CO119

Exploring the Association Between Treatment Effects on Cure Rates and Overall Survival (OS) in Resected Stage II/III Melanoma: A Correlation Meta Analysis (CMA) of Randomized Controlled Trials (RCTs)

Paul Serafini, Victoria Wan, Mir Sohail Fazeli, Murat Kurt
Evidinno Outcomes Research Inc., Vancouver, BC, Canada

Background

- The treatment landscape for resectable melanoma has been recently transformed by the introduction of targeted therapies and immune checkpoint inhibitors (ICIs), which led to substantial improvements in survival.[1]
- As a result, time to collect statistically mature overall survival (OS) data in randomized settings has increased. For example, the first publicly released OS data from the CheckMate 238 had minimum four years of follow-up, with number of death events below prespecified threshold for statistical maturity.[2] Similarly, mature OS data from KEYNOTE-054 are still pending.[3]
- As a result, identifying surrogate endpoints—intermediate outcomes which substitute for or predict OS[4]—is gaining more importance in oncology research to accelerate the assessment of new therapies.
- Elston and Taylor proposed three criteria for validation of a surrogate endpoint: biological plausibility of a causal mechanism between the surrogate and true endpoint, association between the surrogate and true endpoint at the individual patient level, and association between treatment effects on the surrogate and true endpoint across trials.[1]
- In adjuvant treatment of melanoma, association between the treatment effects on recurrence-free survival (RFS) and OS, measured by hazard ratios (HRs), has been investigated previously.[5,6]
- Prior research on RFS-OS surrogacy in resected stage II–III melanoma indicated that improvements in OS can be reliably predicted from improvements in RFS for a given pair of experimental-control therapies.
- Evaluating RFS HR as a predictor of OS HR in a clinical trial indirectly relies on the proportional hazards (PH) assumption, which may be violated in the presence of long-term survivors in a clinical trial or differences between the mechanisms of actions of experimental and control therapies
- As the long-term intent of adjuvant therapy is to prolong survival while offering clinical cure; "cure rate", defined as the proportion of patients who would be free from the risk of melanoma-related deaths and recurrence, can be an appealing alternative surrogate to RFS.[7] Due to its dichotomous nature, assessment of cure rates does not require the PH assumption. Furthermore, as the majority of recurrences and disease-related deaths occur within 3 to 5 years after tumor resection,[8] reliable assessment of cure rates does not require excessively long follow-up.

Objective

- To evaluate cure rate as a surrogate endpoint for OS in resected stage II/III melanoma at the trial-level, and assess how strength of surrogacy varies with key factors on long-term treatment benefit (subsequent treatment patterns, mechanism of action of adjuvant therapy, and the PH assumption).

Methods

Evidence Base

- Included studies were drawn from a previous analysis of RFS as a surrogate endpoint for OS in resected stage II–III melanoma,[5] which identified studies in a targeted literature review (TLR) searching Embase via Ovid from inception to October 10, 2022.
- Eligible studies were randomized controlled trials of patients with resected stage II or III melanoma which reported Kaplan Meier (KM) curves for RFS and either HRs or Kaplan-Meier (KM) curves for OS.
- For each study, the odds ratio (OR) of cure between arms was calculated using 5-year RFS rate as a proxy for cure rate. Studies not reporting 5-year RFS rates or with a null 5-year RFS in either of the arms were excluded.
- For studies which did not report the HR of OS (HR_{OS}), HR_{OS} were calculated using a Cox-PH model from pseudo individual patient data (IPD) reconstructed from the digitized KM curves.[9] For all studies with KM curves for OS, the PH assumption, which was required for the calculation of HR_{OS}, was tested on pseudo IPD using the Schoenfeld test.[10]

