (−1.80, 1.35)	0.60 (−1.90, 1.90)	0.20 (−1.53, 1.30)	−0.90 (−3.03, 1.23)
ORR^{ratio}	1.07 (0.76, 1.40)	1.04 (0.77, 1.35)	1.07 (0.70, 1.83)	1.05 (0.77, 1.35)	0.74 (0.67, 1.22)
ΔORR (%)	3.0% (−10.0, 11.2%)	2.5% (−10.5, 11.6%)	3.0% (−13.8, 9.9%)	2.6% (−8.8, 13.5%)	−9.3% (−16.5, 9.5%)
HR^{OS}	0.90 (0.80, 1.08)	0.89 (0.80, 1.08)	0.86 (0.73, 0.94)	0.94 (0.83, 1.14)	0.87 (0.75, 0.94)
ΔOS (mo)	0.60 (−1.20, 2.58)	0.55 (−1.23, 2.35)	2.60 (0.85, 3.15)	0.20 (−1.30, 1.90)	1.10 (−1.10, 2.50)

HR = hazard ratio; ICI = immune checkpoint inhibitor; mUC = metastatic urothelial cancer; ORR = objective response rate; OS = overall survival; PFS = progression-free survival.
Longer follow-up indicates trials with follow-up of ≥24 mo.

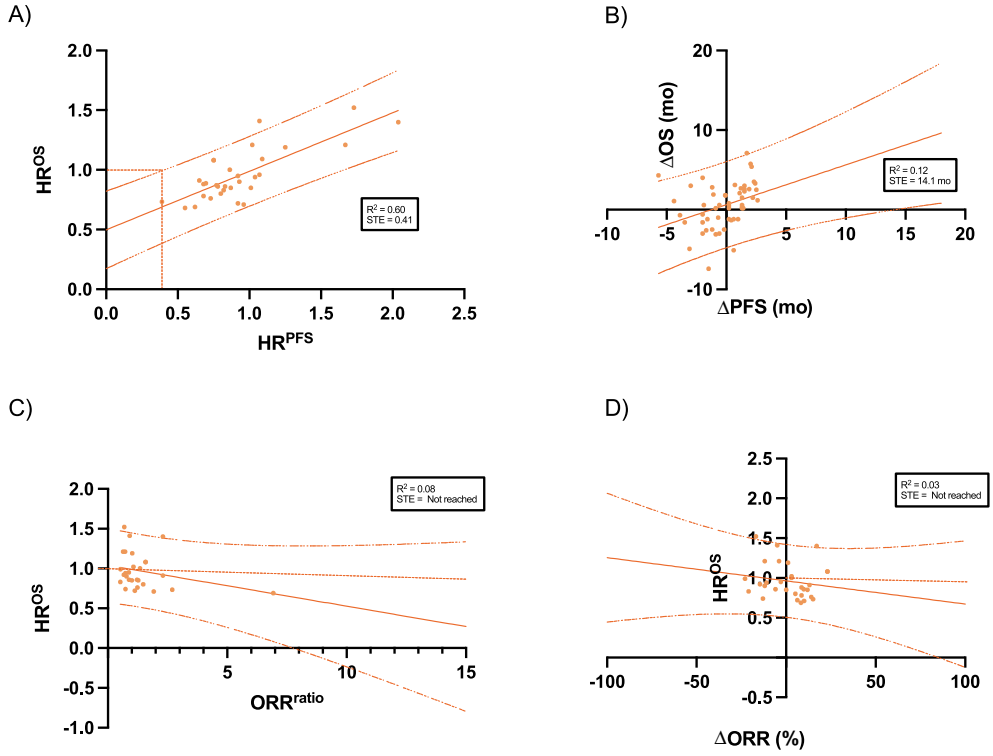


Fig. 2 – Linear regression analysis between (A) HR^{PFS} and HR^{OS}, (B) ΔPFS and ΔOS, (C) ORR^{ratio} and HR^{OS}, and (D) ΔORR and HR^{OS} among all trial comparisons. Longer follow-up indicates trials with follow-up of ≥24 mo. HR = hazard ratio; IO = immunotherapy; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; STE = surrogate threshold effect.

reached at any value (Fig. 2) or in any subgroup (Table 2). Linear regression for ΔORR and HR^{OS} demonstrated a poor correlation with an R^2 value of 0.03, and an STE of 95% was not reached (Fig. 2). Similarly, a subgroup analysis of ΔORR and HR^{OS} revealed low R^2 values, and no STE was found for any subgroup (Table 2).

