264

Predictive Value of Intermediate Endpoints for Overall Survival in High-Risk Localized and Locally Advanced Prostate Cancer

Uwe Siebert[1,2], Francesco De Solda[3], Kimberly Hofer[4], James McCallion[3], Sharon A McCarthy[3], Suneel D Mundle[3], Mir-Masoud Pourrahmat[4], Victoria Wan[4], Beate Jahn[1]

[1]UMIT TIROL – University for Health Sciences and Technology, Hall in Tirol, Austria;
[2]ONCOTYROL – Center for Personalized Cancer Medicine, Innsbruck, Austria;
[3]Johnson & Johnson, Raritan, NJ, USA; [4]Evidinno Outcomes Research Inc., Vancouver, BC

Key Takeaways

To our knowledge, this is the first correlation analysis of NED and pCR as ICEs for OS in PC. Our results support existing evidence for MFS as a surrogate for OS in HR-LPC/LAPC

Despite using rigorous BRMA of RCT/non-RCT data showing consistent directional associations between OS and each ICE, our results did not meet HTA agencies' thresholds for a strong correlation, suggesting a need to better understand how agencies evaluate surrogacy thresholds

Conclusions

The strength of MFS correlation for OS is consistent with the publications by the ICECaP Working Group[5,6]

The moderate strength correlation estimates between EFS and NED with OS are novel findings. This is important because previous studies have shown weak correlation between EFS and OS, while the relationship between NED and OS has not been previously investigated

Our findings highlight a need for

- Universal ICE thresholds for HTA bodies, as we observed different interpretations of strength of correlation
- Ensuring there are health decision science-based approaches for ICE validation that include disease- and treatment-specific outcomes
- Further analysis to prove correlations, including analyses using real-world data

We are currently assessing MFS, EFS, and NED correlation estimates for OS using real-world data of patients with LPC/LAPC

Acknowledgments
Writing assistance was provided by Ann Tighe, PhD, of Parexel, and was funded by Johnson & Johnson.

Disclosures
A Valencia is an employee of Johnson & Johnson.

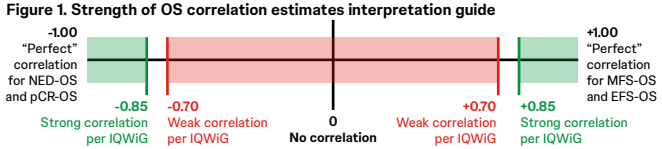
Introduction

- Overall survival (OS) is a well-established endpoint in studies of patients with advanced prostate cancer (PC)
- In high-risk localized or locally advanced PC (HR-LPC/LAPC), demonstrating meaningful OS treatment benefit requires a prolonged duration for metastases to develop or deaths to occur in earlier disease stages[1]
- Several intermediate clinical endpoints (ICEs) for OS have been used to assess early treatment outcomes in patients with PC, eg, metastasis-free survival (MFS), event-free survival (EFS), no evidence of disease (NED), and pathological complete response (pCR)
- The Institute for Quality and Efficiency in Health Care (IQWiG)[2] and NICE[3] provide frameworks and thresholds for strong correlation of ICEs for OS
- These frameworks provide structured approaches to analyze and validate the surrogacy of ICEs in PC research and health technology assessment (HTA)
- **Objective:** We aimed to evaluate the predictive value of ICEs for OS in HR-LPC/LAPC using data from randomized and non-randomized clinical studies (RCTs/non-RCTs)

Methods

- We conducted a systematic literature review using Embase®, MEDLINE®, CENTRAL, and gray literature to identify RCTs/non-RCTs investigating therapeutic options for adults with HR-LPC/LAPC. MFS, EFS, NED, and pCR were ICEs of interest
- English-language publications from database inception to March 22, 2024, reporting treatment effects for OS (the OS hazard ratio [HR_{OS}]) or ICEs (the HR or odds ratio [OR], HR_{MFS}, HR_{EFS}, OR_{NED}, OR_{pCR}) were included
- OS surrogacy validation entailed assessing the strength of correlation and predictive accuracy
- Strength of correlation estimates between each ICE and OS was evaluated using Bayesian bivariate random-effects meta-analysis (BRMA), employing an uninformative prior distribution. Strength of correlation estimates was assessed against IQWiG and Lassere et al.[4] criteria
- Predictive accuracy of each ICE was determined by assessing the observed versus predicted treatment effects of OS of each study via leave-one-out cross validation (LOOCV) as recommended by NICE[3]
- For strength of correlation estimates, either a positive or negative between-study correlation can indicate clinically meaningful treatment effects for ICEs, depending on a surrogacy equation of HR or OR (Figure 1)

