European Journal of Cancer 223 (2025) 115513

Contents lists available at ScienceDirect

European Journal of Cancer

journal homepage: www.ejcancer.com

Review

Radiological progression-free survival as a surrogate for overall survival in patients with metastatic hormone-sensitive prostate cancer: A bivariate meta-analysis

Neal Shore [a,*], Alicia K. Morgans [b], Martin Boegemann [c], Elaine Gallagher [d], Noman Paracha [d], Paul Serafini [e], Divya Pushkarna [e], Mir-Masoud Pourrahmat [e], Murat Kurt [e], Keith R. Abrams [f]

[a] START Carolinas/Carolina Urologic Research Center, Myrtle Beach, SC, USA
[b] Dana-Farber Cancer Institute, USA
[c] University Hospital of Münster, Department of Urology, Germany
[d] Bayer Pharmaceuticals, Basel, Switzerland
[e] Evidinno Outcomes Research Inc., Vancouver, BC, Canada
[f] Department of Statistics & Warwick Medical School, University of Warwick, Coventry, UK

ARTICLE INFO

Keywords:
Surrogate endpoint
Radiological progression-free survival
Overall survival
Metastatic prostate cancer

ABSTRACT

Background: Overall survival (OS) is the standard efficacy endpoint in various solid tumor trials; however, it requires longer follow-up time for assessment than potential intermediate endpoints. This study evaluated radiological progression-free survival (rPFS) as a surrogate for OS in metastatic hormone-sensitive prostate cancer (mHSPC) using aggregate-level data from randomized controlled trials (RCTs).

Methods: A systematic literature review identified mHSPC RCTs published through December 2023, reporting hazard ratios for rPFS (HR_{rPFS}) and OS (HR_{OS}). Correlation between HR_{rPFS} and HR_{OS} was assessed using bivariate random-effects meta-analysis (BRMA). Predictive validity was assessed with leave-one-out cross-validation (LOOCV). The surrogate threshold effect (STE), or minimum rPFS benefit predicting an OS benefit, was estimated using recent mHSPC trial sample sizes. Sensitivity analyses (1) omitted trials that had only one of the endpoints reported, (2) omitted HRs that violated proportional hazards assumptions, (3) omitted trials that allowed cross-over and (4) investigated different assumed values of the within-study correlation.

Results: The primary analysis included 35 treatment comparisons from 31 trials. The estimated rPFS-OS correlation was 0.95 (95 % CrI: 0.75, 1.00). LOOCV confirmed HR_{OS} were within 95 % prediction intervals. The estimated STE ranged from 0.55 to 0.71 depending on the trial size being predicted. Sensitivity analyses produced strong but slightly lower correlations (0.87, 0.89, 0.91) than the primary analysis, with full coverage of the reported HR_{OS} in cross validation. Increasing within-study correlation slightly reduced between-study correlation.

Conclusions: The derived surrogacy equation enables OS estimation based on reported rPFS benefits in mHSPC, meeting NICE's 95 % surrogate validity threshold. These findings support rPFS as a reliable surrogate for OS, facilitating prediction of OS benefits in future mHSPC trials.

1. Introduction

Prostate cancer ranks fourth in incidence and eighth in mortality among cancers globally [1]. Metastatic hormone-sensitive prostate cancer (mHSPC), prostate cancer which has spread beyond the prostate and is sensitive to androgen deprivation therapy (ADT), is invariably lethal, presenting as de novo or recurrent, with de novo cases having poor prognosis [2]. ADT has been the standard of care for treating mHSPC, however, recent phase 3 studies have shown that combining ADT with docetaxel [3], or androgen receptor pathway inhibitor (ARPI) [4–7] leads to improved outcomes. Recent phase 3 trials also support triplet therapy, such as abiraterone [8] or darolutamide [9] with

* Correspondence to: Carolina Urologic Research Center, 823 82nd Parkway, Myrtle Beach, SC, USA.
E-mail address: nshore@auclinics.com (N. Shore).

https://doi.org/10.1016/j.ejca.2025.115513
Received 21 January 2025; Received in revised form 7 May 2025; Accepted 10 May 2025
Available online 16 May 2025

docetaxel and ADT, and thus doublet and triplet therapies for mHSPC are now incorporated in all major guidelines [10].

Overall survival (OS) is the gold standard measure of efficacy of many, but not all[a] new anticancer drugs in various solid tumor trials, and the approval of these drugs therefore relies on the demonstration of clinical or statistically significant OS benefit. However, the growing availability of effective life prolonging treatments for mHSPC has extended the time needed for trials to measure OS outcomes [11], which can lead to delayed drug approvals, restricted patient access, increased trial costs, and prolonged negotiation times for reimbursement and pricing between the agencies and manufacturers. Consequently, intermediate endpoints like radiological progression-free survival (rPFS), complete response (CR), objective response rate (ORR), and prostate-specific antigen (PSA) response are used as secondary trial endpoints and guide real-world therapeutic decisions as they reach statistical maturity sooner with lower sample sizes [12–16]. The importance of surrogate endpoints to regulatory and HTA bodies is reflected in their presence in decision making. In the regulatory space, the US Food and Drug Administration (FDA) made 342 drug approvals in non-hematologic solid cancers between 2006 and 2022, of which 71.3 % were based on surrogate endpoints [17]. The European Medicines Agency (EMA) issued 125 marketing authorization decisions in oncology between 2009 and 2017, in which 65.6 % of the supporting trials did not use OS as a primary endpoint [18]. In the HTA space, a review of technology appraisals by NICE found that of 47 appraisals, 18 (38 %) used surrogate endpoints [19].

