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

Uwe Siebert^{1,2}, Francesco De Solda³, Kimberly Hofer⁴, James McCallion³, Sharon A McCarthy³, Suneel D Mundle³, Mir-Masoud Pourrahmat⁴, Victoria Wan⁴, Beate Jahn¹

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

Kev Takeawavs

264

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
- (i) 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

Acknowledaments

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 stages1
- Several intermediate clinical endpoints (ICEs) for OS have been used to assess early treatment outcomes in patients with PC, eq. 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 NICE³ 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)

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

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

Strength of OS correlation estimate interpretation **BRMA** between-study ICE correlation (95% Crl) Our interpretation IQWiG criteria Lassere criteria MFS-OS 0.69 (-0.35 to 0.98) Strongd Unclear Good EFS-OS 0.53 (-0.23 to 0.96) Moderate Fair Unclear NED-OS -0.53 (-0.95 to 0.39) Moderate Unclear Fair pCR-OS 0.06 (-0.89 to 0.94) Unclear Unclear Fair

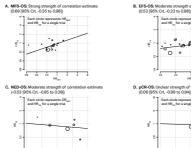
Methods

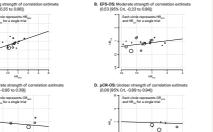
were ICEs of interest

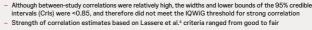
predictive accuracy

"Based on BRMA between-study correlation. HOWIG requires a lower bound of between-study correlation estimate of 20.85 for strong correlation and an upper bound of \$0.70 to be weak. Otherwise correlation is unclear. Lassere et al.⁴ defines excellent association as p² 206, good association a04 and <06, fair 202 and <04, and poor <02, p² 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. Consistent will Intermediate Clinical Endpoints in Cancer of the Prostate (ICECaP) Working Group publications

Figure 2. Linear relationship between natural log-transformed HRs and ORs of ICEs and HRos







Predictive accuracy

assessed against IQWiG and Lassere et al.4 criteria

cross validation (LOOCV) as recommended by NICE3

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

English-language publications from database inception to March 22, 2024,

reporting treatment effects for OS (the OS hazard ratio [HRoe]) or ICEs

· OS surrogacy validation entailed assessing the strength of correlation and

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

Predictive accuracy of each ICE was determined by assessing the observed

versus predicted treatment effects of OS of each study via leave-one-out

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)

(the HR or odds ratio [OR], HR_{MES}, HR_{FES}, OR_{NED}, OR_{nCR}) were included

- LOOCV showed good predictive accuracy; observed HRos 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 3. Validity of surrogacy for OS based on LOOCV and interpretation based on published criteria

	Validity of surrogacy for OS						
ICE	Proportion of studies HR _{os} aligned with prediction, %	Our interpretation based on LOOCV°	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				

Validity based on similar studies with ICECaP Working Group.*NICE criteria define a valid surrogate endpoint when >95% of the study's observed HRs are captured by the 95% PI from their I OOCV model: otherwise validity is not met

