ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

IMMUNO-ONCOLOGY AND TECHNOLOGY

ORIGINAL ARTICLE

Challenges and opportunities of predicting overall survival benefit from improvements to recurrence-free survival in stage II/III melanoma: a correlation meta-analysis

L. Leung[1], J. M. Kirkwood[2], S. Srinivasan[3], M. Dyer[4], A. Qian[1], M.-M. Pourrahmat[1]*, E. Kasireddy[1], J. R. May[4], A. Moshyk[3] & M. Kurt[3]

[1]Evidinno Outcomes Research Inc., Vancouver, Canada; [2]Department of Medicine, University of Pittsburgh, Pittsburgh; [3]Bristol Myers Squibb, Princeton, USA; [4]Bristol Myers Squibb, Uxbridge, UK

Available online 3 February 2025

Background: We evaluated the association between treatment effects on recurrence-free survival (RFS) and overall survival (OS) in randomized controlled trials (RCTs) studying resected stage II/III melanoma.

Methods: Hazard ratios (HRs) of RFS and OS were obtained from a literature review. Bivariate random-effects meta-analysis (BRMA) and weighted linear regression (WLR) models estimated correlations [95% confidence interval (CI)] between HR_{RFS} and HR_{OS}. Slopes and intercepts of surrogacy equations were estimated. Surrogate threshold effect was derived from WLR for various sample sizes. Validity and predictive performance of WLR were assessed using leave-one-out cross-validation. Sensitivity analyses evaluated impact of RCTs violating proportional hazards assumption, publication year, treatments' mechanism of action, and cancer stage.

Results: Across 30 RCTs, treatments included interferon-α ($n = 17$), other immunotherapy-containing regimens ($n = 10$), immune checkpoint inhibitors ($n = 3$), and targeted therapies ($n = 2$). BRMA (0.68, 95% CI 0.45-0.82) and WLR (0.71, 95% CI 0.42-0.87) estimated moderate correlation between HR_{RFS} and HR_{OS}. Surrogate threshold effect was 0.66/0.68 for studies with 800/1000 patients. Slope coefficients were statistically significant in both models (95% CI 0.09-0.61 BRMA; 95% CI 0.41-0.92 WLR). The 95% prediction intervals around the HR_{OS} predicted by WLR accurately contained 29/31 (93.5%) of observed HR_{OS}. Across sensitivity analyses correlations ranged between 0.69 and 0.84 (BRMA) and 0.55 and 0.77 (WLR).

Conclusions: Statistically meaningful correlation between HR_{RFS} and HR_{OS} can assist earlier predictions of OS benefit from improvements in RFS for RCTs in resected stage II/III melanoma and provide insights for the earlier evaluation of emerging therapies. Primary model predictions should be approached with caution as nearly half of the evidence base comprised interferon-α trials.

Key words: surrogate endpoints, melanoma, meta-analysis, survival analysis, immunotherapy, adjuvant, targeted therapy

INTRODUCTION

Treatment of localized or regional melanoma generally consists of surgical resection and may include lymph node biopsy/dissection as well as adjuvant systemic therapy [including radiation and systemic therapy with cytokines, immune checkpoint inhibitors (ICIs), or targeted therapies] for patients with high-risk disease. Adjuvant treatment of melanoma has been recently transformed with BRAF—MEK inhibitors and ICIs.[1] ICIs including nivolumab, ipilimumab, and pembrolizumab were investigated in resected stage III/IV melanoma.[2-4] Among these, nivolumab and pembrolizumab are also being studied in resected stage II melanoma.[5,6] Besides ICIs, as a BRAF—MEK inhibitor combination therapy, dabrafenib plus trametinib,[7] was also evaluated in resected stage III melanoma.

Overall survival (OS) is widely recognized as the gold standard measure by regulatory and reimbursement agencies as well as payers in evaluating randomized, controlled trials (RCTs) in oncology. Positive OS results from an RCT provide irrefutable evidence that the treatment modality under investigation extends the life of the disease population.[8] The time required to observe statistically

*Correspondence to: Mir-Masoud Pourrahmat, Evidinno Outcomes Research Inc., 411 − 63 West 6th Avenue, Vancouver, BC V5Y 1K2, Canada. Tel: +1 604 992 0439

E-mail: mpourrahmat@evidinno.com (M.-M. Pourrahmat).

mature OS in the resected melanoma setting, however, is longer than other time-to-event outcomes. For example, in the adjuvant treatment setting, the first OS data publicly reported from CheckMate 238 study had 4 years of minimum follow-up,[2] while for KEYNOTE-054, OS outcomes have not been reported yet.[9] In contrast, the initial recurrence-free survival (RFS) data from these trials, which led to regulatory approval of nivolumab and pembrolizumab in the United States and Europe, had 1.5-years[10] and 1.25-years[4] of minimum follow-up, respectively. Therefore, RFS can be evaluated in place of OS to evaluate the clinical value of emerging treatments and determine the best choice of therapeutic agents for patients in a shorter time frame.[11] Unlike OS benefit, which may be subject to confounding effects of subsequent treatments depending on their local reimbursement statuses, improvements in RFS are observed sooner and free of the effects of post-recurrence treatments.

In randomized settings, RFS is generally defined as the time from randomization until the date of the first recurrence (local, regional, or distant) or death from any cause,[9] and is a common primary endpoint in resected stage II/III melanoma trials. Prior work investigating RFS-OS association in resected stage II/III melanoma trials using individual patient-level data (IPD) from adjuvant interferon-α and ipilimumab trials has shown a strong correlation at the patient level, but a moderate association at the trial level.[12,13] Since the conduct of the RCTs studied in these meta-analyses, the standard of care in resected stage II/III melanoma has changed significantly with the approval of BRAF—MEK inhibitors and modern ICIs. Despite the availability of OS data from some of these recent RCTs,[2,3,7] the strength of the association between RFS and OS has not been re-evaluated in the most recent treatment landscape.

This study aimed to fill this gap by assessing the strength of association between the treatment effects on RFS and OS in patients with resected stage II/III melanoma receiving adjuvant treatment using aggregate-level trial data.

MATERIALS AND METHODS

Targeted literature review

A targeted literature review was conducted by searching Embase via Ovid on 10 October 2022, without date limits. Studies were included if they either reported hazard ratios (HRs) or Kaplan—Meier (KM) curves, from which HR estimates could be derived, on both RFS and OS. Extraction tables included study design, baseline patient characteristics, and efficacy results.

Data and outcomes

The surrogate endpoint RFS was defined as the time from randomization until any type of recurrence or all-cause death, whichever occurs first. The targeted literature review also included RCTs using alternative endpoints such as disease-free survival (DFS), and disease-free interval (DFI) as surrogates if they were defined analogously to RFS or matched its primary definition (Supplementary Appendix A, available at https://doi.org/10.1016/j.iotech.2025.101042). Therefore, in the rest of this manuscript, 'RFS' will also be used to cover all these analogous surrogate time-to-event outcomes in the evidence base.

For the correlation meta-analysis (CMA), the estimates of comparative treatment effects on RFS and OS, measured by HRs, and the corresponding 95% confidence intervals (CIs) were extracted from all the RCTs in the evidence base. For all RCTs, KM curves were identified for both endpoints where possible and were digitized using WebPlotDigitizer.[14] For trials that had more than two treatment arms, only one contrast including the designated control arm and one of the randomly selected intervention arms was considered in CMA. For both endpoints, digitized survival data from the KM curves were utilized with the corresponding number at risk data to reconstruct the underlying time-to-event data in each arm of each RCT using the Guyot algorithm.[15] Notably, because the reported HRs for RFS and OS in each RCT were derived from the Cox proportional hazards model, which relies on the proportionality of hazards across arms over time, the reconstruction of time-to-event data aimed to test the proportional hazards assumption with the Schoenfeld test.[16] The reconstructed time-to-event data were also used to estimate the HRs for RFS and OS for the RCTs that have not reported these efficacy measures.