Surrogacy Analysis

- The association between the natural log OR of cure ($logOR_{cure}$) and the natural log HR of OS ($logHR_{OS}$) was evaluated using two different models:
  - Sample size-weighted linear regression (WLR).
  - Frequentist Bivariate Random Effects Meta-Analysis (BRMA).[11]
- The strength of association between $logOR_{cure}$ and $logHR_{OS}$ was quantified by the correlation estimates from WLR and BRMA.
  - Strength of association was classified as weak, moderate, or strong per German Institute for Quality and Efficiency in Health Care (IQWiG) criteria: strong when the upper limit of the 95% CI of the correlation was between –0.85 and -1, weak when the lower limit of the 95% CI of the correlation was between –0.70 and 0, and moderate otherwise.[12]
- Surrogacy equations predicting HR_{OS} from OR_{cure} were derived from both BRMA and WLR models.
- The predictive validity of the WLR model was evaluated using leave-one-out cross validation (LOOCV) in which, one study at a time, a new model was fitted to the data after removing that study from the evidence base and the HR_{OS} for the study was predicted from its OR_{cure} using the resulting model. For each omitted study, the predicted HR_{OS} and 95% prediction interval (PI) around it were compared to the observed HR_{OS} and its 95% confidence interval (CI).
- Additionally, the surrogate threshold effect (STE)—the minimum value of OR_{cure} (i.e., the minimum treatment effect on cure rates) to predict a statistically significant HR_{OS} at the default 95% confidence level—was calculated for studies with sample sizes of 800 and 1000. The sample sizes chosen for the STE calculation reflect the sizes of recent trials (CheckMate 238, EORTC 18071, KEYNOTE-054, COMBI-AD, CheckMate 76K, and KEYNOTE-716) that have transformed the treatment of early-stage melanoma.

Analysis Sets

- The following sensitivity analyses were conducted to investigate the impact of key statistical and clinical factors on the evaluation of long-term treatment effect
  1. Excluding studies violating the PH assumption in the OS data from the evidence base, to assess the impact of accurate description of input data for the calculation of HR_{OS} in the Cox PH model
  2. Restricting the evidence base to studies investigating interferon-α as the experimental therapy, to assess the impact of mechanism of action
  3. Restricting the evidence base to studies published after 2012, to assess the impact of changes in the metastatic treatment landscape
  4. Using 3-year RFS rates instead of 5-year RFS rates in the calculation of OR_{cure}, to evaluate the impact of the maturity of RFS data and the choice of proxy for cure rates

Results

- The TLR identified 30 studies published between 1978 and 2022 (median: 2013). Of these, 8 were excluded from the primary analysis. Among the excluded studies, three did not report RFS curves, four did not report 5-year RFS rates, and one reported null 5-year RFS rates. Among the 22 studies included in the primary analysis, four enrolled stage II disease patients, nine enrolled stage III disease patients, and nine enrolled a mix of stage II and stage III disease patients.
- The results of the primary and sensitivity analyses are summarized in **Table 1**.

Primary Analysis

- Both BRMA and WLR estimated moderate (per published IQWiG criteria) correlations of −0.78 (95% CI: −0.91, −0.53) and −0.82 (95% CI: −0.93, −0.56), respectively, with negative correlations indicating that higher values of OR_{cure} were associated with lower values of HR_{OS}. While results of BRMA and WLR were consistent with each other, neither of the correlation estimates met the IQWiG criterion for strong correlation.
- BRMA estimated a surrogacy equation of $logHR_{OS} = 0.11 – 0.64 \times logOR_{cure}$ with 95% CIs of (0.04 to 0.19) and (–0.98, –0.33) for the intercept and slope, respectively; emphasizing statistical significance of both of the parameters defining the surogacy equation
- WLR estimated a surrogacy equation of $logHR_{OS} = 0.09 – 0.79 \times logOR_{cure}$ with 95% CIs of (0.02 to 0.16) and (–1.05 to –0.54) for the intercept and slope, respectively. See **Figure 1**. In WLR, similar to the BRMA, both the intercept and slope were estimated to be statistically significant
- The STEs for studies with 800 and 1000 patients were estimated as 1.65 and 1.61, respectively, implying that OR_{cure} >1.65 (for 800 patients) and OR_{cure} >1.61 (for 1000 patients) would predict HR_{OS} < 1 with at least 95% probability.
- In LOOCV, the 95% PIs around the model-predicted HR_{OS}'s captured the observed HR_{OS}'s for 91% of studies, slightly less than the expected value of 95%.
- In LOOCV, the statistical significance of predicted and observed HR_{OS}'s were aligned for all studies.