Figure 1. Strength of OS correlation estimates interpretation guide



- Improvements in NED and pCR correspond to less evidence of disease and residual cancer calls, respectively. A perfect correlation for OS is -1.00
- NED assesses the proportion of patients without evidence of disease and no subsequent therapies at certain landmarks. The higher the proportion, the more patients with NED, and the better the OS
- pCR assesses the presence of residual tumor by measuring the absence of local disease following radical prostatectomy. The less residual tumor, the better long-term outcomes, including OS
- Longer duration of MFS and EFS are typically associated with better OS, indicating a direct proportional relationship. A perfect correlation for OS is +1.00
- MFS and EFS are time-to-event ICEs that measure the duration patients remain free of disease recurrence or progression

Results

- We selected 137 unique studies (RCTs, n=89; non-RCTs, n=48) for BRMA (Table 1)
- **Strength of correlation**
  - BRMA showed a strong strength of correlation estimate for MFS-OS and moderate strength correlation estimates EFS-OS and NED-OS (Table 2 and Figure 2). The strength of correlation estimate for pCR-OS was unclear
  - OS correlation estimates based on BRMA were not statistically significant. An RCT-only sensitivity analysis showed consistent results
  - Although between-study correlations were relatively high, the widths and lower bounds of the 95% credible intervals (CrIs) were <0.85, and therefore did not meet the IQWiG threshold for strong correlation
  - Strength of correlation estimates based on Lassere et al.[4] criteria ranged from good to fair
- **Predictive accuracy**
  - LOOCV showed good predictive accuracy; observed HR_{OS} were within their 95% prediction intervals (PIs) predicted from MFS, EFS, NED, and pCR for 83% to 100% of studies in LOOCV (Table 3 and Figure 3). Only NED met NICE criteria for validity as a surrogate for OS (Table 3)

Table 1. Number of studies and patients in analysis set

ICE	RCT/non-RCT studies, n	Patients, n
MFS	79	51,746
EFS	85	53,965
NED	112	70,228
pCR	100	56,187

Table 2. Strength of OS correlation estimates based on BRMA and interpretations from published criteria

ICE	BRMA between-study correlation (95% CrI)	Strength of OS correlation estimate interpretation		
		Our interpretation[a]	IQWiG criteria[b]	Lassere criteria[c]
MFS-OS	0.69 (-0.35 to 0.98)	Strong[d]	Unclear	Good
EFS-OS	0.53 (-0.23 to 0.96)	Moderate	Unclear	Fair
NED-OS	-0.53 (-0.95 to 0.39)	Moderate	Unclear	Fair
pCR-OS	0.06 (-0.89 to 0.94)	Unclear	Unclear	Fair

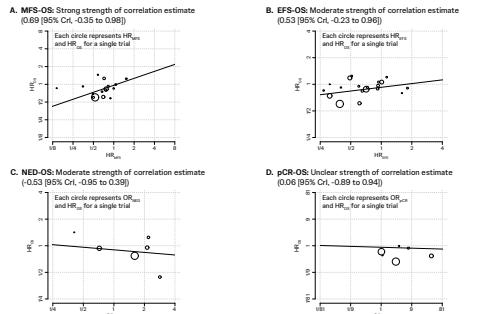
[a]Based on BRMA between-study correlation. [b]IQWiG requires a lower bound of between-study correlation estimate of ≥0.85 for strong correlation and an upper bound of ≤0.70 to be weak. Otherwise, correlation is unclear. [c]Lassere et al.[4] defines excellent association as r^2 ≥0.6, good association ≥0.4 and <0.6, fair ≥0.2 and <0.4, and poor <0.2. r^2 is not shown but can be derived by squaring the between-study correlation point estimate for assessing against the Lassere criteria. Note: Lassere criteria were developed for biomarker outcomes and not time-to-event outcomes. [d]Consistent with Intermediate Clinical Endpoints in Cancer of the Prostate (ICECaP) Working Group publications.[5,6]