4. Discussion

We report a systematic trial-level analysis of the relationship between candidate SEs and OS in MUC. Among available comparisons, there was a moderate correlation between HR^{PFS} and HR^{OR}, but a weak correlation between ΔPFS and ΔOS, and ΔORR or ORR^{ratio} with HR^{OS}. Despite a moderately strong R^2, an STE of 0.41 is computed for HR^{PFS}, while an STE for ORR was not reached. Taken together, our findings indicate that PFS is a moderately good surrogate for OS and an observed HR^{PFS} of ≤0.41 provides 95% confidence of an improvement in OS. An ORR, conversely, represents a poor surrogate for OS.

SEs are increasingly used in lieu of OS as primary endpoints in studies of novel cancer therapies [4,9]. Chen et al. [4] reported that since 1996, the rate of FDA drug approval based on SEs alone has increased dramatically, with <30% ultimately reporting requisite postmarket OS or quality of life data [21]. This trend is a response to FDA policy shifts implemented in the 1990s in the form of an accelerated approval tract intended for therapies to urgently life-

Table 2 – Coefficient of determination, R^2, and STE for PFS and ORR with OS including all trials, as well as key subgroups

PFS	*N*	R^2 HR^{PFS}	STE HR^{PFS}	*N*	R^2 ΔPFS	STE ΔPFS (mo)
All trials	31	0.60	0.41	53	0.12	14.10
No crossover	27	0.65	0.44	45	0.13	15.42
ICI	6	<0.01	NR	11	0.07	NR
Non-ICI	24	0.63	0.33	41	0.15	16.10
Longer follow-up	9	0.76	0.59	14	0.21	9.94
First line	16	0.48	0.24	28	0.02	NR
Non–first line	15	0.74	0.34	22	0.41	4.67
ORR with OS	*N*	**R^2ORR^{ratio}**	**STE ORR^{ratio}**	***N***	**R^2ΔORR**	**STE ΔORR (% difference)**
All trials	30	0.08	NR	30	0.03	NR
No crossover	25	0.05	NR	25	0.02	NR
ICI	10	0.17	NR	10	0.07	NR
Non–ICI	19	0.16	NR	19	0.09	NR
Longer follow-up[a]	11	0.17	NR	11	0.30	NR
First line	31	<0.01	NR	29	<0.01	NR
Non–first line	18	0.20	NR	20	0.24	NR

HR = hazard ratio; ICI = immune checkpoint inhibitor; NR = not reached; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; STE = surrogate threshold effect.
[a] Longer follow-up indicates trials with follow-up of ≥24 mo.

threatening conditions, such as acquired immunodeficiency syndrome, and later expanded to include cancers. Among several changes, the FDA accepted the use of SEs as the basis for approval of new drugs rather than OS [13]. The validity of this increasing reliance on SEs hinges on a strong correlation between SEs and OS since cancer therapies that improve only PFS or ORR, but do not extend patients' lives, are of questionable clinical benefit [1,2].

Validation attempts of SEs in other malignancies have yielded mixed results, and meta-analyses find that most tumors are characterized by a poor correlation between SEs and OS [3,14]. Nevertheless, of medicines approved on the bases of SEs alone, 61% had insufficient or no prior SE validation, and 16% occur in a setting where validations have found a poor correlation with OS [4]. mUC is an example of the former.

Given this gap in knowledge, our finding of a moderate correlation between HR^{PFS} and HR^{OS} in MUC is a key to reliance on this endpoint in clinical trials. With an R^2 value of 0.60 among all trials, one can be confident that a significant proportion of the variance in OS can be explained by PFS in MUC. Importantly, an STE of 0.41 suggests that the observed HR^{PFS} below this value provides 95% confidence of an expected HR^{OS} of <1.0. Put another way, a study reporting PFS must achieve an HR of ≤0.41 to provide 95% confidence of OS benefit. These findings measure favorably with validated surrogates in other malignancies. For colorectal cancer, Buyse et al. [22] found an R^2 of 0.55 with an STE of 0.77 for HR^{PFS}, which was sufficient to validate PFS in that context. Belin et al [18] reported a methodological systematic review of strategies for PFS surrogacy assessment and argued for an R^2 of ≥0.6 as a threshold for validation. Earlier work specific to bladder cancer has not addressed the STE directly, but instead has evaluated PFS time points as predictors of OS. Using patient-level data from seven chemotherapy trials, Galsky et al [23] reported improved OS for those with PFS >6 and >9 mo (HR 2.49 [95% CI 1.55–3.89] and HR 2.84 [95% CI 1.81, 4.24], respectively). This corroborates our finding of a moderate correlation between PFS with OS. Our results may support the use of PFS as a surrogate for OS in MUC, although a strong PFS benefit (HR ≤0.41) is necessary to provide 95% confidence of the predicted OS benefit.