Methodologies have been developed to validate these intermediate endpoints as surrogate endpoints from which OS or OS benefit can be predicted. Prentice (1989) introduced three criteria for assessing a surrogate endpoint: individual patient association, trial level association, and biological plausibility [20]. Individual patient association assesses the prognostic relationship between the surrogate and OS, and therefore does not require contrast data. On the other hand, trial level association investigates the association between aggregate-level treatment effects and therefore requires the study of treatment contrasts. Biological plausibility evaluates the existence of a causal mechanism linking the surrogate endpoint to the true endpoint. A recent white paper on surrogacy, developed by a working group of health technology assessment (HTA) bodies, highlights that when addressing the strength of association, most HTA bodies' surrogacy guidelines typicallyreference these three distinct levels of evidence [21]. Additionally, the National Institute of Health and Care Excellence (NICE) has developed and recommended a bivariate random effects meta-analysis (BRMA) framework for addressing the trial-level association [22].

A recent study by Halabi et al. (2024) [23] assessed rPFS as a surrogate for OS in mHSPC using individual patient data (IPD). They estimated a strong correlation between rPFS and OS at both the individual patient level and the treatment effect level and concluded that rPFS is a promising surrogate for OS in mHSPC. However, a limitation of the analysis was that it only focused on studies with ADT as the backbone therapy, most of which concluded enrolment before 2010, a fact which prompted a letter to the editor asking whether it is premature to accept rPFS as a surrogate for OS [24]. Additionally, Gharzai et al. (2023) explored surrogate endpoints for OS in advanced prostate cancer using a broader evidence base, which included a subgroup analysis for rPFS in mHSPC [25]. They did not meet their pre-specified threshold for surrogacy. However, since rPFS and mHSPC were not the primary objective of Gharzai et al.'s study, key parameters such as the prediction equation were not reported for this subgroup. Therefore, there isa need for further comprehensive validation of rPFS as a surrogate for OS in mHSPC, incorporating a wider evidence base that includes more recent studies on both ADT and ARPIs.

The purpose of this study was to evaluate rPFS as a surrogate endpoint for OS in patients with mHSPC, through BRMA, utilizing aggregate-level data from RCTs in concordance with NICE TSD DSU 20 guidelines [22]. Secondary objectives were to assess the robustness of the correlation with respect to parameter uncertainty, missing data and key structural assumptions made for the analysis.

2. Materials and methods

2.1. Systematic literature review (SLR)

The SLR was conducted following the Cochrane guidelines [26]. Adhering to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [27], RCTs assessing the efficacy, safety and tolerability of approved and upcoming treatments of mHSPC were identified by searching MEDLINE®, Embase®, CENTRAL and the Cochrane Database of Systematic Reviews (CDSR) from database inception to December 12, 2023 using predefined search strategies. Grey literature and hand searches of key clinical conferences (2020–2023) such as American Society of Clinical Oncology (ASCO), American Society of Clinical Oncology – Genitourinary Cancer (ASCO-GU), European Society for Medical Oncology (ESMO), American Urologic Association (AUA) and European Association of Urology (EAU) were conducted. Searches of reference lists of previously published literature reviews on similar topics were also conducted.

Study selection and data extraction were performed by two independent reviewers. If hazard ratios (HRs) were not reported but Kaplan-Meier (KM) curves were available, the KM curves were digitized and pseudo IPD was generated using the Guyot algorithm, then a Cox proportional hazards (PH) model was used to calculate HRs [28]. For trials that did not publish either HRs or KM curves for one or both of rPFS and OS, where possible HRs for both endpoints were taken from Halabi et al. (2024) [23], which reported unpublished HRs calculated from IPD.

2.2. Definition of end points

OS is defined as the interval from the date of randomization to the date of death from any cause. rPFS is defined as the time from randomization to radiographic progression or death from any cause, whichever occurs first [23], as distinct from clinical PFS. Patients are censored at the date of last follow-up if they did not experience an event in the observation period. However, the index date and event definitions may vary between studies.

2.3. Statistical methods

The association between the natural log transformed HRs of rPFS and OS was assessed using BRMA as per the NICE TSD DSU 20 [22]. A standard BRMA model was fitted as the base case, and the product normal formulation (PNF) was fitted as a secondary analysis, to assess the sensitivity of the results under different parameterizations of a BRMA model. Trials were included if they reported an HR for one or both of rPFS (HR_{rPFS}) and OS (HR_{OS}) for one or more independent treatment comparisons. We assigned the heterogeneity parameters τ_1 and τ_2 priors of Unif(0, 2), the between-study correlation ρ a prior of Unif(-0.999, 0.999), and the within-study correlation ρ_{wi} a prior of Unif (0, 0.999).

Predictive validity was assessed using leave-one-out cross validation (LOOCV) by fitting a new model per study reporting both rPFS and OS and omitting OS data for that study. The 95 % prediction interval (PIs) generated from the leave-one-out models were then evaluated to determine whether they captured the omitted OS HRs at least 95 % of the time [22]. The strength of association was classified based on the lower limit of the 95 % Credible Interval (CrI) for ρ^2 (ρ^2_{LL}) per BSES3 criteria [29]. Additionally, the surrogate threshold effect (STE) was estimated for studies with $\ln HR_{OS}$ variances of 0.01 and 0.03 (the first and third quartiles from the evidence base), defined as the minimum rPFS benefit required to predict a statistically significant OS benefit (with 95 % confidence interval [CI] excluding 1).

In addition to the primary analysis, four sensitivity analyses were conducted: 1) including studies that only reported both HR_{rPFS} and HR_{OS}, to assess the impact of excluding studies with missing outcome data; 2) including studies in which HRs met the PH assumption. If the PH assumption was met by one endpoint but not the other for a given treatment comparison, the former was retained, and the latter was excluded; 3) including studies with no treatment crossovers; 4) by varying within-study correlation at 0.05, 0.2, 0.4, 0.6, 0.8, and 0.95 to evaluate the impact of correlation choice on the results, using the same evidence base as the primary analysis.