A total of 30 RCTs, published between 1978 and 2022 (median: 2013), were identified through the targeted literature review and included in the subsequent CMA (Supplementary Appendix B, available at https://doi.org/10.1016/j.iotech.2025.101042). Most of the studies ($n = 23$) were phase III trials, while there was only one phase II trial and six trials with unreported phase information. Fifteen of the RCTs were multinational, while 9 were conducted exclusively in the United States, followed by 2 in the UK, and 1 each in China, France, Italy, and Scotland. Most trials ($n = 26$) reported RFS as the primary endpoint. HRs were reported as comparative efficacy measures by 21 trials. For the remaining trials, HRs were estimated from the Cox proportional hazards models applied to the reconstructed time-to-event data.

Treatment comparisons from all RCTs are presented in Table 1. Treatments in most studies involved interferon-α [n (number of studies with at least one arm in the class) = 17], followed by other immunotherapy-containing regimens agents ($n = 10$), ICIs ($n = 3$), and targeted therapies ($n = 2$). In five studies, HRs were reported for multiple subgroups with differing stages of disease, giving rise to a total of 39 different contrasts for the CMA. The primary analysis included 31 contrasts. For each trial, details on the efficacy data used in analyses and the stages of the corresponding disease population, together with the source of the HRs (i.e. reported or estimated from KM curves), are presented in Table 2.

Statistical methods

Associations between the treatment effects on RFS and OS were assessed using two statistical models, a bivariate

Table 1. Trials in the evidence base and treatments they investigated

Trial	Intervention	Comparator
AIM HIGH	IFN-α-2a	PBO
AVAST-M[a]	Bevacizumab	PBO
Cameron et al. 2001	IFN-α-2b	PBO
Cascinelli et al. 2001	IFN-α-2b	PBO
CheckMate 238[a]	Nivolumab	Ipilimumab
COMBI-AD[a]	Dabrafenib + Trametinib	PBO
Creagan et al. 1995	IFN-α-2a	PBO
ECOG-ACRIN E1609[a]	Ipilimumab	High-dose IFN-α
ECOG 1690	IFN-α-2b (low dose)	PBO
Eigentler et al. 2016[a]	Pegylated IFN-α-2a	IFN-α-2a
EORTC 18071[a]	Ipilimumab	PBO
EORTC 18081[a]	Pegylated IFN-α-2b	PBO
EORTC 18871/DKG 80-1	Iscador-M®	PBO
EORTC 18952	IFN-α-2b (lower dose)	PBO
EORTC 18961[a]	GM2-KLH/QS-21 vaccine	PBO
EORTC 18991	Pegylated IFN-α-2b	PBO
EORTC E1697[a]	IFN-α-2b	PBO
Flaherty et al. 2014[a]	Biochemotherapy (dacarbazine, cisplatin, vinblastine, interleukin-2, IFN-α-2b and granulocyte colony-stimulating factor)	High-dose IFN
Garbe et al. 2008	IFN-α-2a	PBO
Gonzalez et al. 1978[a]	Levamisole	PBO
Khammari et al. 2020[a]	Adoptive tumor-infiltrating lymphocytes therapy and interleukin-2	Abstention (did not receive any other melanoma treatments before inclusion)
Kim et al. 2009[a]	Biochemotherapy (cisplatin, vinblastine, dacarbazine, IFN-α-2b, interleukin-2)	IFN-α-2b (high or intermediate dose)
Lian et al. 2013[a]	Temozolomide + cisplatin	IFN-α-2b (high dose)
MAVIS[a]	Seviprotimut-L	PBO
Miller et al. 1988[a]	Transfer factor	PBO
Mohr et al. 2015[a]	Intermittent high-dose IFN-α-2b	IFN-α-2b (high dose)
Nordic IFN trial[a]	IFN-α-2b (3 years)	PBO
Oratz et al. 1991[a]	Melanoma antigen vaccine + cyclophosphamide	Melanoma antigen vaccine
SWOG-9035[a]	Melacine cell lysate + DETOX	PBO
Wallack et al. 1995[a]	Vaccinia melanoma oncolysate	Vaccinia vaccine virus - PBO

DETOX, detoxified Freund adjuvant, containing mycobacterial cell wall skeleton plus monophosphoryl lipid A; IFN, interferon-α; PBO, this term broadly refers to placebo, no treatment, or observation.
[a]Denotes trials that were not included in the meta-analysis conducted by Suciu et al.[12]

random-effects meta-analysis (BRMA) model and weighted linear regression (WLR). The BRMA methodology proposed by Riley et al.[17] was used, as suggested by the National Institute for Health and Care Excellence[18] for surrogacy assessments when the within-study correlations in meta-analyses are unknown. This methodology is the appropriate choice to handle aggregate-level data in the absence of within-study correlations between the endpoints which requires IPD. Inputs for the models were log-transformed (i.e. natural log) HR_{RFS} and HR_{OS} from the contrasts and their 95% CIs. In the WLR, each contrast was weighted by the total number of patients it included. Associations were measured by the correlation parameter from the BRMA as well as Pearson's correlation from the WLR. To account for the impact of disease staging (i.e. stage II or III) on the strength of correlation, a separate meta-regression considering disease stage as a binary covariate in the WLR framework was also employed. Statistical significance of slopes and intercepts of the surrogacy equations generated from BRMA and WLR was also derived. See Supplementary Appendix C, available at https://doi.org/10.1016/j.iotech.2025.101042 for further details on the statistical methods.

From the WLR, the surrogate threshold effect, defined as the minimum treatment effect on RFS that would translate into a statistically significant and positive treatment effect on OS, was computed. Surrogate threshold effect was derived at a default 95% significance level for a range of practical sample sizes of RCTs conducted in stage II/III melanoma. Compared with a simple correlation measure, which may not provide interpretable insights for the expected OS benefit in a given RCT, the surrogate threshold effect can provide immediate clinical insights on the long-term potential of observed RFS benefit and its translation to a significant OS benefit at a given confidence level.[19]

The surrogacy relationship between RFS and OS in resected stage II/III melanoma was previously assessed in a meta-analysis conducted by Suciu et al.[12] using IPD from various RCTs investigating interferon-α or ICIs. In this prior assessment, RFS was identified as a valid surrogate endpoint for OS. This surrogacy analysis focused specifically on the trials comparing interferon-α to observation, however, and therefore does not reflect the current treatment landscape in melanoma which was transformed by the introduction of ICIs and targeted therapies. Ipilimumab, the first ICI to be approved for patients with advanced melanoma, was approved in 2011 in the USA and 2012 in Europe, followed by nivolumab, pembrolizumab, nivolumab + ipilimumab, and dabrafenib + trametinib in

Table 2. Breakdown of the evidence base in terms of disease stages studied by the RCTs and the contrasts generated from each RCT along with the source of the hazard ratio estimates

