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)

Leung L, MSc[1]; Chou E, MSc[2]; Kurt M, PhD[2]; Moshyk A, MSc[2]; Hamilton M, MPH[2]; Pourrahmat M, BSc[1]; Kanters S, PhD[1]

[1]Evidinno Outcomes Research Inc., Vancouver, BC, Canada; [2]Bristol Myers Squibb, Princeton, NJ, USA

Introduction

- Melanoma is the fifth most common cancer in the United States (US). In 2021, an estimated 106,110 new cases of melanoma will have been diagnosed in the US, and 7,180 American patients with melanoma are estimated to die.[1]
- Many factors influence the prognosis of patients with metastatic melanoma, such as the site of metastasis[2]—patients with brain metastases have the worst prognosis with estimated median survival times ranging between 4.9 months and 34.1 months.[3]
- The three standard classes of FDA-approved first-line (1L) treatments include BRAF-targeted therapies, anti-programmed cell death protein 1 (PD-1) monotherapy and combination of anti-PD-1 plus anti-CTLA-4 therapy.[2] Phase III randomized controlled trials (RCTs) have demonstrated objective response rates of up to 67% and 3-year overall survival (OS) in up to 58% of patients.[2]
- OS is generally the gold standard endpoint for oncology trials. However, observing a benefit on OS may require considerable time for patient follow-up.
- Evaluation and validation of earlier endpoints such as progression-free survival (PFS) as surrogates of OS may support research necessary to achieve timely market access for novel, life-extending drugs by reducing the clinical development time while enabling greater statistical power for efficacy measurements without being influenced by changing treatment landscapes in subsequent lines.
- PFS is listed on the FDA *Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure* for melanoma,[4] and prior research has been conducted to assess the validity of PFS and other surrogate endpoints (SEs) in early and late-stage melanoma (for a comprehensive literature review on published surrogacy analyses in advanced melanoma please see Poster POSC24 at ISPOR EU 2021 Conference). Some examples of this prior research include:
  - Flaherty et al 2014, which reported PFS to be a robust surrogate for OS in dacarbazine-controlled trials of metastatic melanoma.[5]
  - Nie et al 2020, which proposed that PFS may be an appropriate surrogate for OS in anti-PD-1/PD-L1 trials of metastatic melanoma.[6]
- Although the surrogacy of PFS for OS in melanoma has been previously assessed,[5,6] researchers have mainly taken an agnostic approach to treatment line. Hence, there is a gap of analyses assessing PFS-OS relationship among treatment-naive (i.e., receiving 1L therapies) advanced melanoma patients in the literature. Therefore, we sought to evaluate the relationship between PFS and OS in this population.

Objective

- To evaluate PFS as an SE for OS by modeling the association between the treatment effects, conventionally measured or calculated as hazard ratios (HRs) on each endpoint, using RCTs investigating 1L therapies for the treatment of advanced melanoma.

Methods

Systematic literature review (SLR)

- RCTs included in this study were identified through an SLR conducted using methodology adapted from the Cochrane Handbook for Systematic Reviews of Interventions.[7]
- Study selection was guided by a PICO framework summarized as:
  - Population: Adult (≥18 years of age) patients with previously untreated advanced, unresectable stage III or IV melanoma
  - Interventions and Comparators: Any monotherapy or combination with other treatments provided in the 1L treatment setting (e.g., anti-CTLA-4 therapies, BRAF inhibitors, PD-1 and PD-L1 inhibitors)
  - Outcomes: OS and PFS
- To be eligible in the surrogacy analyses, included RCTs must have reported both HRs for OS (HR_{OS}) and PFS (HR_{PFS}), and must have met the proportional hazards (PH) assumption.
- Relevant studies were identified by searching the MEDLINE®, Embase, and CENTRAL databases up to October 19, 2020 using predefined search strategies via the Ovid platform. The searches were limited to English language studies, and no publication date limits were applied.
- Searches for grey literature included conference proceedings between 2018-2020. The list of searched conferences included:
  - American Society of Clinical Oncology (ASCO)
  - European Society for Medical Oncology (ESMO)
  - Society for Immunotherapy of Cancer (SITC)
  - Society of Melanoma Research (SMR)
  - American Association for Cancer Research (AACR)

