Assessing the surrogacy relationship between disease-free survival and overall survival for patients in clinical trials for adjuvant treatment of muscle-invasive urothelial carcinoma following radical resection

[POSB302]

Barkema AE[1], Simons CL[2], Kurt M[3], Broughton E[3], Teitsson S[4], May JR[4], Kroep S[1]

[1] OPEN Health, Rotterdam, Netherlands; [2] OPEN Health, York, United Kingdom; [3] Bristol Myers Squibb, Princeton, NJ, United States; [4] Bristol Myers Squibb, Uxbridge, United Kingdom

Background

- Bladder cancer (BC) is the 10th most common cancer in the world, with the most common type of BC being urothelial cancer (UC). For patients with muscle-invasive urothelial cancer (MIUC), the predominant form of UC, radical resection is standard therapy [1,2,3].
- Historically, most of the existing trials in adjuvant MIUC have investigated chemotherapy agents with overall survival (OS) being a primary endpoint; however, recent FDA approval of nivolumab, an immuno-oncology (I-O) agent, was based on disease-free survival (DFS).
- Due to their differing mechanisms of action, chemotherapies and I-O agents may induce fundamentally different endpoint associations as I-O agents often deliver improved efficacy and higher long-term survivorship rates but with a delayed response [4].
- For both patients and regulatory agencies, it is clinically beneficial to establish the relationship between intermediate endpoints and OS, particularly in early-stage clinical trials in oncology. Clinically proven surrogacy relationships to allow patients' access to treatments that are efficacious at earlier stage diseases [5]. Furthermore, due to growing use of I-O agents, it is of practical importance to assess the surrogacy relationship between DFS and OS with trial level data and determine its applicability to I-O trial-level data.
- Using the framework of Ciani et al. (2017) [6], and validation steps suggested by Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) [7], a meta-analytic approach can be used to evaluate the strength of association and determine the validity of predicting treatment effects on OS based on an observed intermediate endpoint such as DFS.
- Validity of a surrogacy relationship between DFS and OS is often studied in two major steps: 1) Endpoint correlation: DFS should have an effect on OS; and 2) Correlation between the hazard ratios (HRs): treatment effect on DFS should adequately predict the treatment effect on OS. The current study is focused on the latter type of surrogacy.
- Development of surrogate endpoints for OS can: (a) Provide greater statistical power for efficacy measurements; (b) Be minimally influenced by rapidly changing treatment landscapes in subsequent lines; (c) Reduce the overall clinical development time for new interventions; and (d) Help payers provide earlier access to new therapies.
- This study aims to assess feasibility of predicting treatment effects on OS based on treatment effects on DFS using trial-level data from patients receiving adjuvant treatment of MIUC following radical resection.

Methods

- A systematic literature review (SLR) [8] was conducted to identify RCTs (published up to February 13, 2021) in adjuvant treatments for patients with invasive UC (including MIUC of the bladder) who had been treated with radical resection and were eligible for adjuvant cisplatin-based therapy. A total of 21 RCTs qualified for inclusion in the surrogacy analysis.
- After excluding RCTs based on study quality and the availability of reported endpoints of interest, 5 and 6 chemotherapy trials were selected to evaluate the strength of association in the base case analyses of univariate and bivariate meta-analyses, respectively (Table 1).
  - Birtle et al. (2020) [9] reported OS (HR) only. Therefore, this study was only eligible for inclusion in the evidence base of the bivariate meta-analysis approach.
  - Two studies, Freiha et al. (1996) [11] and Lehmann et al. (2006) [13], were omitted from the base case analyses due to their low sample sizes (50 and 49 participants, respectively). These 2 trials were included in sensitivity analyses of both univariate and bivariate meta-analyses.
  - The IMVigor-010 trial [18] - comparing adjuvant atezolizumab with observation in patients with MIUC with reported DFS and OS data - was used to assess the predictive performance of the surrogacy model for an IO trial since the evidence base entirely consists of chemotherapy trials.
  - For all but one of the included studies, proportional hazards assumption was verified.
  - The types of treatments studied in each trial as well as the number of participants in each treatment arm of each trial are summarized in Table 1 below. The table also summarizes the list of studies analyzed in the base case and sensitivity analyses of each modelling approach.