Trial	Disease stages studied	HR_{OS} source	HR_{RFS} source	Surrogate reported as RFS	Primary analysis (k = 31)	Sensitivity analysis 1 (k = 22)	Sensitivity analysis 2 (k = 26)	Sensitivity analysis 3 (k = 15)	Sensitivity analysis 4 (k = 17)	Meta-regression (k = 28)
AIM HIGH	II and III	KM[a]	KM[a]	Yes	✔	✔	✔	✔	✖	✖
AVAST-M	II and III	HR[a]	HR[a]	No (DFI)	✔	✔	✔	✖	✔	✖
Cameron et al. 2001	II and III	KM[a]	KM[a]	No (DFS)	✔	✖	✔	✔	✖	✖
Cascinelli et al. 2001	III	KM[a]	KM[a]	No (DFS)	✔	✔	✔	✔	✖	✔
CheckMate 238	IIIb and IIIc	HR[a]	HR[a]	Yes	✔	✔	✔	✖	✔	✔
COMBI-AD	III	HR[a]	HR[a]	Yes	✔	✖	✔	✖	✔	✔
Creagan et al. 1995	II	KM[a]	KM[a]	Yes	✔	✔	✖	✔	✖	✔
ECOG-ACRIN E1609	IIIb	HR[a]	HR[a]	Yes	✔	✔	✔	✖	✔	✔
ECOG-ACRIN E1609	IIIc	HR[a]	HR[a]	Yes	✔	✔	✔	✖	✔	✔
ECOG 1690	II and III	HR[a]	HR[a]	Yes	✔	✖	✔	✔	✖	✖
Eigentler et al. 2016	II and III	HR[a]	HR[a]	No (DFS)	✔	✖	✔	✔	✔	✖
EORTC 18071	III	HR[a]	HR[a]	Yes	✔	✔	✔	✖	✔	✔
EORTC 18081	IIb and IIc	KM[a]	HR[a]	Yes	✔	✔	✔	✔	✔	✔
EORTC 18871/ DKG 80-1	II and III	HR[a]	HR[a]	No (DFI)	✔	✔	✔	✖	✖	✖
EORTC 18871/ DKG 80-1	IIb	HR[a]	HR[a]	No (DFI)	✖	✖	✖	✖	✖	✔
EORTC 18871/ DKG 80-1	III	HR[a]	HR[a]	No (DFI)	✖	✖	✖	✖	✖	✔
EORTC 18952	II and III	HR[a]	HR[a]	Yes	✔	✖	✔	✖	✔	✖
EORTC 18952	III	HR[a]	HR[a]	Yes	✖	✖	✖	✔	✖	✔
EORTC 18961	II	HR[a]	HR[a]	Yes	✔	✔	✔	✖	✔	✔
EORTC 18991	III	HR[a]	HR[a]	Yes	✔	✔	✔	✔	✖	✔
EORTC E1697	II and III	HR[a]	HR[a]	Yes	✔	✔	✔	✔	✔	✖
Flaherty et al. 2014	III	HR[a]	HR[a]	Yes	✔	✔	✔	✔	✔	✔
Garbe et al. 2008	III	HR[a]	HR[a]	No (DFS)	✔	✔	✔	✔	✖	✔
Gonzalez et al. 1978	II	KM[a]	KM[a]	No (DFS)	✔	✖	✖	✖	✖	✔
Khammari et al. 2020	III	HR[a]	HR[a]	No (DFS)	✔	✔	✔	✖	✔	✔
Kim et al. 2009	III	KM[a]	KM[a]	Yes	✔	✔	✔	✔	✖	✔
Lian et al. 2013	II	HR[a]	HR[a]	Yes	✖	✖	✖	✖	✖	✔
Lian et al. 2013	II and III	KM[a]	KM[a]	Yes	✔	✖	✔	✖	✔	✖
Lian et al. 2013	III	HR[a]	HR[a]	Yes	✖	✖	✖	✖	✖	✔
MAVIS	II and III	HR[a]	HR[a]	Yes	✔	✔	✔	✖	✔	✖
MAVIS	IIb and IIc	HR[a]	HR[a]	Yes	✖	✖	✖	✖	✖	✔
MAVIS	IIIa	HR[a]	HR[a]	Yes	✖	✖	✖	✖	✖	✔
MAVIS	IIIb and IIIc	HR[a]	HR[a]	Yes	✖	✖	✖	✖	✖	✔
Miller et al. 1988	II	KM[a]	KM[a]	No (DFI)	✔	✔	✖	✖	✖	✔
Mohr et al. 2015	III	HR[a]	HR[a]	Yes	✔	✔	✔	✔	✔	✔
Nordic IFN trial	II and III	HR[a]	HR[a]	Yes	✔	✔	✔	✔	✖	✖
Oratz et al. 1991	II	KM[a]	KM[a]	No (DFS)	✔	✖	✖	✖	✖	✔
SWOG-9035	II	HR[a]	HR[a]	Yes	✔	✔	✔	✖	✔	✔
Wallack et al. 1995	II	KM[a]	KM[a]	No (DFI)	✔	✖	✖	✖	✖	✔

Note: 25 trials reported data for one disease stage only, while one reported data for four stages, two reported data for three stages, and two reported data for two stages, giving rise to a total of 39 different contrasts.

Sensitivity analysis 1: restricted to trials that did not fail the proportionality test. Sensitivity analysis 2: restricted to trials which were published after year 2000. Sensitivity analysis 3: restricted to trials that investigated interferon-α as the experimental treatment. Sensitivity analysis 4: restricted to trials which were published after year 2012. Meta-regression analysis: analysis adjusting for stage II or stage III disease.

DFI, disease-free interval; DFS, disease-free survival; HR, hazard ratio; k, number of contrasts; KM, Kaplan—Meier curve; OS, overall survival; RFS, recurrence-free survival.

[a]Estimate measure used in the surrogacy analysis.

subsequent years—yet none of these therapies were represented in the evidence base of Suciu et al.[12] To compare the impact of the evidence base on the significance and strength of the correlation, we analyzed the aggregate level data from this prior assessment through WLR.[12] In this prior meta-analysis, standard errors or 95% CIs around the slope and intercept of the published surrogacy equation were not reported.[12] Furthermore, the weights used for each RCT in their WLR analysis were unclear, whereas the current analysis used the total number of patients in each contrast as weights in WLR. Therefore, to facilitate a head-to-head comparison between the two evidence bases, we fitted a WLR to the aggregate level RFS and OS data used by the prior meta-analysis[12] and assumed sample sizes of the RCTs as their weights.

Datasets for primary and sensitivity analyses

The analyses were classified into two sets as 'primary' and 'sensitivity' with respect to the input data analyzed. The primary analysis included all trials in the evidence base. Three sets of sensitivity analyses and a meta-regression were carried out to investigate the stability of the results with respect to changes in the evidence base. Sensitivity analyses were restricted to RCTs that were (i) not violating the proportional hazards assumption, (ii) published after the year 2000, (iii) investigated interferon-α as the experimental arm therapy, and (iv) published after the approval of ipilimumab as the first ICI for metastatic treatment of melanoma in 2012. A fifth sensitivity analysis employed a meta-regression considering disease stage at baseline as a binary covariate with two levels (0—stage II; 1—stage III). The analysis set for the meta-regression included all HRs reported or computed from either stage II or stage III patients exclusively. Contrasts including a mixture of stage II and III patients were not considered in the meta-regression.

Assessment of the validity and strength of surrogacy

Leave-one-out cross-validation (LOOCV) was used to assess the predictive performance of the surrogacy equation obtained from the WLR.[20-22] In the LOOCV, for each contrast, the corresponding data were removed and the WLR model was re-parameterized with the remaining data from other contrasts. The surrogacy equation obtained from the new regression model was used to predict the HR_{OS} of the removed contrast via its observed HR_{RFS}. The 95% prediction interval (PI) generated by the model provides predictive inference for the HR_{OS} of a prospective trial. More specifically, it provides an estimate of the interval into which the HR_{OS} of a given trial will fall with 95% chance. In LOOCV, validity of the model was assessed by the total fraction of the contrasts for which the observed HR_{OS} were covered by the 95% PIs generated.[18] Additionally, external validity of the WLR model using the evidence base of the primary analysis was assessed on two phase III trials (SWOG 1404 and BRIM8) that were not included in our evidence base.[23,24]

In the absence of universally accepted criteria or guidelines published by a regulatory or reimbursement authorization agency to classify the strength of a surrogacy relationship, we adopted relatively stringent guidelines published by the German Institute for Quality and Efficiency in Health Care (IQWiG) to assess the strength of correlation.[25] According to IQWiG, a correlation is classified as 'strong' if the lower limit of its 95% CI is $\geq$0.85, 'weak' if the upper limit of its 95% CI is $\leq$0.7, and 'moderate' otherwise.