Lastly, we conducted a two-fold comparison of our results with previous meta-analysis by Halabi et al. (2024) [23]. First, we compared the original weighted linear regression (WLR) results from Halabi et al. and second, we compared a BRMA using the same evidence base to our analysis, to account for any methodological differences.

The details on model definitions, model fitting and data selection process are provided in the Supplementary Methods.

3. Results

A total of 43 trials were identified, 39 trials from the SLR and 4 hand-searched trials (CALGB 90202, MRC PR05, SWOG S9346, and ZAPCA)

Table 1
Summary of RCTs identified through SLR.

Trial	Phase	Blinding	Crossover	Region	N Randomized	Median Follow-Up (Months)	Experimental Arm Class	Primary Endpoint(s)
ARANOTE	III	Double blind	Planned #	Intercontinental	669	25, 25.3	ADT + ARPI	rPFS
ARASENS	III	Double blind	No	Intercontinental	1306	42.4, 43.7	ADT + ARPI + CT	OS
ARCHES	III	Double blind	DSMB	Intercontinental	1150	44.6	ADT + ARPI	rPFS
Ayoub 2023	III	N/R	No	Egypt	63	12	ADT + RT	HRQoL
BonEnza	II	Open label	N/R	Europe	126	N/R	ADT + ARPI + Other	Bone response
Bruun 1996	III	Open label	No	Denmark	149	N/R	ADT	N/R
CALGB 90202	III	Open label	N/R	North America	645	38.4	ADT + Other	SRE
CHAARTED	III	Open label	No	United States	790	54	ADT + CT	OS
Chang 1996	N/R	Double blind	No	N/R	92	59	ARPI	N/R
CHART	III	Open label	No	N/R	654	20.4,22.1	ADT + ARPI	rPFS, OS
China ARCHES	III	Double blind	Post-Radiographic Progression #	China	180	N/R	ADT + ARPI	PSA-PFS
Chodak 1995	N/R	Open label	No	United States	516	17	ARPI	TTP, OS, TTF
ENZAMET	III	Open label	Post-mCRPC Progression	Intercontinental	1125	68	ADT + ARPI + /- CT	OS
EORTC−30892	III	Open label	No	Europe	310	103.2	ARPI	OS
GETUG-AFU 15	III	Open label	Post-mCRPC Progression	Europe	385	83.9	ADT + CT	OS
Iversen 1996	III	Open label	No	Europe	376	15.4, 17.3	ARPI	TTP, OS, TTF
Kaisary 1995	N/R	Open label	No	Europe	304	14, 15.2	ARPI	TTP, OS, TTF
KEYNOTE−991	III	Double blind	N/R	Intercontinental	1251	21.1	ADT + ARPI + IO	rPFS, OS
KYUCOG−1401	N/R	N/R	No	N/R	200	N/R	ADT	PSA-PFS
LATITUDE	III	Double blind	DSMB #	Intercontinental	1199	30.4	ADT + ARPI + Other	rPFS, OS
MRC PR05	III	Double blind	N/R	Intercontinental	311	79.2	ADT + Other	Symptomatic BPFS
PEACE−1	III	Open label	Post-mCRPC Progression	Europe	1173	42.7, 52.7	ADT + ARPI + /- CT + /- RT	rPFS, OS
PROSTRATEGY	II/III	Open label	No	Spain	150	N/R	ADT + CT + IO	OS
STAMPEDE A-B	II/III	Open label	No	Europe	1090	60.0	(ADT or ARPI) + Other	OS
STAMPEDE A-C					1086	78.0	(ADT or ARPI) + CT	
STAMPEDE A-D					565	81.6	(ADT or ARPI) + Other	
STAMPEDE A-E					1090	76.8	(ADT or ARPI) + CT + Other	
STAMPEDE A-F					567	81.6	(ADT or ARPI) + Other	
SWOG S1216	III	Open label	No	United States	1313	58.8	ADT + Other	OS
SWOG S9346	III	Open label	N/R	Canada	1535	117.6	ADT	N/R
TITAN	III	Double blind	DSMB #	Intercontinental	1052	24 *	ADT + ARPI	rPFS, OS
Vaishampayan 2021	II	Open label	No	United States	71	39	ADT + ARPI	PSA response
Vogelzang 1995	III	Open label	No	United States	283	48	ADT	N/R
Zalcberg 1996	N/R	Double blind	No	Australia	222	N/R	ADT + ARPI	N/R
ZAPCA	III	Open label	N/R	Japan	227	46.8	ADT + ARPI	TTF

Abbreviations: * – Milestone rather than median follow-up, # – Crossover did not impact the surrogacy analysis because pre-crossover data were used, ADT – Androgen Deprivation Therapy, ARPI – Androgen Receptor Pathway Inhibitor, BPFS – Bone Progression-Free Survival, CT – Chemotherapy, DSMB – Recommended by a Data and Safety Monitoring Board, HRQoL – Health-Related Quality of Life, IO – Immunotherapy, N – Sample size, N/R – Not Reported, PSA – Prostate-Specific Androgen, RT – Radiotherapy, SLR – Systematic Literature Review, SRE – Skeletal-Related Events, TTF – Time to Treatment Failure, TTP – Time to Progression.
Note: STAMPEDE A-B, A-C, A-D, A-E, and A-F are independent pseudo trials generated by Halabi et al. (2024) [23] by randomly splitting the control arm between the treatment comparisons. "Registry #" refers to the ClinicalTrials.gov registry (NCT number).