Sensitivity Analyses

- With the exception of the sensitivity analysis using 3-year RFS rates as a proxy for cure rates, compared to primary analyses, all sensitivity analyses resulted in similar or stronger correlations, with smaller absolute average deviation (AAD) between the predicted and observed HR_{OS} (i.e., more accurate predictions), and STEs closer to 1 (i.e., smaller treatment effect on cure rates necessary to predict a statistically significant OS benefit).
- When 3-year RFS rates were used as a proxy for cure rates, compared to primary analysis, the correlation estimated from the BRMA was weaker, and AAD and the STE estimates were higher. The sensitivity of the results with respect to the proxy for cure rate emphasizes the importance of statistical maturity of data in assessing cure rates.
- Compared to the primary analysis, alignment rates between the 95% PIs around the predicted HR_{OS}'s and observed HR_{OS}'s in LOOCV were mixed across sensitivity analyses. Using only the studies with no violation of PH assumption in their OS data or 3-year RFS rate as a proxy for cure rate in the main evidence base led to higher alignment rates, whereas other settings in sensitivity analyses generated lower alignment rates.
- As in primary analysis, alignment rates between the statistical significance of the predicted and observed HR_{OS}'s were 100% across all sensitivity analyses, indicating the strong predictive potential of cure rates for the significance of OS benefit by experimental therapies in adjuvant treatment of resectable melanoma.

Table 1: Summary table of results from primary and sensitivity analyses

Analysis Set	N	Correlation Coefficient (95% CI)		Alignment in LOOCV		AAD*	Surrogate Threshold Effect	
		BRMA	WLR	95% PIs vs Observed HR_{OS}'s	Significance of Predicted vs Observed HR_{OS}'s		800 Patients	1000 Patients
Primary	22	−0.78 (−0.91, −0.53)	−0.82 (−0.93, −0.56)	20/22 (90.9%)	22/22 (100.00%)	0.14	1.65	1.61
Studies without a PH violation for OS	16	−0.93 (−0.98, −0.81)	−0.90 (−0.97, −0.67)	15/16 (93.8%)	16/16 (100.00%)	0.11	1.26	1.25
Studies with interferon-α as the experimental arm	15	−0.80 (−0.93, −0.51)	−0.81 (−0.94, −0.43)	13/15 (86.7%)	15/15 (100.00%)	0.12	1.47	1.42
Studies published after 2012	12	−0.95 (−0.99, −0.81)	−0.93 (−0.98, −0.74)	10/12 (83.3%)	12/12 (100.00%)	0.13	1.37	1.34
Using 3-year RFS rate as a proxy for cure rate	25	−0.64 (−0.80, −0.38)	−0.80 (−0.92, −0.54)	23/25 (92.0%)	25/25 (100.00%)	0.28	1.70	1.63

* – The average absolute difference between the observed and predicted HR_{OS} across the studies.

Abbreviations: AAD – Average Absolute Deviation, BRMA – Bivariate Random Effects Meta-Analysis, CI – Confidence Interval, N – Number of studies, OS – Overall Survival, PH – Proportional Hazards, RFS – Recurrence-Free Survival, WLR – Weighted Linear Regression, LOOCV – Leave-one-out cross validation, Year 2012 represents the introduction of immune checkpoint inhibitors for the treatment of metastatic melanoma.

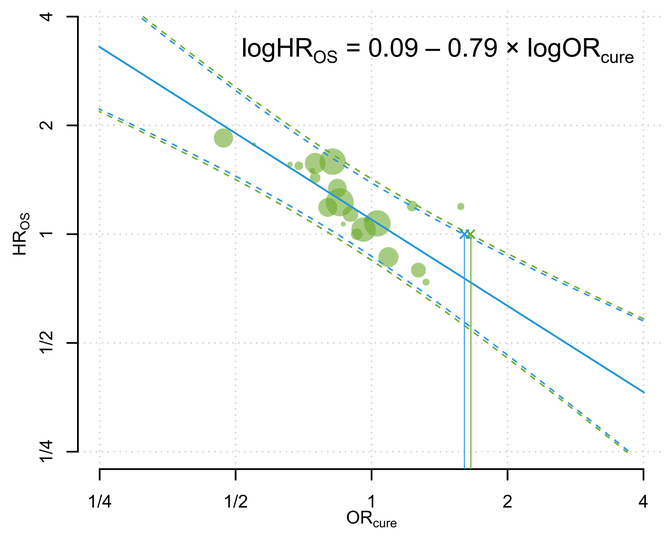

Figure 1: Scatterplot for the primary WLR analysis.