Table 1. Trials included in the analysis per modelling method

Study	Sample size		Inclusion for univariate Frequentist/Bayesian meta-regression		Inclusion for bivariate Bayesian meta-analysis	
	Intervention	Comparator	Base case	Scenario analysis	Base case	Scenario analysis
Birtle et al.[9]	132	129	×	×	✓	✓
Cognetti et al.[10]	97	86	✓	✓	✓	✓
†Freiha et al.[11]	20	20	×	✓	×	✓
Lehmann et al.[12]	163	164	✓	✓	✓	✓
†Lehmann et al.[13]	26	23	×	✓	×	✓
†Skinner et al.[14]	44	47	✓	✓	✓	✓
Sternberg et al.[15]	141	143	✓	✓	✓	✓
Zhegalik et al.[16]	46	46	✓	✓	✓	✓
Total included			5	7	6	8

✓ Included, × Not included
HRs were calculated from reconstructed survival data (Guyot et al. [17]) after KM digitization

- The evaluation steps suggested by IQWiG [7] were used to assess the validity of surrogacy:
  1. A feasibility study was conducted to evaluate quality of the included trials and similarity of the endpoint definitions across trials.
  2. The strength of the evidence base and Pearson's correlation coefficient (ρ) between the treatment effects on DFS and OS (i.e. HR of DFS [DFS_{HR}] and OS [OS_{HR}]) in the selected chemotherapy trials were assessed.
  3. Multiple meta-analytic approaches were explored to generate the surrogacy equations

Frequentist univariate meta-regression (IQWiG [7] and European Network on Health Technology Assessment (EUnetHTA) [19] guidelines)

- Ordinary Least Squares (OLS)
- Weighted Least Squares (WLS), where trials were weighted by inverse-variance of their DFS_{HR} and their sample sizes

Bayesian meta-analytic approaches (Bayesian methods are based on Technical Support Document (TSD) 20 from the National Institute for Health and Care Excellence (NICE) [20])

- Univariate meta-regression, in which OS_{HR} is explained by using the surrogate endpoint as a covariate
- Bivariate meta-analysis, in which the trial-specific treatment effects of DFS and OS are modeled simultaneously considering a correlation between treatment effects on DFS and OS as well as uncertainty related to all parameters describing the surrogate relationship.

- HRs were log-transformed before conducting the analyses.
- When the estimated slope coefficient of these surrogacy models is significant, one can conclude that an association between DFS_{HR} and OS_{HR} is present. In this case, the surrogate threshold effect (STE) can be estimated as the minimum DFS_{HR} which would translate into a statistically significant OS benefit at a default 95 % confidence level (i.e. the 95% prediction interval [PI] of the model-predicted OS_{HR} should not include 1).
- When the estimated slope coefficient is not significant, changes in the surrogate endpoint are not expected to impact changes in the final endpoint, hence the STE cannot be defined.
- The accuracy of the models was tested by leave-one-out cross validation and comparing the predicted OS_{HR} with the reported OS_{HR} in the IMvigor-010 trial [18].

Results

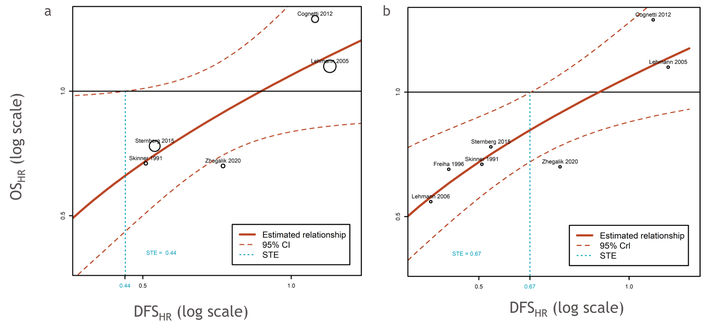
- Based on the feasibility study, the data used for both base case and sensitive analyses were of moderate reliability.[7]
  - The evidence base was weak, and few studies were included, most with a low level of quality.
  - Heterogeneity in multiple characteristics, such as endpoint definitions, length of follow up, and study design, across studies was present.
- The estimated correlation between DFS_{HR} and OS_{HR} was 0.84 (95% credible interval [CrI]: -0.18 ; 0.99) and 0.93 (95% CrI: 0.57 ; 0.99) for the base case and sensitivity analyses, respectively.
  - Due to limited number of studies in the evidence base and their relatively low sample sizes, PIs around the estimated surrogacy equations were wide.
  - Although the sensitivity analyses point out a higher correlation coefficient with narrower CrIs, the estimates of the 2 additional studies included here are highly uncertain due to the low sample size in these trials.
  - These are considered moderate strength correlations based on the IQWiG criteria, and do not independently prove a surrogacy relationship between the two endpoints.