(Supplementary Figure). Among these, 31 reported an HR for one or both of rPFS and OS and therefore contributed to the surrogacy analysis. Data for STAMPEDE trial were taken from Halabi et al. (2024) [23], which divided STAMPEDE into five pseudo-trials, splitting the control arm across the treatment comparisons. Thus, in total the evidence base for the primary analysis consisted of 35 treatment comparisons from 31 trials comprising a total of 18,900 patients, described in Table 1**.** There were 20 studies for which both rPFS and OS HRs were available and were included in the first sensitivity analysis, along with 13 studies that reported OS HRs and 2 studies with only rPFS HRs. The second sensitivity analysis included 25 studies where at least one HR passed the PH assumption test. The third sensitivity analysis included 27 studies without treatment cross-over. The fourth sensitivity analysis included the same evidence base as the primary analysis and varied within-study correlations. The input data for the included trials, along with PH test results for each HR, are provided in the Supplementary Table.

The BRMA estimated a between-study correlation of 0.945 (95 % CrI: 0.747, 0.996), corresponding to an R^2 of 89.4 % (95 % CrI: 55.8 %, 99.2 %), meeting the BSES3 criteria for "good" correlation (Table 2).

The estimated surrogacy equation was $\ln HR_{OS} = -0.006 + 0.556 \times \ln HR_{rPFS}$ with a near-zero intercept and statistically significant slope (Fig. 1). A scatterplot color coding ARPI/immunotherapy studies and primary ADT studies suggested that the relationship did not differ substantially by drug class (**Supplementary Results**). For variances of the lnHR of OS ranging from 0.01 to 0.03 (standard error [SE]: 0.1–0.173), STEs ranged from 0.711 to 0.550. In other words, for a study with $\ln HR_{OS}$ variance of 0.01 (SE: 0.1), HR_{rPFS} of approximately 0.711 or less predicts a statistically significant HR_{OS}. In LOOCV (Fig. 2), every trial's reported HR_{OS} was captured by the 95 % PI around the model's predicted HR_{OS}, meeting NICE's 95 % threshold for validity and indicating that HR_{rPFS} accurately predicts HR_{OS} for new studies.

The median between-study correlation was slightly lower (0.873, 0.885, 0.914) when the evidence base was restricted to 20 studies that reported information on both endpoints, 25 studies reporting endpoint data that met the PH assumption and 27 studies without treatment cross-over (Table 2). When sensitivity analyses were conducted specifying a range of possible values of the within-study correlation, the median between-study correlation was stable with overlapping CrIs although at the highest levels (0.8 and 0.95) the median between-study correlation was lower (Table 3). Scatterplots and LOOCV plots for the sensitivity analyses are provided in the **Supplementary Results**.

Results using the PNF model were nearly identical to the standard BRMA model in the primary and sensitivity analyses (**Supplementary Results)**.

Table 2
Summary estimates of key model outcomes/parameters for the primary analysis and sensitivity analyses (1−3).

Parameter	Primary Analysis	Sensitivity Analysis 1	Sensitivity Analysis 2	Sensitivity Analysis 3
	Median (95 % CrI)	**Median (95 % CrI)**	**Median (95 % CrI)**	**Median (95 % CrI)**
ρ (between-study correlation)	0.95 (0.75, 1.00)	0.89 (0.55, 0.99)	0.87 (0.40, 0.99)	0.97 (0.84, 1.00)
λ_0 (intercept of the surrogacy equation)	−0.01 (−0.08, 0.08)	−0.05 (−0.15, 0.04)	−0.06 (−0.17, 0.04)	−0.01 (−0.07, 0.07)
λ_1 (slope of the surrogacy equation)	0.56 (0.38, 0.75)	0.48 (0.25, 0.70)	0.46 (0.17, 0.72)	0.57 (0.41, 0.74)
ψ_2^2 (conditional variance)	0.01 (0.00, 0.03)	0.01 (0.00, 0.02)	0.01 (0.001, 0.027)	0.00 (0.00, 0.02)

Abbreviations: CrI – Credible Interval, HR – Hazard Ratio, OS – Overall Survival, rPFS – Radiographic Progression-Free Survival.
Sensitivity Analysis #1: Studies with Complete Data; Sensitivity Analysis #2: Trial with rPFS and OS data which met the PH Assumption #3 Studies without treatment cross-over
Note: R^2 can be derived by squaring ρ.

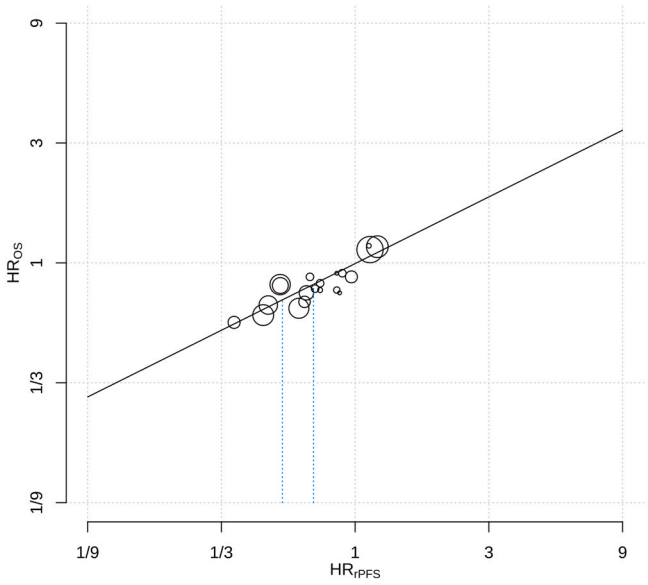

Fig. 1. Scatterplot of the estimated surrogacy equation from the standard BRMA model for the primary analysis. Note: Each circle represents the HR of rPFS and HR of OS for a single trial. The size of the circle is scaled to the sample size of the trial. The solid line represents the estimated surrogacy equation for predicting HR_{OS} from HR_{rPFS}. The x- and y-axes are on the logarithmic scale. The estimated surrogate threshold effects are represented by blue, vertical dashed lines corresponding to $\ln HR_{OS}$ variances of 0.01 and 0.03. Abbreviations: BRMA – Bivariate Random Effects Meta-Analysis, HR – Hazard Ratio, OS – Overall Survival, rPFS – Radiographic Progression-Free Survival.