RESULTS

The primary analysis included 31 contrasts in total from all trials. The correlation in the primary analysis was estimated as 0.68 (95% CI 0.45—0.82) by the BRMA. Using the WLR (Figure 1), Pearson's correlation was estimated to be 0.71 (95% CI 0.42—0.87). The surrogacy equation estimated from the BRMA was $\log(HR_{OS}) = -0.06 + 0.33 \times \log(HR_{RFS})$, where 95% CIs for the slope and the intercept were (0.09—0.61) and (−0.12 to 0.00), respectively. For example, for an HR_{RFS} of 0.70, the predicted HR_{OS} from the BRMA model was $\exp[-0.06 + 0.33 \times \log(0.70)] = 0.84$. In this equation for each unit increase in the $\log(HR_{RFS})$ was associated with 0.33 units of increase in $\log(HR_{OS})$ highlighting a non-linear and monotone relationship between the two HRs where the rate of increase in HR_{OS} is decreasing with higher value of HR_{RFS}. For instance, when HR_{RFS} was increased from 0.90 to 1.00, the corresponding increase in the HR_{OS} was 0.03 units (from 0.91 to 0.94), whereas when the HR_{RFS} was increased from 0.50 to 0.60, the corresponding increase in the HR_{OS} was 0.05 units (from 0.75 to 0.80). The surrogacy equation estimated from WLR was $\log(HR_{OS}) = -0.01 + 0.67 \times \log(HR_{RFS})$, where 95% CIs for the slope and the intercept were (0.41—0.92) and (−0.09 to 0.06), respectively. Although the intercepts from the two equations were similar, the slopes were not which could be due to methodological differences between the two approaches. Overall, the equations from both approaches agreed with each other in terms of statistical significance of their slopes and intercepts. While both equations had a statistically significant slope, the intercepts of the equations were either statistically insignificant or at the borderline of statistical insignificance.

Surrogate threshold effects were calculated for two hypothetical trials with 800 and 1000 patients. These sample sizes were chosen as the representative range of the sample sizes of recent positive RCTs (CheckMate 238, EORTC 18071, KEYNOTE-054, COMBI-AD, CheckMate 76K, and KEYNOTE-716) shaping the current standard of care in adjuvant treatment of melanoma. For a hypothetical RCT with a sample size of 800/1000 patients, the estimated surrogate threshold effect was 0.66/0.68, indicating that a reported $HR_{RFS} \leq 0.66/0.68$ would lead to a statistically significant HR_{OS} at the 95% confidence level. Given an RCT, although the point estimate of the HR_{OS} predicted from the HR_{RFS} is independent of the sample size for both BRMA and WLR, the 95% PI of the HR_{OS} tends to shrink with increasing sample sizes. Therefore, intuitively, the surrogate threshold

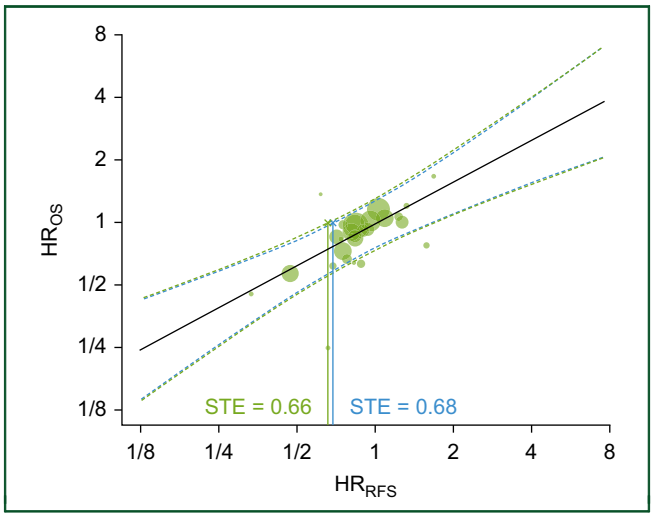

Figure 1. Summary of WLR and corresponding regression line for primary analysis using the total number of patients within each comparison as weights (primary analysis). The predictive surrogacy equation is graphed as the solid straight line in black. Each of the plotted green circles represents the (HR_{RFS}, HR_{OS}) pair from a contrast. Sizes of the circles are proportional to the total number of patients within each contrast. The dotted curves refer to the 95% PIs for the HR_{OS} for a range of HR_{RFS} 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 HR_{RFS} at which the upper bound of the 95% PI of the HR_{OS} crosses 1.
HR, hazard ratio; OS, overall survival; PI, prediction interval; RFS, recurrence-free survival; STE, surrogate threshold effect; WLR, weighted linear regression.

effect depends on the sample size of the RCTs and is expected to be relatively higher in larger trials. Nevertheless, the modest increase in the surrogate threshold effect within the range of practical sample sizes mentioned earlier signifies the robustness of the model.

Another piece of evidence solidifying the robustness of the model was its predictive performance during cross-validation. In LOOCV, observed HR_{OS} were within the 95% PIs generated by the model for 29 out of 31 (93.5%) contrasts in the primary analysis (Figure 2). Status of significance between observed and predicted HR_{OS} was aligned in 27 out of 31 (87.1%) contrasts. Across all trials, the average absolute gap between observed and predicted HR_{OS} was 0.15.

To investigate the predictive performance of the WLR calibrated from all trials in the evidence base, external validation was conducted using data from phase III SWOG 1404 and BRIM-8 trials. In the SWOG 1404 trial, the efficacy of adjuvant pembrolizumab was investigated versus standard of care (interferon-α or ipilimumab), whereas in the BRIM-8 trial the efficacy of adjuvant vemurafenib was investigated versus placebo.[23,24] Predicted HR_{OS} from the WLR were very close to their reported counterparts in both of these trials. Specifically, for the SWOG 1404 trial, the WLR predicted an HR_{OS} of 0.83 (95% PI 0.60—1.13) from an HR_{RFS} of 0.77 where the reported estimate of HR_{OS} was 0.82 (95% CI 0.72—0.94). For the BRIM8 trial, however, the surrogacy equation derived from the WLR predicted an HR_{OS} of 0.74 (95% PI 0.51—1.06) from an HR_{RFS} of 0.65 where the reported estimate of HR_{OS} was of 0.76 (95% CI 0.49—1.18).

In sensitivity analysis restricted to trials not violating the proportional hazards assumption, the correlation was estimated as 0.74 (95% CI: 0.48, 0.89) by the BRMA and 0.65 (95% CI: 0.25, 0.86) by the WLR. These estimates were similar to those obtained in the primary analysis. In sensitivity analysis restricted to RCTs published after the year 2000, the correlation was estimated as 0.81 (95% CI 0.62—0.91) by the BRMA and 0.77 (95% CI 0.50—0.90) by the WLR. These estimates point out stronger correlations compared with those from the primary analysis. The meta-regression generated similar correlation estimates to the primary analysis. Specifically, the estimated correlation was 0.69 (95% CI 0.46—0.84) by the BRMA and 0.72 (95% CI 0.39—0.89) by the WLR.