Data analysis

Trial-level surrogacy models and analysis sets

- The validity of an SE is typically studied in two major steps[8]: Analysis of the correlation of endpoints themselves, which is often referred to as individual level surrogacy; and analysis of correlation of treatment effects, which is often referred to as trial-level surrogacy.
- The surrogacy of PFS for OS at the trial-level was assessed using two meta-analysis models.
- Logarithmic transformations on HRs were conducted to linearize the HRs to be consistent with the linearity assumption for the relationship between the treatment effects.
- The first model was based on a bivariate random-effects meta-analysis (BRMA) model proposed by Riley et al 2007[9], which provides an overall correlation measure between the log-transformed HRs of PFS and OS (i.e., $\log\text{-}HR_{PFS}$ and $\log\text{-}HR_{OS}$). This alternative modeling framework is similar to the general BRMA[10] but assumes one overall correlation parameter rather than two separate parameters accounting for within-study and between-study correlations.
- The second model was a weighted linear regression (WLR) model using sample size of treatment arms in each comparison. The association between $\log\text{-}HR_{PFS}$ and $\log\text{-}HR_{OS}$ was measured by the Pearson correlation coefficient based on a linear surrogate equation.
- Sensitivity analyses were conducted by restricting the evidence base according to the phases of the trials, use of immune checkpoint inhibitors (ICIs) in the trials, or the status and permission of crossover within the trial. Specifically, the sets of clinical trials in the evidence base were re-classified below to conduct the sensitivity analyses:
  - Phase II and III ICI trials (this analysis set was explored as immunotherapy is one of the most efficient therapeutic strategies against melanoma[11])
  - Phase III trials
  - Trials with no crossover permitted or trials that provided adjusted estimates due to crossover (referred to as "Crossover analysis I" in this poster)
  - Trials in Crossover analysis I set, plus trials with unreported crossover status or descriptions (referred to as "Crossover analysis II" in this poster) *

* Note that the term 'crossover' in this analysis is defined as patients permitted to switch between treatment arms during the study period.

Assessing the surrogacy equation and the correlation estimates

- The German Institute of Quality and Efficiency in Health Care (IQWiG) guidelines were used to assess the strength of the correlation estimates[12]:
  - High correlation is achieved when the lower limit of the 95% confidence interval (CI) for correlation is ≥0.85
  - Low correlation is achieved when the upper limit of the 95% CI for correlation is ≤0.7.
  - Otherwise, the correlation is considered moderate
- The validity of the model was assessed by using a leave-one-out cross-validation (LOOCV) approach based on WLR models, whereby the observed HR_{OS} was compared to the default 95% prediction interval (PI) from the model. Based on the WLR models, prediction accuracies on the statistical significance of HR_{OS} as a binary outcome were also reported for each analysis.
- The National Institute for Health and Care Excellence (NICE) Decision Support Unit Technical Support Document 20 was used as a guide to assess model validity, whereby at least 95% of the trials in the respective evidence base should have their reported HR_{OS} to be within the 95% PI of the predicted HR_{OS} (i.e., coverage is ≥95%).[13]

Results

- Out of 11,015 records that were identified by the main database search and grey literature searches, 70 publications pertaining to 34 unique RCTs were included in the SLR.
- Of the 34 RCTs, 23 were eligible to be included in the surrogacy analysis after satisfying two criteria: reporting both HR_{PFS} and HR_{OS} estimates (and/or KM curves to enable the HR calculation through a Cox PH model), and meeting the PH assumption.
- List of included trials in each analysis set is provided in **Table 1**, and results from each analysis set is presented in **Table 2**.

Full analysis set

- The full analysis set included 23 studies and 27 comparisons (CheckMate 067 and PACMEL trials provided two treatment comparison estimates, and COLUMBUS provided three comparison estimates).
- Analyzing the results from WLR, three contradicting data points were identified (i.e., the directions of OS estimates in terms of pointing statistical significance were opposite from those of PFS): one from KEYNOTE-029 trial, and two from PACMEL trial.
- Using LOOCV, observed HR_{OS} in two (KEYNOTE-029 and NEMO) out of 27 comparisons laid outside of the model-predicted 95% PI.
- The results of WLR model and LOOCV on the full analysis set are presented in **Figure 1** and **Figure 2**, respectively.

Phase II and III ICI trials

- This analysis included five studies, all of which were trials with immunotherapy arms.
- There were six comparisons in total. Using LOOCV, observed HR_{OS} in all comparisons laid within the model-predicted 95% PI.

Phase III trials

- This analysis was on phase III trials in the full analysis set. A total of 20 comparisons from 17 studies were included.
- Using LOOCV, observed HR_{OS} in two (BRIM-3 and NEMO) out of 20 comparisons laid outside the model-predicted 95% PI.

Crossover analysis I

- This analysis included seven trials allowing no crossover or providing adjusted crossover estimates.
- There were seven comparisons in total. Using LOOCV, observed HR_{OS} in all comparisons laid within the model-predicted 95% PI.