Abbreviations: CrI – Credible Interval, HR – Hazard Ratio, OS – Overall Survival, rPFS – Radiographic Progression-Free Survival.

The original WLR results from Halabi et al. (2024) [23], and the BRMA estimate was generated for this publication to enhance the comparability of the results between the two evidence bases. The detailed comparison is presented in Table 4.

4. Discussion

This BRMA evaluated rPFS as a surrogate for OS in mHSPC. The point estimate of the between-study correlation demonstrated "good" correlation as per BSES3 in the primary analysis. The predictive accuracy in LOOCV was 100 %, meeting NICE's 95 % threshold for surrogate validity. The STE demonstrated that a significant OS benefit can be predicted from rPFS benefit in a large trial. Sensitivity analyses, restricted to studies that reported both rPFS and OS, HRs that met the PH assumption, and studies with no treatment cross-over, resulted in only slightly lower correlations than the primary analysis. Further, analysis with varying within-study correlations showed a slight inverse relationship between within-study and between-study correlations, with consistency maintained for both slope and intercept. Overall, these suggest a strong association and potential for consideration of rPFS as a valid surrogate for OS in this setting.

A previous meta-analysis by Halabi et al. (2024) [23] also assessed rPFS as a surrogate for OS in mHSPC, using a frequentist WLR approach, as opposed to Bayesian BRMA method used in this analysis. The present analysis included all but two studies from Halabi et al. (NTR 130 [30] and HOG GU-0421 [31] were excluded because the population was unclear, and the population was castration-resistant prostate cancer, respectively), and included an additional nine studies reporting both rPFS and OS (ARANOTE, ARCHES, CHART, ENZAMET, KEYNOTE-991, KYUCOG-1401, LATITUDE, PEACE-1, and TITAN) and fifteen RCTs that reported just one of rPFS and OS.