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

edictive accuracy of MFS for	OS					B. Predictive accuracy of EFS for OS					
idy .	N	Alignment*		Predicted OS (95% PI)	Observed OS (95% CI)	Study	N	Alignment*		Predicted OS (95% PI)	Observed OS (95%
D 96-02/AUO AP 09/95 (2009)	388	×		1325 (0.675-2.599)	1.243 (0.673-2.296)	ARO 98-02(AUO AP 09/95 (2009)	388	×		0.805 (0.412-1.572)	1243 (0.673-2.296
ance/CALGB 90203 (2020)	788	×		0.839 (0.504-1.396)	0.610 (0.398-0.935)	Alliance/CALGB 90203 (2020)	788	×		0.883 (0.522-1427)	0.610 (0.398-0.935
angeli (2010)	168	0		0.700 (0.367-1.339)	1450 (0.806-2.609)	Carles (2018)	152	× -		1829 (0.425-8.244)	0795 (0.213-2.98)
mico (2021)	350	×		1114 (0.670-1.851)	0.990 (0.650-1.509)	EORTC 22911 (2005)	1005	0		0.802 (0.556-1156)	1180 (0.910-1.530
RT01/05 (2015)	355	×		0.973 (0.521-1.818)	0.580 (0.332-1.015)	GETUG 12 (2015)	413	×		0.908 (0.587-1.458)	0.923 (0.827-1.35
RTC 22863 (1997)	415	×		0.682 (0.460-1.012)	0.800 (0.450-0.800)	GETUG-01 (2016)	446	×		1.272 (0.682-2.372)	0.905 (0.516-1.58
ME (2021)	671	×		0.828 (0.501-1.369)	1260 (0.828-1916)	Ito (2020)	303	×		1.064 (0.572-1.981)	1204 (0.689-210
TUG 12 (2015)	413	×		0.925 (0.574-1.492)	0.923 (0.827-1358)	MRC RT01(2014)	843	×		0.852 (0.552-1.315)	0.950 (0.677-1.35
an (2011)	301	0	· · · · · · · · · · · · · · · · · · ·	0.279 (0.141-0.552)	0.860 (0.461-1.604)	NCIC CTG PR.3/MRC UK PR07 trial/intergroup T94-0110 (2011)	1206	×		0760 (0.550-1.048)	0740 (0.620-0.8
102668718	250	×		0760 (0.309-1866)	0.690 (0.294-1.621)	POP-RT (2021)	224	×		0.700 (0.300-1.636)	0920 (0.411-2.0
P-RT (2021)	224	× –		0.520 (0.222-1.219)	0.920 (0.411-2.057)	RTOG 0521 (2015)	612	×		0.953 (0.659-1.378)	0.890 (0.897-11
MPEDE (2022)	5488	×		0.793 (0.579-1.087)	0.600 (0.487-0.740)	RTOG 9202 (2003)	1554	×		0.884 (0.655-1193)	0.880 (0790-03
06 8794 (2006)	425	×		0.878 (0.588-1.310)	0.800 (0.602-1.063)	RT0G 94(3 (2003)	1323	×		0.936 (0706-1.242)	0.968 (0.904-10
32 (2012)	264	×		1081 (0.670-1742)	0.851/0.579-1.249/	RTOG 9902 (2015)	397	×		0.955 (0.629-1.451)	1080 (0767-144
0G 03.04 RADAR (2014)	1071	×		0882 (0623-1249)	0.830 (0.678-1.017)	STAMPEDE (2022)	5488	×		0.817 (0.598-1117)	0.600 (0.487-0)
Cooperative Study #553/CAP (2020)	298	×		0.841 (0.440-1607)	0750 (0.419-1.342)	SW0G 9921 (2018)	983	×		0.955 (0.640-1425)	1060 (0788-14
				_		Sathya (2005)	138	0		0.528 (0.309-0.903)	1000 0.629-15
		1/8 1/4		8		TAP 32 (2012)	264	2		0.615 (0.385-0.984)	0.851 (0.579-12
			HR _{os}			TROG 96.01 (2005)	818	×		0.877 (0.603-1.274)	0.840 (0.652-1)
								1/8	14 1/2 1 2 4	3	
edictive accuracy of NED for	05					D. Predictive accuracy of pCR for OS			HR _{os}		
dv	N	Alignment*		Predicted OS (95% PI)	Observed OS (95% CI)	Study	N	Alignment*		Predicted OS (95% Pf)	Observed OS (95
0 96-02/AUO AP 09/95 (2009)	388			0.754 (0.388-1472)	1243 (0.673-2.296)	Aliance/CAL08 90203 (2020)	788	Auguman		0.928 (0.561-1.536)	0.610 (0.398-0.1
thelet (2005)	300			0.719 (0.434-1291)	0.442 (0.287-0.683)	Fuita (2017)	613	- 1		0.328 (0.361-1536) 0.758 (0.218-2.635)	0.270 (0.080-0
teleftir (2005) Isdottir (2009)	494			0.488 (0.331-0.719)	0.641 (0.484-0.847)	Homma (2004)	176			0.426 (0.225-0.809)	0.820 (0.463-14
water (2009) zven (2003)	296			- 1926 (1.066-3.478)	1418 (0.834-2.410)	Kole (2015)	660			0.481(0.094-2.447)	0.432 (0.087-2)
Jyan (2003) DG 7506 (1986)	607	- 1		1037 (0719-1496)	0.935 (0.738-1185)	Nose (2015) Nosey (2016)	44			1087 (0.358-3.305)	0.448 (0.153-1.2
DG 8531 (1997)	977	- 1		0.867 (0.634-1384)	0769 (0.658-0.901)	Silberstein (2015)	34	- 1		0.677 (0.271-1686)	0.980 (0.409-2.2
DG 8610 (1995)		- 1				Suberstein (2015)	34	· _		0.677 (0.271-0666)	0.960 (0.409-2.
	471			0.768 (0.538-1.091)	0.952 (0763-1389)			1/16	1/4 1 4 1		

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 lin represents the estimated surrogacy equation for predicting HR_{op} from HR_{ber}, HR_{tor}, oR_{ber}, and OR_{pok}. Positive slopes are expected for MFS and EFS; negative slopes are expected for 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% Pls, whereas the blue diamonds and their error bars represent the reported HR, 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. Alignment between 95% Pl and observed HR_{ge} green checkmark indicates that the observed HR_{ge} was captured by the 95% Pl of the predicted HR_{ge}; red square indicates that the observed HR_{ge} was not captured by the 95% Pl of the predicted HR_{ge}.

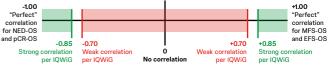
References

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. Buikiewicz 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



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

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

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		