Examining the difference in median PFS (ΔPFS) compared with ΔOS yielded weaker support for surrogacy compared with HRs. We find an R^2 of 0.12 and an STE of 14.1 mo, a threshold that none of the examined trials achieved. The discrepancy between surrogacy validation for HR^{PFS} and ΔPFS highlights an important methodological issue in surrogacy validation of this type. Many SE validations in the literature performed analyses using ΔPFS and did not analyze HR^{PFS} in the analysis [3,18,24]. Indeed, guidance on SE validation from regulatory agencies is scant and quite vague regarding statistical details. The Institute for Quality and Efficiency in Health Care, for example, highlights the importance of strong correlations and STE, but does not comment on the specific parameter to consider in the regressions, that is, HR or difference of medians [25]. Additionally, HR^{PFS} is not universally reported in trials, even when PFS is the primary endpoint. In this study, of the 53 comparisons reporting PFS, 22 (42%) did not report HR and instead presented only median PFS of each arm. The staggering difference in results between these two methods of analyzing similar data highlights the need for standardization of surrogacy validation prior to their use in drug approval. In mUC, proportional hazard regression should be performed and HR^{PFS} should be reported for trials using PFS as an SE.

The ORR performed poorly as a surrogate for OS in mUC. Regardless of the method of analysis, R^2 values were low and the STE at 95% confidence was not reached. This was reproduced within subsets of trials including trials without crossover, immunotherapy (IO) and non-IO trials, or trials with follow-up of >24 mo. Our findings suggest that ORR alone should not be used as a surrogate for OS, especially when justifying the approval of new therapies.

Several trial design tools influence the relationship between SEs and OS, and thus warranted a separate analysis. Trials with follow-up of longer than 24 mo demonstrated an improved R^2 of 0.76 and a more favorable STE of 0.59 for HR^{PFS}. This finding is intuitive as longer trials allow for accrual of additional mortality events and likely more completely capture differential OS, and thus may bet-

ter reflect the PFS/OS relationship. Importantly, ΔPFS and ORR improved as well among trials with longer follow-up, but did not reach a threshold to make this SE a strong predictor of OS, and the STE was still not reached at 95% confidence for ORR.

Similarly, trials evaluating therapeutics in the first-line setting may have different performance of SEs from those in the non–first-line setting. Since patients evaluated for non–first-line therapeutics have failed prior therapy and are further along in the natural history of their metastatic disease, they will have shorter median survival and thus a smaller interval between capture of SEs and OS. We find that SEs tend to perform modestly better in the second-line setting. Most notably, ΔPFS demonstrated a reasonable STE of 4.67 mo in non–first-line trials, significantly better than that in the first-line space (Table 2). ORR continued to perform poorly regardless of the line of therapeutic being evaluated.

Crossover between the control and treatment arms is another factor that influences the SE/OS correlation. In lung cancer, Hashim et al [24] reported very poor correlation coefficients for ORR and PFS among 146 clinical trials examined (R = 0.18 and R = 0.25, respectively), but significantly improved correlation coefficients among trials where crossover was not allowed (R = 0.53 for ORR and R = 0.78 for PFS). This is also intuitive as control arms that cross over to receive an experimental treatment after the primary SE is captured may receive the benefits of the therapy reflected in their OS measurement, thus biasing any OS difference toward zero, while preserving a strong PFS difference. However, when trials with crossover were excluded from our data, our findings did not change significantly. Still, this highlights two important points. First, SE validation should consider a subanalysis in crossover-restricted trials to avoid underestimating SE/OS correlations. Second, the use of crossover should be avoided in trials investigating the first instance of the use of therapy in a particular disease, as this not only contaminates a subsequent OS analysis, but also potentially delays the access to more established second-line therapies to expose control-arm patients to a still unproven intervention. Conversely, trials investigating therapy *advancement* in a particular disease (ie, second- to first-line therapy) should ideally provide crossover to the control arm upon progression in order to reflect standard-of-care treatment [26].

SEs suffer from important limitations that might explain their poor correlation with survival. Examples include important statistical considerations such as the disproportionate impact of missing data on PFS compared with OS [27,28]. Additionally, as PFS/ORR is determined with the use of cross-sectional imaging, while OS is more obvious to capture, the inherent limitations of imaging confer added challenges on these endpoints. Target lesion identification, measurement, and classification within the Response Evaluation Criteria in Solid Tumors (RECIST) system, for example, are frequent sources of error and can thus contribute to the observed poor reproducibility and high rates of inconsistency in assessments among various trial practitioners and central reviewers [29–31]. Further, the intensity of surveillance imaging and duration of follow-up time can influence PFS/ORR. Despite continued efforts to address many of these challenges [32,33], technical and conceptual problems continue to plague SE reliance and highlight the urgent need to ideally validate SEs prior to their isolated use in clinical decision-making.