Compared to the Halabi et al. analysis using WLR, our BRMA using

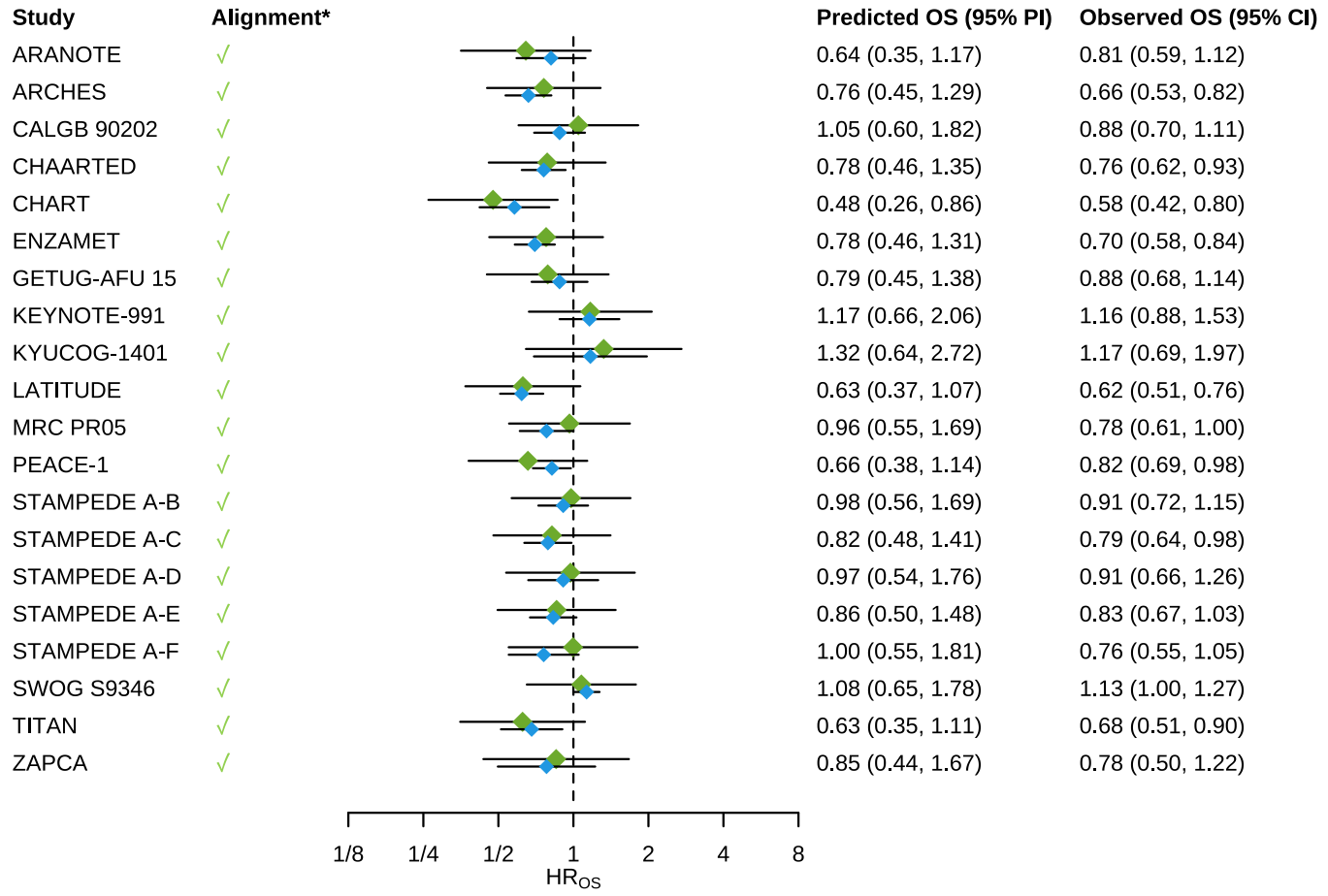

Fig. 2. Forest plot comparing the model predictions from the standard BRMA model to the observed data during LOOCV for the primary analysis. Trials are ordered alphabetically. Note: Blue diamonds and their error bars represent the reported HR_{OS}'s from the trials and their 95 % CIs, respectively, whereas the green diamonds and their error bars are the predicted HR_{OS}'s and their 95 % PIs. Checkmarks next to trial names indicate alignment between the 95 % prediction interval and reported HR_{OS}. HRs in the "Observed" column may not match the associated publications due to the imprecision and resulting asymmetry of published CIs. Abbreviations: * – Alignment between predicted and observed OS; a green checkmark indicates that the OS HR was captured by the 95 % PI of the predicted OS HR, BRMA – Bivariate Random Effects Meta-Analysis, CI – Confidence Interval, HR – Hazard Ratio, OS – Overall Survival, PH – Proportional Hazards, PI – Prediction Interval, rPFS – Radiographic Progression-Free Survival.

Table 3
Summary estimates of key model outcomes/parameters for the sensitivity analyses with varied within-study correlation.

Within-Study Correlation **Parameter**	0.05 **Median (95 % CrI)**	0.2 **Median (95 % CrI)**	0.4 **Median (95 % CrI)**	0.6 **Median (95 % CrI)**	0.8 **Median (95 % CrI)**	0.95 **Median (95 % CrI)**
ρ (between-study correlation)	0.97 (0.75, 1.00)	0.97 (0.77, 1.00)	0.96 (0.77, 1.00)	0.96 (0.76, 1.00)	0.94 (0.74, 1.00)	0.93 (0.73, 0.98)
λ_0 (intercept of the surrogacy equation)	0.02 (−0.07, 0.11)	0.01 (−0.07, 0.11)	0.01 (−0.07, 0.09)	−0.00 (−0.08, 0.08)	−0.01 (−0.08, 0.07)	−0.02 (−0.09, 0.06)
λ_1 (slope of the surrogacy equation)	0.59 (0.38, 0.81)	0.59 (0.39, 0.81)	0.58 (0.39, 0.77)	0.57 (0.38, 0.74)	0.55 (0.38, 0.72)	0.53 (0.37, 0.70)

an expanded evidence base found a similar correlation. However, although the intercept of the surrogacy equation estimated in our BRMA was similar to Halabi et al., its slope was shallower, leading to more conservative predictions: a greater rPFS benefit was required to predict a significant OS benefit in our analysis. When we conducted a BRMA using the restricted evidence base from Halabi et al. to assess the impact of methodology on the results, compared to their original WLR analysis we found a weaker correlation and a shallower slope. This suggests that the BRMA methodology is more conservative than WLR, and this may explain the shallower slope in our primary BRMA using an expanded evidence base.

The earlier surrogacy analysis conducted by Gharzai et al. (2023) also used WLR to estimate an R^2 of 0.54 (95 % CI: 0.15–0.80) for rPFS in mHSPC [25]. This estimate is weaker than those reported by Halabi et al.'s ($R^2 = 0.83$) and our own analysis ($R^2 = 0.894$). However, Gharzai et al.'s estimate is only marginally excluded from the upper bound of the 95 % CrI on R^2 our analysis (0.558, 0.992). This discrepancy may be attributed to study inclusion, as Gharzai et al.'s methodology was similar to Halabi et al., however further explanation is difficult, as Gharzai et al did not report which studies were included or how many studies were part of their subgroup analysis of rPFS in mHSPC.

A key strength of this surrogacy analysis compared to the previous analysis by Halabi et al. (2024) is the use of the Bayesian BRMA method advocated by NICE. Unlike the WLR approach, BRMA accounts for all relevant uncertainty (both within- and between-study), by incorporating the uncertainty (in terms of the variances) from both endpoints into the analysis and including studies that only report one of the two endpoints. Additionally, this BRMA analysis enhances the findings of Halabi et al.

Table 4
Comparison of primary analysis results to Halabi et al. (2024).

Outcome	Halabi et al. (2024) Original WLR Results	Halabi et al. (2024) BRMA Results *	Current Study BRMA Results
	Median (95 % CI)	**Median (95 % CI)**	**Median (95 % CI)**
R^2	0.83 (0.64, 0.98)	0.680 (0.036, 0.977)	0.894 (0.558, 0.992)
Intercept	0 * *	−0.039 (−0.142, 0.055)	−0.006 (−0.080, 0.079)
Slope	0.781 * *	0.655 (0.073, 1.148)	0.556 (0.378, 0.747)
STE	0.80 †	0.635–0.788 * * *, †	0.550–0.711 * * *, †

Abbreviations: * – BRMA results derived from the same evidence base used in Halabi et al. (2024) [23]; * * – Digitized from figure; CIs not reported. * * * – A range of STEs based on assumed variances around HR_{OS}, CI – Confidence Interval, CrI – Credible Interval, STE – Surrogate Threshold Effect, WLR – Weighted Linear Regression, † – Because STE represents the threshold rPFS benefit that would translate into statistically significant OS benefit at a given confidence level, 95 % CIs are often not generated for them.

(2024) by demonstrating stable results across different methodological approaches. It further strengthens the analysis by incorporating both ADT and ARPI therapies from all published studies.