Note: The predictive surrogacy equation is graphed as the solid straight line. Each of the plotted green circles represents the (OR_{cure}, HR_{OS}) pair from a study. Sizes of the circles are proportional to the total number of patients within each study. The dotted curves refer to the 95% PIs for the HR_{OS} for a range of OR_{cure} for two hypothetical trials with sample sizes 800 (green) and 1000 (blue). Solid lines connecting the crosses to the x-axis indicate the STEs calculated for two hypothetical trials with sample sizes 800 (green) and 1000 patients (blue). In statistical terms it corresponds to the OR_{cure} at which the upper bound of the 95% PI of the HR_{OS} crosses 1.

Abbreviations: HR – Hazard Ratio, OR – Odds Ratio, OS – Overall Survival, PI – Prediction Interval, STE – Surrogate Threshold Effect, WLR – Weighted Linear Regression.

Conclusions

- The stability and strength of correlation between OR_{cure} and HR_{OS}, and cross-validation results highlight the potential of cure rates to assist earlier assessment of OS benefit in resected stage II/III melanoma. Sensitivity analyses were overall supportive of the findings from the base case, emphasizing the robust predictive potential of cure rates for long-term OS benefit. Notably the correlation between OR_{cure} and HR_{OS} was stronger when the evidence base was restricted to studies conducted after the introduction of ICIs.
- This study is the first meta-analysis in the literature to demonstrate that cure rates can serve as valid surrogates for OS in stage II/III melanoma. With its generalizable nature, the proposed meta-analytic approach can also be applied to other early-stage cancers.
- Results from the primary analysis should be applied with caution in future settings with modern therapies as the evidence base predominantly consists of interferon-α studies which predate the introduction of targeted therapies and ICIs in adjuvant and metastatic melanoma.
- Unlike RFS, cure rates are not easily measurable clinical outcomes. Therefore, they are not often used as primary or secondary endpoints of early-stage cancer trials. Because of this, application of our approach requires rigorous post-hoc estimation of cure rates from clinical trial data. A promising future research direction can extend our proposed approach and use mixture cure models to formally estimate cure rates from the RFS data, and this extension can enable the integration of treatment effects on surrogate outcomes for the uncured subgroup as a predictor into the meta-analysis.

References

1. Ernst M and Guibellion A. *Biomedicine*. 2022. 10(4):822
2. Ascierto PA et al. *Lancet Oncol*. 2020. 21(11):1465–77
3. Eggermont AMM et al. *NEJM Evid*. 2022. 1(11).
4. Taylor RS and Elston J. *NIHR Health Technology Assessment*. 2009. 13(8):1–50.
5. Leung L et al. *Immunooncol Technol*. 2025. 25.
6. Suciui S et al. *J Natl Cancer Inst*. 2018. 110(1):87–96.
7. Lambert et al. *Biostatistics*. 2007. 8(3):576–594.
8. Peirano et al. *Dermatol Pract Concept*. 2023. 13(4).
9. Guyot et al. *BMC Med Res Methodol*. 2012. 12(1).
10. Therneau TM and Grambsch PM. 2020. Modeling Survival Data: Extending the Cox Model. Springer.
11. Riley RD et al. *Biostatistics*. 2008. 9(1):172–186.
12. Institue for Quality and Efficiency in Health Care: Executive Summaries. 2011. V1.1.

Acknowledgments

Authors report employment with Evidinno Outcomes Research Inc. (Vancouver, BC, Canada). This research was conducted during Murat Kurt's employment with Evidinno Outcomes Research Inc.

EVIDINNO
EVIDENCE NAVIGATION & SYNTHESIS

Presented at ISPOR Europe 2025; November 9–12; Glasgow, Scotland, UK.

Email: mfazeli@evidinno.com