Table 3. Validity of surrogacy for OS based on LOOCV and interpretation based on published criteria

ICE	Validity of surrogacy for OS		
	Proportion of studies HR_{OS} aligned with prediction, %[a]	Our interpretation based on LOOCV[a]	Interpretation based on NICE criteria[b]
MFS-OS	89.5	Validity met	Validity not met
EFS-OS	89.5	Validity met	Validity not met
NED-OS	100	Validity met	Validity met
pCR-OS	83.3	Validity met	Validity not met

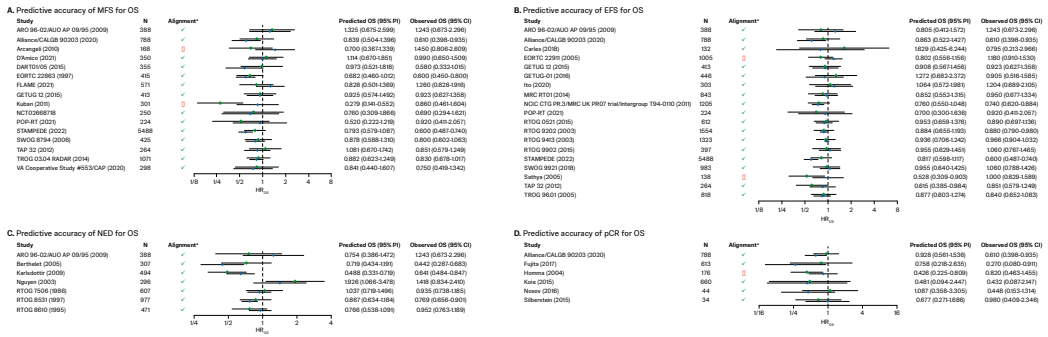
[a]Validity based on similar studies with ICECaP Working Group. [b]NICE criteria define a valid surrogate endpoint when ≥95% of the study's observed HRs are captured by the 95% PI from their LOOCV model; otherwise, validity is not met.

Figure 2. Linear relationship between natural log-transformed HRs and ORs of ICEs and HR_{OS}



The size of the circles is scaled to the sample size of the trial. The x- and y-axes are on the logarithmic scale. The solid line represents the estimated surrogacy equation for predicting HR_{OS} from HR_{MFS}, HR_{EFS}, OR_{NED}, and OR_{pCR}. Positive slopes are expected for MFS and EFS; negative slopes are expected for NED and pCR.

Figure 3. Predictive accuracy of MFS, EFS, NED, and pCR



Forest plots show model predictions to observed data during LOOCV for each ICE-OS. Trials are ordered alphabetically. Green diamonds and their error bars are the predicted HR_{OS} and their 95% PIs, whereas the blue diamonds and their error bars represent the reported HR_{OS} from the trials and their 95% confidence intervals (CIs), respectively. HRs in the observed OS column may not match the associated publications due to the imprecision and resulting asymmetry of published CIs. [a]Alignment between 95% PI and observed HR_{OS}: green checkmark indicates that the observed HR_{OS} was captured by the 95% PI of the predicted HR_{OS}; red square indicates that the observed HR_{OS} was not captured by the 95% PI of the predicted HR_{OS}.

References

1. Brookman-May S, et al. *Eur Urol*. 2024;86:81-87. 2. Institute for Quality and Efficiency in Health Care: Executive Summaries. 2005. 3. Bujkiewicz S, et al. NICE DSU technical support document. 2019;20. 4. Lassere MN, et al. *BMC Med Res Methodol*. 2012;12:27. 5. Xie W, et al. *J Clin Oncol*. 2017;35:3097-3104. 6. ICECaP Working Group, Sweeney C, et al. *J Natl Cancer Inst*. 2015;107:djv261.

Prostate Cancer

Presented by A Valencia at HTAi 2025 Annual Meeting; June 14-18, 2025; Buenos Aires, Argentina