Table 2. Summary results for the estimated surrogacy equations' parameters

Method	Slope: Point estimate (p-value) or [95% CrI]	Intercept: Point estimate (p-value) or [95% CrI]
Univariate frequentist (Studies unweighted)	0.622 (0.078)	0.049 (0.663)
Univariate frequentist (Studies weighted by inverse-variance of their DFS_{HR})	0.572 (0.063)	0.042 (0.609)
Univariate frequentist (Studies weighted by their sample sizes)	0.582 (0.052)	0.063 (0.468)
Univariate Bayesian	0.573 [0.096 ; 1.048]	0.023 [-0.144 ; 0.191]
Bivariate Bayesian	0.585 [-0.495 ; 1.711]	-0.046 [-0.443 ; 0.341]

- Results of the meta-analytic approaches generating surrogacy equations are presented in Table 2.
  - If we are expecting an association between DFS_{HR} and OS_{HR}, the slope coefficient should not be equal to zero, while the intercept coefficient should be zero.
  - Only the slope coefficient of the resulting equation for the univariate Bayesian meta-regression was found to be significant, and the slope coefficient of the univariate frequentist model weighted by sample size was borderline significant.
  - As for the leave-one-out cross validation, the 95% PIs for the OS_{HR} for each study covered the observed OS_{HR}. This surpasses the 95% threshold required for a 5% false positivity rate and is therefore considered a desirable result for model validity [20]. However, as the PIs from the model were wide and subject to high level of, this validation exercise should be interpreted with caution.
  - For the base case analysis, STE was non-existent for all models except the univariate frequentist approach using sample sizes of the studies as their weights and univariate Bayesian meta-regression models (Figure 1).
  - Estimated STEs from the univariate frequentist approach using sample sizes of the studies as their weights and univariate Bayesian meta-regression were both lower than the reported DFS_{HR} of the IMVigor-010 trial, meaning that no significant treatment effect on OS_{HR} can be expected. As the observed OS benefit in IMVigor-010 was statistically insignificant, the estimated STEs verified the trial outcome.

Figure 1. Demonstration of STE in both modelling approaches
(a) Univariate frequentist (weighted by sample size) (b) Univariate Bayesian



- In the base case analysis, for the IMVigor-010 trial, the predicted OS_{HR} from the univariate Bayesian meta-analysis model was 0.99 (95% confidence interval [CI]: 0.84 ; 1.17) whereas the observed OS_{HR} of the trial was 0.85 (95% CI: 0.66 ; 1.09). Therefore, the model overestimated the observed OS_{HR} (i.e. underestimated the treatment effect on OS) point estimate by 0.14.
- In the sensitivity analysis, the slope coefficient was found to be significant for all models, and the OS_{HR} of IMVigor-010 was overestimated by a range of [0.14 ; 0.21].

Conclusions

- According to IQWiG criteria, the correlation between the treatment effects can only be classified as moderate.
- OS_{HR} predictions from the estimated surrogacy equations using DFS_{HR} should be interpreted with a high degree of caution due to the limited scope of evidence and moderate association estimated between DFS_{HR} and OS_{HR} in our analyses.
- Based on the results of this study, predicting OS_{HR} based on DFS_{HR} in I-O trials in MIUC using surrogacy equations derived from adjuvant chemotherapy trials is not recommended.
- This study used trial-level data from publicly available resources. Therefore, HR calculations from the digitized data may be subject to precision errors which could be avoided if patient level data from the trials in the evidence base was available. Availability of patient level data could also inform endpoint correlation in addition to trial level correlation.

References

1. Bellmunt J, et al. Bladder cancer: ESMO Practice Guidelines for diagnosis, treatment and follow-up† on behalf of the ESMO Guidelines Working Group* incidence and epidemiology. 2014
2. Bajorin D, et al. Poster presentation at ESMO 2017; 10 Sept 2017 #2111
3. Cancer Research UK. Bladder cancer statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bladder-cancer#heading-Zero2019.
4. Anagnostou V, et al. Clin Cancer Res. 2017;23(17):4959-4969
5. Zhang, J et al. B J Cancer 2020;123:333-334.
6. Ciani O et al. Value Health. 2017;20(3):487-495
7. Validity of surrogate parameters in oncology. IQWiG-Berichte 80; 2011.
8. Trip AM, et al. Poster presentation at ISPOR EU 2021; Nov 2021 # POSB8
9. Birtle A, et al. The Lancet. 2020; 10232(395):1268-1277
10. Cognetti F, et al. Annals of Oncol. 2012;23(3):695-700
11. Freiha F, et al. J of urology. 1996;155(2):495-500
12. Lehmann J, et al. J Clin Oncol. 2005;23(22):4963-4974
13. Lehmann J, et al. BJU international. 2006;97(1):42-47
14. Skinner DG, et al. J of urology. 1991;145(3):459-464
15. Sternberg CN, et al. The lancet Oncol. 2015;16(1):76-86
16. Zhegalik AG, et al. C Eur J of Urology. 2020;73(1):26
17. Guyot P, et al. BMC MRM. 2012;12(1):9
18. Bellmunt et al. The Lancet. 2021;22(4): 525-537
19. Grigore, B et al. Pharmacoeconomics. 2020;38(10):1055-1070
20. Bujkiewicz S, et al. NICE DSU TSD 20. 2019.

Email: arnebarkema@openhealthgroup.com