This study is limited by the number of trials that have been performed in mUC, and that present both an SE and OS, which is a relatively small sample size. Additionally, our analysis correlates SEs with OS only, and does not consider important potential relationships with health-related or overall quality of life endpoints and patient-reported outcomes. The criteria for PFS/ORR definitions, almost universally the RECIST system, have undergone several versions of modifications and represent an additional source of heterogeneity when comparing trials [32–34]. Finally, some validation strategies utilize patient-level data to estimate the relationships between surrogates and clinical endpoints [35]. Although trial-level analyses are significantly more frequent in the surrogacy literature [3,18], both strategies have a role in answering these important questions [25].

5. Conclusions

mUC trials demonstrating a significant improvement in HR^{PFS} can be expected to represent in improvement in HR^{OS}. However, there is a poor correlation of ΔPFS with ΔOS, and ΔORR or ORR^{ratio} with HR^{OS}; thus, improvements in these surrogates alone should be interpreted with caution and should be de-emphasized in mUC trials. The large variability in the results when comparing ΔPFS/ΔOS and HR^{PFS} and HR^{OR} highlights the need to standardize the validation of surrogacy biomarkers in MUC.

***Author contributions*:** Fady Ghali had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Ghali, Wright, Patel.
Acquisition of data: Ghali, Jewel.
Analysis and interpretation of data: Ghali, Etzioni, Zhao, Gore, Grivas, Wright.
Drafting of the manuscript: Grivas, Gore.
Critical revision of the manuscript for important intellectual content: Yu, Wright, Ghali.
Statistical analysis: Ghali, Etzioni, Zhao.
Obtaining funding: Montgomery, Wright.
Administrative, technical, or material support: Montgomery, Wright.
Supervision: Wright.
Other: None.

***Financial disclosures*:** Fady Ghali certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Petros Grivas has done paid consulting with Aadi Bioscience, AstraZeneca, Astellas Pharma, Boston Gene, Bristol Myers Squibb, Dyania Health, EMD Serono, Exelixis, Fresenius Kabi, Foundation Medicine, Genentech/Roche, Genzyme, GlaxoSmithK-

line, Guardant Health, Gilead Sciences, Infinity Pharmaceuticals, Janssen, Lucence, Merck & Co., Mirati Therapeutics, Pfizer, PureTech, QED Therapeutics, Regeneron Pharmaceuticals, Seattle Genetics, Silverback Therapeutics, UroGen, and 4D Pharma PLC; his institution has received grants from Bavarian Nordic, Bristol Myers Squibb, Clovis Oncology, Debiopharm, EMD Serono, G1 Therapeutics, Gilead Sciences, GlaxoSmithKline, Merck & Co., Mirati Therapeutics, Pfizer, and QED Therapeutics. Evan Yu has received research funding to institution from Daiichi-Sankyo, Taiho, Dendreon, Merck, Seattle Genetics, Blue Earth, Bayer - DAROL and citDNA, and Lantheus; and consulting with honorarium (in the past 3 yr) from Jansen, Merck, Advanced Accelerator Applications, Bayer, Exelixis, Clovis, Abbvie, and Sanofi-Genzyme. Bruce Montgomery has received institutional grants from AstraZeneca, Janssen Oncology, Clovis Oncology, Astellas Pharma, and BeiGene. Jonathan Wright has received institutional grants from Merck & Co., Jannsen, BMS, Altor Biosciences, Nucleix, Pacific Edge, Veracyte, and royalties from UpToDate.

Funding/Support and role of the sponsor: This work was supported by Seattle Translational Tumor Research and MXD Championships via institutional funds.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.euros.2022.11.003.