However, this study is subject to several limitations. First, due to the inability to estimate within-study correlation from IPD, a positive correlation between 0 and 1 was assumed. Nevertheless, a positive within-study correlation between HR_{rPFS} and HR_{OS} is suggested, as both individual-level and trial-level correlation from Halabi et al. were positive [23]. This assumption is conservative as it does not infer the absolute magnitude of correlation. Second, the lack of access to IPD prevented the assessment of individual-level correlation, which along with the aggregate-level surrogacy analysis would fulfill two of the three criteria for surrogacy. However, conducting both analyses simultaneously is not necessary, as the individual-level correlation was recently investigated by Halabi et al. [23], though their analysis did not include all published studies. Third, although the predictive validity in LOOCV was 100 %, some point estimates from the models were somewhat skewed. For instance, the predicted HR_{OS} for the ARANOTE trial (0.644) was more favorable than the currently observed value (0.813). This discrepancy may be attributed to the relative immaturity of OS data for ARANOTE, as its follow-up period is shorter than that of other studies. However, this fact highlights a benefit of this analysis, which is that it can help predict long-term OS in studies where it may be biased due to crossover and predict significance in studies that are underpowered for OS. Fourth, intermediate OS for TITAN and LATTITUDE were used in this analysis instead of final OS in order to use rPFS and OS from the same cut-off date. Lastly, the heterogeneity of the included trials could have influenced the results. For example, the surrogacy relation could differ based on tumor volume, drug class, or local reimbursement patterns for subsequent treatments. However, Halabi et al. (2024) conducted subgroup analyses stratified by disease volume, finding an R^2 of 0.87 (95 % CI: 0.06, 1.00) among patients with high-volume disease and 0.85 (95 % CI: 0.16, 1.00) among patients with low-volume disease. These results indicate that the surrogacy relationship between rPFS and OS is not significantly affected by tumor volume. Additionally, although we did not conduct subgroup analyses by drug class, highlighting primary ADT vs ARPI/immunotherapy studies in a scatterplot showed a similar relationship between HRs of rPFS and OS. Nonetheless, further validation of the effects of tumor volume, drug class, and regional reimbursement patterns for subsequent treatments is recommended.

5. Conclusions

The derived surrogacy equation allows for the estimation of OS benefits based on the reported rPFS benefits in mHSPC, meeting NICE's 95 % surrogate validity threshold. Together with Halabi et al.'s validation of the individual level association between rPFS and OS, these findings support the use of rPFS as a reliable surrogate for OS, helping to predict OS benefits earlier in future trials for mHSPC. Clinically this could enable more timely treatment decisions by enabling accelerated approval and early market access in advance of mature OS data, and aid in the design of future clinical trials.

CRediT authorship contribution statement

Alicia K Morgans: Writing – review & editing. **Martin Boegemann:** Writing – review & editing. **Elaine Gallagher:** Writing – review & editing, Conceptualization. **Noman Paracha:** Writing – review & editing, Conceptualization. **Paul Serafini:** Writing – review & editing, Formal analysis, Data curation. **Divya Pushkarna:** Writing – review & editing, Methodology, Data curation. **Pourrahmat Masoud:** Writing – review & editing, Methodology. **Murat Kurt:** Writing – review & editing, Methodology, Conceptualization. **Abrams Keith:** Writing – review & editing, Methodology. **Neal Shore:** Writing – review & editing, Conceptualization.

Authors' contributions

E.G., N.P., P.S., M.K., contributed to study conceptualization and design. D.P. and M.P. were responsible for the development of the methodology and data acquisition, and P.S. performed the statistical analysis. All authors were involved in interpretation of data and were responsible for the writing, review, and/or revisions of the article. All authors read and approved the final article.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This study was funded by Bayer Pharmaceuticals.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: NS has received honoraria from Bayer, Janssen Scientific Affairs, Dendreon, Tolmar, Ferring, Medivation/Astellas, Amgen, Pfizer, AstraZeneca, Genentech, Myovant Sciences. AM has received honoraria from Astellas, AstraZeneca, AAA, Bayer, Exelixis, Lantheus, Novartis, Pfizer, Myovant, Myriad, Merck, Sanofi, Telix. MB has received honoraria from Astellas, AstraZeneca, AAA, Amgen, Bayer, Exelixis, Novartis, Pfizer, Lilly, MSD, BMS, Merck, Janssen, Gilead, EUSApham, Eisai. KRA is an ad hoc paid consultant to Bayer, as well as providing unrelated methodological and strategic advice to the pharmaceutical and life sciences industry. He has received unrelated research funding from Bayer, Association of the British Pharmaceutical Industry (ABPI), European Federation of Pharmaceutical Industries & Associations (EFPIA), Pfizer, Sanofi and Swiss Precision Diagnostics. He is a Partner and Director of Visible Analytics Limited. He is also a member of the NICE Diagnostics Advisory Committee, the NICE Decision and Technical Support Units, and is a National Institute for Health Research (NIHR) senior investigator emeritus. E.G. and N.P. report employment with Bayer. P.S., D.P. and M.P. report employment, and M.K. reports contractual relationship with Evidinno Outcomes Research Inc., which was commissioned by Bayer to conduct this study. Authors report no other conflicts of interest.

Acknowledgements

We thank Nishu Gaind from Evidinno Outcomes Research Inc. for her support in the digitization of KM curves from the publications. The authors received medical writing support for the preparation of this manuscript provided by Evidinno Outcomes Research Inc. (Vancouver, BC, Canada), funded by Bayer Pharmaceuticals.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejca.2025.115513.