In the evidence base, there were six studies investigating modern therapies such as ICIs (e.g. nivolumab, ipilimumab) and targeted therapies (e.g. bevacizumab, dabrafenib plus trametinib). Despite the feasibility of a separate CMA including only these trials, predictions based on a sparse evidence base including only a handful of studies would be subject to high uncertainty. Therefore, we addressed this limitation by conducting a sensitivity analysis where the evidence base was restricted to trials investigating only interferon-α in the experimental arm where the purpose was to identify the degree of dominance of interferon-α trials on the results of primary analysis. There were 15 contrasts in this sensitivity analysis. Due to conflicting results obtained from the two different methodologies we employed it was not possible to draw a decisive conclusion on whether the correlation between the treatment effects on RFS and OS was stronger or weaker among the interferon-α trials. Specifically, while BRMA estimated a stronger correlation between the treatment effects on two endpoints among interferon-α trials than the correlation obtained in primary analysis including all trials (0.76 versus 0.68), WLR estimated a weaker correlation between the treatment effects on two endpoints in this sensitivity analysis than in the primary analysis (0.55 versus 0.71). Correlations estimated from WLR (95% CI −0.04 to 0.85) and BRMA (95% CI 0.45—0.90) were both classified as moderate according to the IQWiG criteria. During cross-validation, in 14 out of 15 contrasts, the reported HR_{OS} was captured within the 95% PIs for the predicted HR_{OS} generated by the WLR (Table 3).

Compared with the primary analysis, for the range of sample sizes considered, the estimated surrogate threshold effects were substantially lower (0.42/0.47 for sample size of 800/1000 patients) in the subgroup analysis including interferon-α trials implying a substantially higher requirement on the RFS benefit to observe a statistically significant OS benefit from an interferon-α treatment. This drastic difference in the threshold RFS benefit between the two sets of analysis can be partly explained by the lack of effective subsequent treatment options during the interferon-α era in the adjuvant setting, where most of the survival benefit would have to be accrued before a recurrence (Table 3).

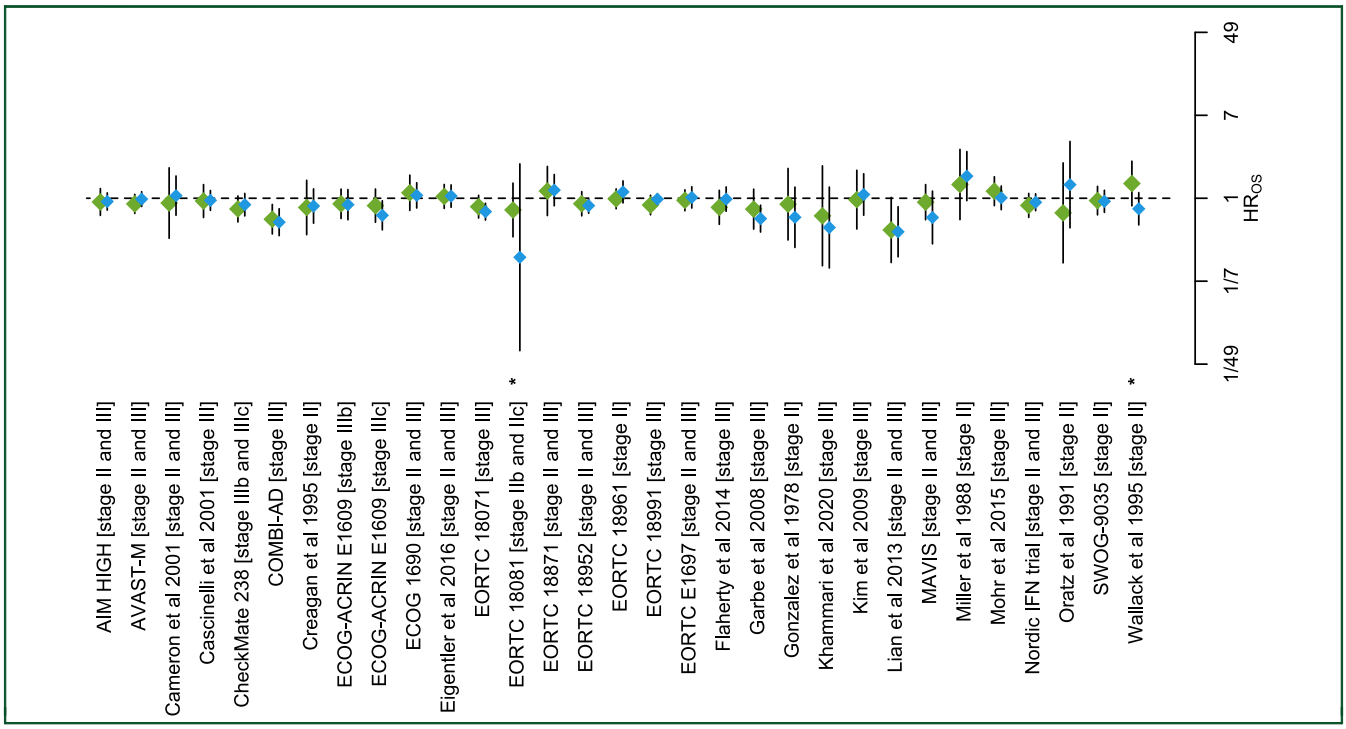

Figure 2. Results from the LOOCV in the primary analysis. The blue diamonds and their error bars represent the HR_{OS} and its 95% CI reported from the trials or calculated from reconstructed survival data, respectively. The green diamonds and their error bars represent the predicted HR_{OS} and their 95% PI obtained from the WLR, respectively. The asterisks indicate the RCTs where the observed HR_{OS} was not covered by the corresponding 95% PI.
CI, confidence interval; HR, hazard ratio; LOOCV, leave-one-out cross-validation; OS, overall survival; RCT, randomized, controlled trial; PI, prediction interval; WLR, weighted linear regression.

In the evidence base 22 out of 39 contrasts were from the trials which were conducted before the approval of ipilimumab as the first ICI for the treatment of metastatic melanoma. Therefore, to measure the impact of modern subsequent treatments on the strength of correlation, we conducted a sensitivity analysis by restricting the evidence base to the trials that were published after 2012. Regardless of the approach (BRMA or WLR), compared with the primary analysis including all trials in the evidence base, in this sensitivity analysis correlation between the treatment effects on RFS and OS was stronger. Specifically, the correlation was estimated as 0.84 (95% CI 0.63—0.93) by the BRMA and 0.77 (95% CI 0.41—0.92) by the WLR. These results indicate that the transformation of metastatic

Table 3. Summary table of results

Analysis set	*N*	Correlation coefficient from BRMA (95% CI)	Correlation coefficient from WLR (95% CI)	Number (%) of aligned contrasts (for the nominal value of OS HR) in LOOCV[a]	Number (%) of aligned contrasts (for the significance of OS HR) in LOOCV[b]	Average absolute deviation[c]	Surrogate threshold effect (based on 800 patients)	Surrogate threshold effect (based on 1000 patients)
Primary analysis	31	0.68 (0.45-0.82)	0.71 (0.42-0.87)	29/31 (93.5)	27/31 (87.1)	0.15	0.66	0.68
Sensitivity analysis 1	22	0.74 (0.48-0.89)	0.65 (0.25-0.86)	21/22 (95.5)	19/22 (86.4)	0.13	0.65	0.68
Sensitivity analysis 2	26	0.81 (0.62-0.91)	0.77 (0.50-0.90)	25/26 (96.2)	23/26 (88.5)	0.11	0.68	0.71
Sensitivity analysis 3	15	0.76 (0.45-0.90)	0.55 (−0.04 to 0.85)	14/15 (93.3)	14/15 (93.3)	0.12	0.42	0.47
Sensitivity analysis 4	17	0.84 (0.63-0.93)	0.77 (0.41-0.92)	16/17 (94.1)	14/17 (82.4)	0.13	0.64	0.66
Meta-regression analysis	28	0.69 (0.46-0.84)	0.72 (0.39-0.89)	25/28 (89.3)	22/28 (78.6)	0.21	0.61 (stage II) 0.63 (stage III)	0.64 (stage II) 0.66 (stage III)