References

[1] Schnipper LE, Davidson NE, Wollins DS, et al. American Society of Clinical Oncology statement: a conceptual framework to assess the value of cancer treatment options. J Clin Oncol 2015;33:2563–77.
[2] Driscoll JJ, Rixe O. Overall survival: still the gold standard: why overall survival remains the definitive end point in cancer clinical trials. Cancer J 2009;15:401–5.
[3] Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. Int J Clin Oncol 2009;14:102–11.
[4] Chen EY, Haslam A, Prasad V. FDA acceptance of surrogate end points for cancer drug approval: 1992–2019. JAMA Intern Med 2020;180:912–4.
[5] Center for Drug Evaluation and Research. Clinical trial endpoints for the approval of cancer drugs and biologics. U.S. Food and Drug Administration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics.
[6] Chen EY, Joshi SK, Tran A, Prasad V. Estimation of study time reduction using surrogate end points rather than overall survival in oncology clinical trials. JAMA Intern Med 2019;179:642–7.
[7] Center for Drug Evaluation and Research. Surrogate endpoint resources for drug and biologic development. FDA. https://www.fda.gov/drugs/development-resources/surrogate-endpoint-resources-drug-and-biologic-development.
[8] Wilson MK, Karakasis K, Oza AM. Outcomes and endpoints in trials of cancer treatment: the past, present, and future. Lancet Oncol 2015;16:e32–42.
[9] Lebwohl D, Kay A, Berg W, Baladi JF, Zheng J. Progression-free survival: gaining on overall survival as a gold standard and accelerating drug development. Cancer J 2009;15:386–94.
[10] Patel RB, Vaduganathan M, Samman-Tahhan A, et al. Trends in utilization of surrogate endpoints in contemporary cardiovascular clinical trials. Am J Cardiol 2016;117:1845–50.
[11] Fauber J, Chu E. FDA approves cancer drugs without proof they're extending lives. http://www.jsonline.com/watchdog/watchdogreports/fda-approves-cancer-drugs-without-proof-theyre-extending-lives-b99348000z1-280437692.html.
[12] Tannock IF, Pond GR, Booth CM. Biased evaluation in cancer drug trials—how use of progression-free survival as the primary end point can mislead. JAMA Oncol 2022;8:679–80.
[13] Stahl J. A history of accelerated approval: overcoming the FDA's bureaucratic barriers in order to expedite desperately needed drugs to critically ill patients. https://dash.harvard.edu/handle/1/8852155.
[14] Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. JAMA Intern Med 2015;175:1389–98.
[15] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin 2021;71:7–33.
[16] Zhang J, Liang W, Liang H, Wang X, He J. Endpoint surrogacy in oncological randomized controlled trials with immunotherapies: a systematic review of trial-level and arm-level meta-analyses. Ann Transl Med 2019;7:244.
[17] Johnson KR, Liauw W, Lassere MND. Evaluating surrogacy metrics and investigating approval decisions of progression-free survival (PFS) in metastatic renal cell cancer: a systematic review. Ann Oncol 2015;26:485–96.
[18] 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.
[19] Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016;104:240–3.
[20] Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006;5:173–86.
[21] Beaver JA, Pazdur R. "Dangling" accelerated approvals in oncology. N Engl J Med 2021;384:e68.
[22] Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. J Clin Oncol 2007;25:5218–24.
[23] Galsky MD, Krege S, Lin CC, et al. Relationship between 6- and 9-month progression-free survival and overall survival in patients with metastatic urothelial cancer treated with first-line cisplatin-based chemotherapy. Cancer 2013;119:3020–6.
[24] Hashim M, Pfeiffer BM, Bartsch R, Postma M, Heeg B. Do surrogate endpoints better correlate with overall survival in studies that did not allow for crossover or reported balanced postprogression treatments? An application in advanced non-small cell lung cancer. Value Health 2018;21:9–17.
[25] Institute for Quality and Efficiency in Health Care (IQWiG). Validity of surrogate endpoints in oncology. 2011. https://www.ncbi.nlm.nih.gov/books/NBK198799/.
[26] Haslam A, Prasad V. When is crossover desirable in cancer drug trials and when is it problematic? Ann Oncol 2018;29:1079–81.
[27] Korn RL, Crowley JJ. Overview: progression-free survival as an endpoint in clinical trials with solid tumors. Clin Cancer Res 2013;19:2607–12.
[28] Sridhara R, Mandrekar SJ, Dodd LE. Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. Clin Cancer Res 2013;19:2613–20.
[29] Sullivan DC, Schwartz LH, Zhao B. The imaging viewpoint: how imaging affects determination of progression-free survival. Clin Cancer Res 2013;19:2621–8.
[30] Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. J Clin Oncol 2003;21:2574–82.
[31] Thiesse P, Ollivier L, Di Stefano-Louineau D, et al. Response rate accuracy in oncology trials: reasons for interobserver variability. Groupe Français d'Immunothérapie of the Fédération Nationale des Centres de Lutte Contre le Cancer. J Clin Oncol 1997;15:3507–14.
[32] Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228–47.
[33] Schwartz LH, Litière S, de Vries E, et al. RECIST 1.1-Update and clarification: from the RECIST committee. Eur J Cancer 2016;62:132–7.
[34] 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.
[35] Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8:431–40.