Data Availability

Not applicable. The data for this literature review was retrieved from published studies listed in the manuscript.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J Clin 2021;71:209–49.
[2] Yudi MB, Clark DJ, Tsang D, Jelinek M, Kalten K, Joshi S, et al. SMARTphone-based, early cardiac REHABilitation in patients with acute coronary syndromes: a randomized controlled trial. Coron Artery Dis 2020;27.
[3] Sweeney CJ, Chen Y-H, Carducci M, Liu G, Jarrard DF, Eisenberger M, et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer. N Engl J Med 2015;373:737–46.
[4] James ND, de Bono JS, Spears MR, Clarke NW, Mason MD, Dearnaley DP, et al. Abiraterone for prostate cancer not previously treated with hormone therapy. N Engl J Med 2017;377:338–51.
[5] Fizazi K, Tran N, Fein L, Matsubara N, Rodriguez-Antolin A, Alekseev BY, et al. Abiraterone plus prednisone in metastatic, castration-sensitive prostate cancer. N Engl J Med 2017;377:352–60.
[6] Stenzl A, Dunshee C, De Giorgi U, Alekseev B, Iguchi T, Szmulewitz RZ, et al. Effect of enzalutamide plus androgen deprivation therapy on health-related quality of life in patients with metastatic hormone-sensitive prostate cancer: an analysis of the ARCHES randomised, placebo-controlled, phase 3 study. Eur Urol 2020;78:603–14.
[7] Chi KN, Agarwal N, Bjartell A, Chung BH, Pereira de Santana Gomes AJ, Given R, et al. Apalutamide for metastatic, castration-sensitive prostate cancer. N Engl J Med 2019;381:13–24.
[8] Fizazi K, Foulon S, Carles J, Roubaud G, McDermott R, Fléchon A, et al. Abiraterone plus prednisone added to androgen deprivation therapy and docetaxel in de novo metastatic castration-sensitive prostate cancer (PEACE-1): a multicentre, open-label, randomised, phase 3 study with a 2×2 factorial design. Lancet 2022;399:1695–707.
[9] Smith MR, Hussain M, Saad F, Fizazi K, Sternberg CN, Crawford ED, et al. Darolutamide and survival in metastatic, hormone-sensitive prostate cancer. N Engl J Med 2022;386:1132–42.
[10] Lowrance W., Dreicer R., Jarrard D. Advanced Prostate Cancer: AUA/SUO Guideline Using AUA Guidelines.
[11] Fiteni F, Westeel V, Bonnetain F. Surrogate endpoints for overall survival in lung cancer trials: a review. Expert Rev Anticancer Ther 2017;17:447–54.
[12] Cooper K, Tappenden P, Cantrell A, Ennis K. A systematic review of meta-analyses assessing the validity of tumour response endpoints as surrogates for progression-free or overall survival in cancer. Br J Cancer 2020;123:1686–96.
[13] Berghmans T, Pasleau F, Paesmans M, Bonduelle Y, Cadranel J, Toth IC, et al. Surrogate markers predicting overall survival for lung cancer: ELCWP recommendations. Eur Respir J 2012;39:9–28.
[14] Larkins E, Blumenthal GM, Chen H, He K, Agarwal R, Gieser G, et al. FDA approval: alectinib for the treatment of metastatic, ALK-positive non-small cell lung cancer following crizotinib. Clin Cancer Res: J Am Assoc Cancer Res 2016;22:5171–6.
[15] Wu Y-L, Herbst RS, Mann H, Rukazenkov Y, Marotti M, Tsuboi M. ADAURA: phase III, double-blind, randomized study of osimertinib versus placebo in EGFR mutation-positive early-stage NSCLC after complete surgical resection. Clin Lung Cancer 2018;19:e533–6.
[16] Halabi S, Armstrong AJ, Sartor O, de Bono J, Kaplan E, Lin CY, et al. Prostate-specific antigen changes as surrogate for overall survival in men with metastatic castration-resistant prostate cancer treated with second-line chemotherapy. J Clin Oncol: J Am Soc Clin Oncol 2013;31:3944–50.
[17] Chitkara A, Rai MP, Thawani R, Chen. EY-s. Recent analysis of frequency of surrogate end points used in oncology clinical trials 2006-2022. J Clin Oncol 2023;41:e13658 (-e).
[18] Kordecka A, Walkiewicz-Żarek E, Łapa J, Sadowska E, Kordecki M. Selection of endpoints in clinical trials: trends in european marketing authorization practice in oncological indications. Value Health 2019;22:884–90.
[19] Wheaton L, Bujkiewicz S. Use of surrogate endpoints in health technology assessment: a review of selected NICE technology appraisals in oncology. Int J Technol Assess Health Care 2025;41:e11.
[20] Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8:431–40.
[21] Cda-Am, Icer, Nice, Adhac, Zin, Iets, et al. Surrogate endpoints in cost-effectiveness analysis for use in health technology assessment. 2025.
[22] Bujkiewicz S, Achana F, Papanikos T, Riley R, Abrams K. Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints. NICE DSU Tech Support Doc 2019:20.
[23] Halabi S, Roy A, Rydzewska L, Guo S, Godolphin P, Hussain M, et al. Radiographic progression-free survival and clinical progression-free survival as potential surrogates for overall survival in men with metastatic hormone-sensitive prostate cancer. J Clin Oncol 2024;42:1044–54.
[24] Garcia JA, Spratt DE. Is it premature to accept radiographic progression-free survival as a surrogate end point in metastatic hormone-sensitive prostate cancer? J Clin Oncol 2024;42:2939–40.
[25] Gharzai LA, Jiang R, Jaworski EM, Rivera KM, Dess RT, Jackson WC, et al. Meta-analysis of candidate surrogate end points in advanced prostate cancer. NEJM Evid 2023;2:EVIDoa2200195.
[26] Higgins J.P.T., Thomas J., Chandler J., Cumpston M., Li T., Page M.J., et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated 2023): Cochrane; 2023.
[27] 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.
[28] Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012;12:9.
[29] Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? An analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the Biomarker-Surrogacy (BioSurrogate) Evaluation Schema (BSES). BMC Med Res Method 2012;12:27.
[30] Verhagen PCMS, Wildhagen MF, Verkerk AM, Vjaters E, Pagi H, Kukk L, et al. Intermittent versus continuous cyproterone acetate in bone metastatic prostate cancer: results of a randomized trial. World J Urol 2014;32:1287–94.
[31] Sweeney C, WMD II, Dreicer R, Chu F, Parks G, Baker K, et al. A randomized placebo-controlled trial of daily high-dose oral risedronate in men with metastatic prostate cancer commencing androgen deprivation therapy (ADT). J Clin Oncol 2010;28:e15000 (-e.).