Crossover analysis II

- The analysis included trials with unclear crossover status or descriptions, trials with no crossover permitted, or trials that provided estimates adjusted for crossover. A total of 19 studies with 23 comparisons were included.
- Correlation measures in crossover analysis II were similar to those in the full analysis set, and the WLR figure identified one data point from KEYNOTE-029 and the two data points from PACMEL to have contradicting treatment effect directions between PFS and OS.
- Using LOOCV, observed HR_{OS} in one (NEMO) out of 23 comparisons laid outside of the model-predicted 95% PI.

Table 1. List of included trials in each analysis set

Analysis set	Number of included studies	Included studies
Full analysis set	23	Algazi 2020, Ascierto 2017, Avril 2004, BREAK-3, BRIM-3, CheckMate 066, CheckMate 067, CheckMate 069, CheckMate 511, coBRIM, COLUMBUS, COMBI-d, COMBI-v, IMspire150, KEYNOTE-006, KEYNOTE-022, Lebbe 2020, KEYNOTE-029, Middleton 2000, NEMO, PACMEL, Patel 2011, Robert 2011
Phase II and III ICI trials	5	CheckMate 066, CheckMate 067, CheckMate 069, CheckMate 511, KEYNOTE-006
Phase III	17	Ascierto 2017, Avril 2004, BREAK-3, BRIM-3, CheckMate 066, CheckMate 067, CheckMate 511, coBRIM, COLUMBUS, COMBI-d, COMBI-v, IMspire150, KEYNOTE-006, Middleton 2000, NEMO, Patel 2011, Robert 2011
Crossover analysis I*	7	Algazi 2020, Ascierto 2017, coBRIM, IMspire150, BREAK-3, BRIM-3, CheckMate 069
Crossover analysis II†	19	Algazi 2020, Ascierto 2017, Avril 2004, BREAK-3, BRIM-3, CheckMate 067, CheckMate 069, CheckMate 511, coBRIM, COLUMBUS, IMspire150, KEYNOTE-006, KEYNOTE-022, KEYNOTE-029, Middleton 2000, NEMO, PACMEL, Patel 2011, Robert 2011

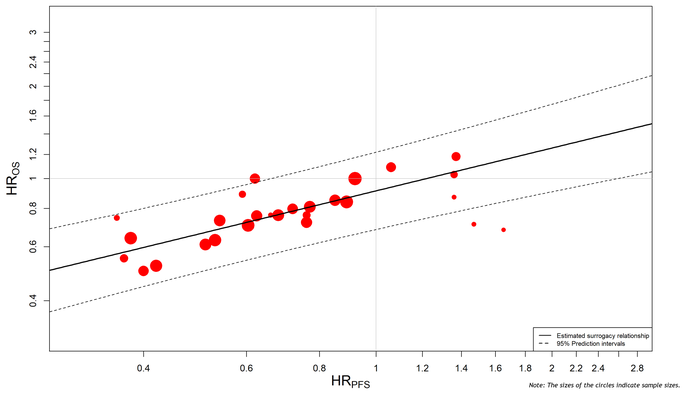
*Crossover analysis I: Trials with no crossover permitted or trials that provided adjusted crossover estimates
†Crossover analysis II: Trials with non-informative crossover descriptions, trials with no crossover permitted, or trials that provided adjusted crossover estimates
ICI: immune checkpoint inhibitor

Table 2. Summary of results

Analysis set	Number of included studies	R_{BRMA} (95% CI)	IQWiG correlation strength for BRMA	R_{WLR} (95% CI)	IQWiG correlation strength for WLR	LOOCV (% Validated)	LOOCV on the accuracy on HR_{OS} significance‡
Primary analysis							
Full analysis set	23	0.72 (0.50, 0.85)	Moderate	0.79 (0.59, 0.91)	Moderate	25/27 (93%)	20/27 (74%)
Sensitivity analysis							
Phase II and III ICI trials	5	0.80 (0.28, 0.96)	Moderate	0.89 (-0.23, 1.00)	Moderate	6/6 (100%)	3/6 (50%)
Phase III	17	0.88 (0.71, 0.95)		0.89 (0.75, 0.97)		18/20 (90%)	14/20 (70%)
Crossover analysis I*	7	0.87 (0.45, 0.98)		0.94 (0.80, 1.00)		7/7 (100%)	4/7 (57%)
Crossover analysis II†	19	0.71 (0.45, 0.86)		0.78 (0.54, 0.91)		22/23 (96%)	17/23 (74%)