Sensitivity analysis 1: restricted to trials that did not fail the proportionality test. Sensitivity analysis 2: restricted to trials which were published after year 2000. Sensitivity analysis 3: restricted to trials that investigated interferon-α as the experimental treatment. Sensitivity analysis 4: restricted to trials which were published after year 2012. Meta-regression analysis: analysis adjusting for stage II or stage III disease.
BRMA, bivariate random-effects meta-analysis; CI, confidence interval; HR_{OS}, overall survival hazard ratio; LOOCV, leave-one-out cross-validation; *N*, number of contrasts; RCT, randomized, controlled trial; WLR, weighted linear regression.
[a]The number (%) of contrasts for which observed HR_{OS} were captured by their 95% prediction intervals from the model in leave-one-out cross-validation.
[b]The number (%) of contrasts for which the significance of observed HR_{OS} were the same as the significance of the predicted HR_{OS}.
[c]The average absolute difference between the observed and predicted HR_{OS} across the RCTs.

https://doi.org/10.1016/j.iotech.2025.101042

melanoma by novel agents has a favorable impact on the association between RFS and OS. During cross-validation, in 16 out of 17 contrasts, the reported HR_{OS} was captured within the 95% PIs for the predicted HR_{OS} generated by the WLR. In this sensitivity analysis, estimated surrogate threshold effects were only marginally lower (0.64/0.66 for sample size of 800/1000 patients) than their counterparts for the range of sample sizes in the primary analysis (Table 3).

According to the IQWiG criteria, across all analyses and meta-regression, correlations were classified as moderate regardless of the approach. Except for the sensitivity analysis investigating only interferon-α trials, the surrogate threshold effect exhibited only marginal changes across all analyses, and ranged between 0.61 and 0.64 and 0.68 and 0.71 for an RCT with 800 and 1000 patients, respectively (Table 3). Across all sensitivity analyses and the meta-regression, the coverage rate of the observed HR_{OS} within 95% PIs generated by WLR ranged from 89.3% to 96.2% in LOOCV. The range for the alignment between the status of significance of observed and predicted HR_{OS} was 78.6%-88.5%, and the range for the average absolute difference between observed and predicted HR_{OS} was 0.11-0.21.

We measured the predictive performance of the surrogacy equations from WLR particularly for the ICI trials, as the evidence base included predominantly relatively older non-ICI trials investigating outdated standard of care treatments with differing mechanisms of action. From the CheckMate 238, ECOG-ACRIN E1609, and EORTC 18071 trials, there were four contrasts contributing to the performance evaluation (two contrasts from the ECOG-ACRIN E1609 trial). In the LOOCV, while the observed HR_{OS} were within their 95% PIs for all four contrasts, OS benefit was overestimated in three of them. The average over-prediction margin of HR_{OS} was consistently between 0.04 and 0.05 across all analyses. Using the evidence base from Suciu et al.,[12] the previously published meta-analysis on RFS-OS surrogacy in stage II/III melanoma, the surrogacy equation obtained from the WLR using trials' weights as their sample sizes was $\log(HR_{OS}) = 0.03 + 1.01 \times \log(HR_{RFS})$.

In this equation, the slope was statistically significant (95% CI 0.48−1.53) unlike the intercept (95% CI −0.05 to 0.11). Based on the comparison of slopes from the two surrogacy equations, the predictions from the WLR in this study were subject to less uncertainty due to a narrower 95% CI around the slope. The difference in the steepness of the slopes (1.01 versus 0.67) could be due to differences in the number of RCTs, presence or absence of recent ICI trials and distributions of patients across the (sub)-stages of disease in the two evidence bases. Variability in the definition of RFS in terms of inclusion or exclusion of new primary cases of melanoma could have contributed to differences in slopes. Despite differences between the two equations in the steepness of and uncertainty around slopes, they were consistent with each other in terms of significance/insignificance of slope and intercept. The surrogacy equation obtained from Suciu et al.[12] estimated a surrogate threshold effect of 0.86 for RCTs with 800/1000 patients, which was considerably higher and therefore more achievable than the range of its counterparts (0.66/0.68 based on 800/1000 patients) derived from the current evidence base. These differences not only emphasize the need for an updated meta-analysis with a greater focus on stage II melanoma patients, but also point to the conservativeness of the WLR developed from the current evidence base in estimating the RFS benefit threshold required to observe a significant OS benefit.

DISCUSSION

Our analyses showed a statistically and clinically meaningful correlation between HR_{RFS} and HR_{OS} in resected stage II/III melanoma patients receiving adjuvant therapy. Most sensitivity analyses and meta-regression generally confirmed the strength and robustness of the results with respect to potential changes in the evidence base. A stronger correlation between the treatment effects on RFS and OS was observed when the evidence base was restricted to relatively more recent RCTs (post-2000 and post-2012), and results were similar to those of the primary analysis when the evidence base was restricted to RCTs that did not fail the proportional hazards assumption. When the evidence base was restricted to studies investigating interferon-α as experimental therapy, compared with the primary analysis, a stronger correlation was observed from the BRMA but a weaker correlation was observed from the WLR. The conflicting findings from the two models on the restricted evidence base implied no clear impact of the mechanism of action of adjuvant therapy on the RFS-OS association, which necessitates further research solely focusing on trials investigating modern therapies to better understand the evolution of RFS-OS over time. External validation of the primary model against the SWOG 1404 and BRIM8 trials generated highly accurate predictions. While the surrogacy equation obtained from the WLR in primary analysis was comparable to the surrogacy equation generated from a WLR using a relatively narrower evidence base from Suciu et al.,[12] its narrower 95% CI around its slope was indicative of narrower bands of uncertainty and hence more stability on the predictions.

Our study expanded on previously published research evaluating RFS-OS surrogacy in the adjuvant setting[12,13] by encompassing a wider range of treatments, a larger and more recent set of trials including 16 252 patients from 30 trials, compared with 6815 patients from 13 trials in Suciu et al.[12] Despite reporting on a larger set of trials, the current study included only 10[26-34] of the 13 trials analyzed in Suciu et al.[12] due to lack of direct reporting on HRs or KM curves for either OS or RFS from the remaining trials. These three trials analyzed by Suciu et al.,[12] however, were not reflective of the current clinical practice as they were published in 2001 or earlier. Overall, Suciu et al.[12] estimated a stronger correlation ($R^2 = 0.91$) between the treatment effects on RFS and OS than the current study ($R^2 = 0.46$) which can be explained by the differences between the two evidence

bases and potentially with the lack of IPD in our case. Our study had a broader and more heterogeneous set of studies. In particular, 9 of 30 studies in our evidence base recruited stage II patients, compared with only 1 of 13 studies included in Suciu et al.[12] Because of this, the majority of the study population (75%) in Suciu et al.[12] had stage III melanoma. Additionally, more recent studies investigating ICIs (e.g. CheckMate-238, EORTC-18071) and targeted therapies (e.g. COMBI-AD) were absent from the evidence base of Suciu et al.[12] Despite the weaker correlation, the model performance in LOOCV was indicative of the high predictive value of RFS. The moderate and relatively imprecise correlation between the treatment effects on RFS and OS led to wide uncertainty around the OS HR predictions from the model.

One of the key strengths of this study is its up-to-date and large evidence base containing 30 RCTs. Another key strength is the extensive set of sensitivity analyses. Sensitivity analyses based on publication date and restricted to studies of interferon-α investigated the impact of changes in the adjuvant and metastatic treatment landscape on correlation strength. Additionally, from a statistical standpoint, an important aspect of this study which differentiates it from prior studies on the subject is the rigorous assessment of the violation of proportional hazards assumption on the strength of correlation. Because the input data to meta-analyses were mostly reported HRs, calculation of which relies on the proportional hazards assumption, digitization of KM curves from the publications and evaluation of the appropriateness of using HRs for relative treatment effects aimed to eliminate a potential source of bias due to violation of proportional hazards assumption. In fact, the proportional hazards assumption was violated in 10 out of 39 contrasts, hence a sensitivity analysis on the subset of RCTs not violating this assumption was essential. The digitization process also allowed the maximum utilization of available data, as for 11 of the 39 contrasts HRs were not directly reported from the RCTs for at least one endpoint. The sensitivity analyses and a meta-regression accounted for key factors that might affect the strength of the correlation. Moreover, employing two different approaches (BRMA and WLR) identified the impact of methodology on the correlation. The BRMA relaxed the need to know within-study correlations across the evidence base. Unlike WLR it also accounted for the uncertainty around the treatment effects on RFS. The WLR and its results were easier to visualize and interpret but the uncertainty around the treatment effects on RFS was not considered in WLR.

Despite their common use, the utility of surrogate endpoints and their predictive ability are often debated due to lack of reported OS benefit in modern adjuvant trials.[2] A key factor influencing the predictive power of RFS in early-stage melanoma is the advent of effective subsequent treatments, including ICIs and targeted therapies. The level of exposure to novel treatments in advanced melanoma may have confounding effects on the OS benefit of adjuvant treatments. Additionally, OS may take extended time to manifest its benefits, particularly when recurrences occur long after adjuvant therapy or when deaths occur long after the recurrence. Despite these challenges in predicting OS benefit, the estimated surrogate threshold effects can aid trial design by informing sample size and statistical power calculations. The surrogacy equation can also assist economic evaluations by generating long-term OS projections for an emerging treatment. This can be achieved by applying the predicted HR_{OS} between the emerging treatment versus a pre-specified reference treatment to the long-term OS projections of the reference treatment. Similarly, the 95% PI generated for the HR_{OS} can be used to generate a 95% confidence band around the estimated OS curve.

The evidence base and the methodological approaches employed in this study had some limitations. First, without the IPD from the RCTs, the scope of this research was limited to correlation between treatment effects and the prognostic role of RFS on OS at the individual level was not studied. Additionally, this constrained the extent of the aggregate-level analysis where the BRMA model distinguishing within-study correlations from between-study correlation could be employed to enhance the accuracy of the correlation estimates between the treatment effects on the endpoints. The lack of IPD also limited the ability to assess the plausibility of randomization sequences, verify data integrity and consistency, standardize the definitions of endpoints, investigate the impact of the length of follow-up on the correlation, carry out sensitivity analyses with alternative endpoint definitions, and assess the impact of subsequent treatment distributions on the strength of correlation.

Second, standard surrogacy approaches in the literature including BRMA and WLR do not account for the impact of subsequent treatments on the surrogacy relationship. In addition to this methodological gap, addressing the impact of subsequent treatment distributions on the correlations was not possible due to lack of reported data from relatively older studies. Third, because most of the RCTs in the evidence base were conducted before the advent of modern ICIs, they are not fully reflective of recent changes in diagnosis and management of resected melanoma, and changes in treatment landscape for metastatic melanoma. In particular, many of the studies investigated interferon-α, which is no longer the standard adjuvant therapy in melanoma. Therefore, predictions for future ICI trials from the surrogacy equations of this study should be approached with caution. Fourth, due to limited stage-specific efficacy data from the RCTs, the majority of the evidence base included data from RCTs studying combined stage II/III populations in all analyses except for the meta-regression. Therefore, predictions from the surrogacy equations for an RCT that has an entirely stage II or entirely stage III population may require further adjustments.

Fifth, RCTs in the evidence base had a wide range of publication dates, during which American Joint Committee on Cancer (AJCC) manuals were revised from 5th edition to 8th edition. Therefore, classification of patients with similar tumor characteristics across disease stages may differ among the studies. Likewise, the evidence base covered a

broad range of RCTs with slightly variant RFS definitions, which may not have accounted for non-disease-related deaths (e.g. AVAST-M trial) or be inconsistent with the RFS definition in more recent ICI trials such as CheckMate 76K, which consider new cases of primary melanoma or melanoma *in situ* in the RFS definition. This discrepancy in the RFS definition among the trials may pose a potential source of bias for predictions.

Sixth, despite extensive sensitivity analyses, the estimated surrogate threshold effects from our study may still fall short of the RFS benefit that would translate into statistically significant OS benefit in the current clinical practice due to presence of several novel agents in both adjuvant and metastatic treatment settings. For a given contemporary trial, using a more detailed model with a state-transition structure (e.g. semi-Markov), with the availability of subsequent treatment distributions, it would be possible to derive more accurate threshold RFS benefits required to observe statistically significant OS benefit.

Finally, treatment modalities studied in the experimental arms showed variation across the evidence base. There were 10 studies in the evidence base with an active comparator arm. While pooling data from RCTs investigating different treatment modalities in a CMA has statistical advantages (e.g. increased sample size, more stable predictions) and has precedent applications in other tumor areas,[12,35-37] the magnitude of treatment effect in an RCT depends on the class of treatments in both arms. Therefore, in an ideal setting, it would be clinically more intuitive and preferrable to have the experimental and/or control arms of the studies to be as similar as possible across the evidence base and conclusions about the validity of RFS-OS correlation across different pairs of experimental-control therapies should be approached with caution. Nevertheless, due to rapid evolution of melanoma treatment in the adjuvant setting with ICIs and targeted therapies, over time it would be unrealistic and uncommon to expect trials with placebo control.

Conclusion

Based on a comprehensive and up-to-date evidence base, our CMA suggests that HR_{RFS} may be used as a surrogate predictor for HR_{OS} in resected stage II/III melanoma. The predictive ability of the surrogacy equation derived from WLR, as demonstrated in internal and external validation, highlighted its potential utility in informing trial design and earlier evaluation of future trials. The evidence base in our study, however, primarily stems from older trials that may not fully reflect the recent transformation of adjuvant and metastatic melanoma treatment landscape with ICIs and targeted treatments, and the impact of these treatments on long-term efficacy outcomes. Moreover, as evidenced by our numerical study, the mechanism of action of adjuvant therapy displayed no clear impact on the RFS-OS correlation. Therefore, extrapolation of findings and insights from our study to future settings involving modern therapies should be interpreted with caution.

FUNDING

This work was supported by Bristol Myers Squibb (no grant number).

DISCLOSURE

The authors have declared no conflicts of interest.