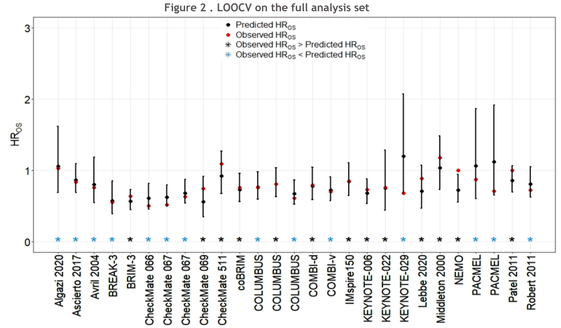
*Crossover analysis I: Trials with no crossover permitted or trials that provided adjusted crossover estimates
†Crossover analysis II: Trials with non-informative crossover descriptions, trials with no crossover permitted, or trials that provided adjusted crossover estimates
‡Accuracy is defined as the proportion of the HR_{OS} significance correctly predicted by the model, out of all predicted HR_{OS}.
BRMA: Bivariate random-effects meta-analysis, HR: Hazard ratio, ICI: immune checkpoint inhibitor, IQWiG: Institute for Quality and Efficiency in Health Care, LOOCV: leave-one-out cross-validation, OS: Overall survival, R_{BRMA}: Pearson correlation estimate based on BRMA, R_{WLR}: Pearson correlation estimate based on WLR, WLR: Weighted linear regression.

Figure 1. WLR model for the full analysis set



Note: The sizes of the circles indicate sample sizes.

Figure 2 . LOOCV on the full analysis set



Note: The black vertical lines indicate the 95% PI of HR_{OS}. trial names are listed below the horizontal axis.

Conclusions

- Results suggest a moderate correlation per IQWiG recommendations, between PFS and OS across all analyses in the 1L advanced melanoma setting.
- Although the full analysis set did not meet the 95% LOOCV coverage threshold as per NICE's recommendation for validity, the sensitivity analyses restricted to phase II and III ICI trials, Crossover analysis I, and Crossover analysis II did meet this criterion.
- These results also provide two additional observations:
  - Sensitivity analysis restricted to immunotherapies provide a greater correlation estimate compared to that obtained from full analysis set. Although limited by the small number of trials in this set (n = 5), this work will further encourage similar analyses once more trials become available in the public domain.
  - The restriction of the evidence base to trials adjusting for crossover led to results aligning with claims from previous surrogacy work in advanced melanoma.[5]

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA: A Cancer Journal for Clinicians. 2021;71(1):7-33.
2. Gibney GT, Atkins MB. Choice of first-line therapy in metastatic melanoma. Cancer. 2019;125(5):666-669.
3. Sperduto PW, Jiang W, Brown PD, et al. Estimating survival in melanoma patients with brain metastases: an update of the graded prognostic assessment for melanoma using molecular markers (Melanoma-molGPA). International Journal of Radiation Oncology* Biology* Physics. 2017;99(4):812-816.
4. FDA. Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure Accessed 12 October 2021.
5. Flaherty KT, Hennig M, Lee SJ, et al. Surrogate endpoints for overall survival in metastatic melanoma: a meta-analysis of randomised controlled trials. The Lancet Oncology. 2014;15(3):297-304.
6. Nie RC, Yuan SQ, Wang Y, et al. Surrogate endpoints for overall survival in anti-programmed death-1 and anti-programmed death ligand 1 trials of advanced melanoma. Therapeutic Advances in Medical Oncology. 2020;12(no pagination).
7. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. Vol 4: John Wiley & Sons; 2011.
8. Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of Surrogate End Points in Multiple Randomized Clinical Trials with Failure Time End Points. Journal of the Royal Statistical Society Series C (Applied Statistics). 2001;50(4):405-422.
9. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. BMC medical research methodology. 2007;7:3.
10. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. Statistics in medicine. 1997;16(17):1965-1982.
11. Ralli M, Botticelli A, Visconti IC, et al. Immunotherapy in the Treatment of Metastatic Melanoma: Current Knowledge and Future Directions. Journal of immunology research. 2020;2020:9235638.
12. Validity of surrogate endpoints in oncology: Executive summary of rapid report A10-05, Version 1.1. Institute for Quality and Efficiency in Health Care: Executive Summaries. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG); 2005.
13. Bujkiewicz, S., Achana, F., Papanikos, T., Riley, R.D., Abrams, K.R. NICE DSU Technical Support Document 20: Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints. 2019; available from http://www.nicedsu.org.uk.

Acknowledgments

- This study was funded by Bristol Myers Squibb (Princeton, NJ, USA) and conducted by Evidinno Outcomes Research Inc. (Vancouver, BC, Canada).
- Engels Chou, Murat Kurt, Andriy Moshyk, and Melissa Hamilton report employment with Bristol Myers Squibb. Lisa Leung and Mir-Masoud Pourrahmat report employment, and Steve Kanters reports contract, with Evidinno Outcomes Research Inc. Authors report no other conflicts of interest.
- All authors contributed to and approved the presentation; poster development was supported by Andi Qian (Evidinno Outcomes Research Inc.).

Email: lleung@evidinno.com