DATA SHARING

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

REFERENCES

1. Ernst M, Giubellino A. The current state of treatment and future directions in cutaneous malignant melanoma. *Biomedicines*. 2022;10(4):822.
2. Ascierto PA, Del Vecchio M, Mandalá M, et al. Adjuvant nivolumab versus ipilimumab in resected stage IIIB-C and stage IV melanoma (CheckMate 238): 4-year results from a multicentre, double-blind, randomised, controlled, phase 3 trial. *Lancet Oncol*. 2020;21(11):1465-1477.
3. Eggermont AM, Chiarion-Sileni V, Grob JJ, et al. Adjuvant ipilimumab versus placebo after complete resection of high-risk stage III melanoma (EORTC 18071): a randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2015;16(5):522-530.
4. Eggermont AMM, Blank CU, Mandala M, et al. Adjuvant pembrolizumab versus placebo in resected stage III melanoma. *N Engl J Med*. 2018;378(19):1789-1801.
5. Luke JJ, Rutkowski P, Queirolo P, et al. Pembrolizumab versus placebo as adjuvant therapy in completely resected stage IIB or IIC melanoma (KEYNOTE-716): a randomised, double-blind, phase 3 trial. *Lancet*. 2022;399(10336):1718-1729.
6. Kirkwood JM, Del Vecchio M, Weber J, et al. Adjuvant nivolumab in resected stage IIB/C melanoma: primary results from the randomized, phase 3 CheckMate 76K trial. *Nat Med*. 2023;29(11):2835-2843.
7. Long GV, Hauschild A, Santinami M, et al. Adjuvant dabrafenib plus trametinib in stage III BRAF-mutated melanoma. *N Engl J Med*. 2017;377(19):1813-1823.
8. Driscoll J, Rixe O. Overall survival: still the gold standard: why overall survival remains the definitive end point in cancer clinical trials. *Cancer J*. 2009;15(5):401-405.
9. Eggermont AMM, Kicinski M, Blank CU, et al. Five-year analysis of adjuvant pembrolizumab or placebo in stage III melanoma. *NEJM Evid*. 2022;1(11):EVIDoa2200214.
10. Weber J, Mandala M, Del Vecchio M, et al. Adjuvant nivolumab versus ipilimumab in resected stage III or IV melanoma. *N Engl J Med*. 2017;377(19):1824-1835.
11. Polley M-YC, Lamborn KR, Chang SM, Butowski N, Clarke JL, Prados M. Six-month progression-free survival as an alternative primary efficacy endpoint to overall survival in newly diagnosed glioblastoma patients receiving temozolomide. *Neuro Oncol*. 2009;12(3):274-282.
12. Suciu S, Eggermont AMM, Lorigan P, et al. Relapse-free survival as a surrogate for overall survival in the evaluation of stage II-III melanoma adjuvant therapy. *J Natl Cancer Inst*. 2018;110(1):87-96.
13. Coart E, Suciu S, Squifflet P, et al. Evaluating the potential of relapse-free survival as a surrogate for overall survival in the adjuvant therapy of melanoma with checkpoint inhibitors. *Eur J Cancer*. 2020;137:171-174.
14. Rohatgi A. WebPlotDigitizer Version 4.6. Available at https://automeris.io/WebPlotDigitizer. Accessed May 18, 2023.
15. 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.
16. Therneau TM, Grambsch PM. Testing proportional hazards. In: *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer New York; 2000. p. 127-152.

17. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*. 2008;9(1):172-186.
18. Bujkiewicz S, Achana F, Papanikos T, Riley R, Abrams K. NICE DSU Technical Support document 20: Multivariate Meta-Analysis of Summary Data for Combining Treatment Effects on Correlated Outcomes and Evaluating Surrogate Endpoints. Available at http://www.nicedsu.org.uk. Accessed May 18, 2023.
19. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat*. 2006;5(3):173-186.
20. Armitage P, Berry G, Matthews J. *Statistical Methods in Medical Research*. 4th ed. Malden, MA: Wiley-Blackwell; 2001.
21. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc*. 1983;78(382):316-331.
22. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*. 2000;1(1):49-67.
23. Grossmann KF, Othus M, Patel SP, et al. Adjuvant pembrolizumab versus IFNα2b or ipilimumab in resected high-risk melanoma. *Cancer Discov*. 2022;12(3):644-653.
24. Maio M, Lewis K, Demidov L, et al. Adjuvant vemurafenib in resected, $BRAF^{V600}$ mutation-positive melanoma (BRIM8): a randomised, double-blind, placebo-controlled, multicentre, phase 3 trial. *Lancet Oncol*. 2018;19(4):510-520.
25. Institute for Quality and Efficiency in Health Care: Executive Summaries. Validity of Surrogate Endpoints in Oncology: Executive Summary of Rapid Report A10-05, Version 1.1. Available at https://www.ncbi.nlm.nih.gov/books/NBK198799/. Accessed May 18, 2023.
26. Cameron DA, Cornbleet MC, Mackie RM, et al. Adjuvant interferon alpha 2b in high risk melanoma - the Scottish study. *Br J Cancer*. 2001;84(9):1146-1149.
27. Cascinelli N, Belli F, MacKie RM, Santinami M, Bufalino R, Morabito A. Effect of long-term adjuvant therapy with interferon alpha-2a in patients with regional node metastases from cutaneous melanoma: a randomised trial. *Lancet*. 2001;358(9285):866-869.
28. Creagan ET, Dalton RJ, Ahmann DL, et al. Randomized, surgical adjuvant clinical trial of recombinant interferon alfa-2a in selected patients with malignant melanoma. *J Clin Oncol*. 1995;13(11):2776-2783.
29. Garbe C, Radny P, Linse R, et al. Adjuvant low-dose interferon {alpha}2a with or without dacarbazine compared with surgery alone: a prospective-randomized phase III DeCOG trial in melanoma patients with regional lymph node metastasis. *Ann Oncol*. 2008;19(6):1195-1201.
30. Hancock BW, Wheatley K, Harris S, et al. Adjuvant interferon in high-risk melanoma: the AIM HIGH Study—United Kingdom Coordinating Committee on Cancer Research randomized study of adjuvant low-dose extended-duration interferon alfa-2a in high-risk resected malignant melanoma. *J Clin Oncol*. 2004;22(1):53-61.
31. Kirkwood JM, Ibrahim JG, Sondak VK, et al. High- and low-dose interferon alfa-2b in high-risk melanoma: first analysis of intergroup trial E1690/S9111/C9190. *J Clin Oncol*. 2000;18(12):2444-2458.
32. Eggermont AMM, Suciu S, MacKie R, et al. Post-surgery adjuvant therapy with intermediate doses of interferon alfa 2b versus observation in patients with stage IIb/III melanoma (EORTC 18952): randomised controlled trial. *Lancet*. 2005;366(9492):1189-1196.
33. Kleeberg UR, Suciu S, Bröcker EB, et al. Final results of the EORTC 18871/DKG 80-1 randomised phase III trial. rIFN-alpha2b versus rIFN-gamma versus ISCADOR M versus observation after surgery in melanoma patients with either high-risk primary (thickness >3 mm) or regional lymph node metastasis. *Eur J Cancer*. 2004;40(3):390-402.
34. Eggermont AM, Suciu S, Santinami M, et al. Adjuvant therapy with pegylated interferon alfa-2b versus observation alone in resected stage III melanoma: final results of EORTC 18991, a randomised phase III trial. *Lancet*. 2008;372(9633):117-126.
35. Ajani JA, Leung L, Singh P, et al. Disease-free survival as a surrogate endpoint for overall survival in adults with resectable esophageal or gastroesophageal junction cancer: a correlation meta-analysis. *Eur J Cancer*. 2022;170:119-130.
36. Leung L, Chou E, Kurt M, et al. POSA24 progression-free survival (PFS) as a surrogate endpoint for overall survival (OS) in previously untreated advanced melanoma: a correlation meta-analysis of randomized controlled trials (RCTS). *Value Health*. 2022;25(1):S22.
37. Leung L, Kanters S, Pourrahmat MM, et al. CO205 evaluating disease-free survival (DFS) as a surrogate endpoint (SE) for overall survival (OS) in muscle-invasive urothelial carcinoma (MIUC): an analysis of surveillance, epidemiology, and end results (SEER)-Medicare data. *Value Health*. 2023;26(6):S54.