

ORIGINAL RESEARCH

Time to next treatment or death as a candidate surrogate endpoint for overall survival in advanced melanoma patients treated with immune checkpoint inhibitors: an insight from the phase III CheckMate 067 trial

S. Branchoux^{1*}, C. L. Sofeu², A.-F. Gaudin¹, M. Kurt³, A. Moshyk³, A. Italiano⁴, C. Bellera⁵ & V. Rondeau²

¹Health Economics and Outcomes Research, Bristol Myers Squibb, Rueil-Malmaison; ²Biostatistic Team, Bordeaux Population Health Center, ISPED, Centre INSERM U1219, INSERM, Bordeaux, France; ³Health Economics and Outcomes Research, Bristol Myers Squibb, Princeton, USA; ⁴Department of Early Phase Trial Unit, Institut Bergonié Comprehensive Cancer Centre, Bordeaux; ⁵Epicene Team (Cancer and Environment), Bordeaux Population Health Center, ISPED, Centre INSERM U1219, INSERM, Bordeaux, France



Available online xxx

Background: Time to next treatment or death (TNT-D) may be a patient-relevant endpoint in patients treated with immune checkpoint inhibitors. This study investigated TNT-D as a surrogate endpoint (SE) for overall survival (OS) in previously untreated advanced melanoma patients.

Methods: Patient-level data from the 60-month results of the CheckMate 067 randomised, controlled trial were used. Analyses were carried out for nivolumab monotherapy or nivolumab with ipilimumab versus ipilimumab monotherapy. The SE 1-step validation method based on a joint frailty-copula model was used where the country of enrolment was applied to define clusters. Kendall's τ and the coefficient of determination (R^2_{trial}) were estimated for respective measurements of association at the individual and cluster levels. The surrogate threshold effect, the maximum threshold hazard ratio for TNT-D that would translate into OS benefit, was estimated. A leave-one-out cross-validation analysis was carried out to evaluate model robustness.

Results: Fifteen clusters of data were generated from 945 patients. For both nivolumab-containing arms, the association between TNT-D and OS was deemed acceptable at the individual level (Kendall's $\tau > 0.60$) and strong at the cluster level, with R^2_{trial} fairly close to 1, with narrow confidence intervals. The estimated surrogate threshold effects were 0.61 for nivolumab versus ipilimumab and 0.49 for nivolumab + ipilimumab versus ipilimumab. Cross-validation results showed minimum variation of the correlation measures and satisfactory predictive accuracy for the model.

Conclusion: Results suggest that TNT-D may be a valuable SE in previously untreated advanced melanoma patients treated with immune checkpoint inhibitors. Surrogacy analyses considering multiple randomised controlled trials are warranted for confirming these findings.

Key words: immune checkpoint inhibitors, nivolumab, ipilimumab, surrogate endpoint, overall survival, time to next treatment, advanced melanoma

INTRODUCTION

Immune checkpoint inhibitors (ICIs) are among the most established and clinically promising therapies in oncology.¹ ICIs are indicated as monotherapy or in combination with other systemic treatments in several types of cancer at different stages, such as locally advanced or metastatic, and in different treatment settings (adjuvant, palliative), and their clinical development is expanding. For some

advanced-stage cancers, such as melanoma and renal cell carcinoma (aRCC), a current treatment strategy is to combine different ICIs targeting two distinct receptors, e.g. nivolumab (anti-programmed cell death protein 1) and ipilimumab (anti-Cytotoxic T-lymphocyte antigen-4).^{2,3}

In oncology, several clinical endpoints have been used to assess treatment efficacy.^{4,5} Overall survival (OS), defined as the time from randomisation to the date of death (any cause), is frequently considered the gold standard for assessing treatment efficacy in randomised, controlled trials (RCTs) of anticancer therapies.⁶⁻⁸ This endpoint may require more extensive follow-up and a greater number of patients to demonstrate a clinically and statistically significantly relevant difference when comparing therapeutic strategies, which leads to higher budget requirements due to the

*Correspondence to: Dr. Sébastien Branchoux, Laboratoire Bristol Myers Squibb, 3, rue Joseph Monier, 92500 Rueil-Malmaison, France. Tel: +33-1-58-83-61-11

E-mail: Sebastien.branchoux@bms.com (S. Branchoux).

2059-7029/© 2021 The Author(s). Published by Elsevier Ltd on behalf of European Society for Medical Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

increased cost of surveillance. In this context, evaluation and validation of surrogate endpoints (SEs) are particularly important to expedite patients' access to innovative and potentially life-extending medicines. An SE is defined as 'a biomarker or an intermediate endpoint intended to substitute for a clinical endpoint'⁴; its validity depends on both the mechanism of action of the treatment and the disease setting.⁹

A recent systematic literature review summarised the current evidence base on clinical alternative endpoints associated with OS in ICI-treated patients.¹⁰ Current published evidence is inconclusive regarding validated SEs for OS in ICI-treated patients ($N = 24$). Traditional alternative endpoints such as progression-free survival (PFS) and overall response rate based on conventional response criteria (RECIST 1.1) were weakly correlated with OS in this specific population, possibly because of delayed treatment effect, pseudo-progression and the medical imaging technique used to monitor disease progression.¹¹⁻¹⁵

In haematologic malignancies and metastatic soft tissue sarcomas, time to next treatment or death (TNT-D) has emerged as a relevant alternative clinical endpoint.¹⁶ TNT-D is defined as the time between baseline (randomisation, inclusion or treatment initiation) and the date of next subsequent systemic treatment initiation or death, whichever occurs first. Only a handful of studies, reporting on advanced melanoma, aRCC and non-small-cell lung cancer, describe TNT-D in ICI-treated patients.¹⁷⁻²⁰ One attractive characteristic of TNT-D is its ability to capture the treatment-free interval, defined as the time from the end of index therapy to the date of initiation of a subsequent line of treatment or death, whichever occurs first.²¹ The treatment-free interval has been described in ICI-treated advanced melanoma^{21,22} and, in the American Society of Clinical Oncology value framework, is considered an integral component of a patient's net health benefit.²³ Although TNT-D appears to be a relevant clinical endpoint for ICI-treated patients, the systematic literature review points out that data on surrogacy analyses based on this alternative endpoint have yet to be studied.¹⁰ Our objective was thus to estimate TNT-D and formally assess the surrogate properties of TNT-D in previously untreated advanced melanoma patients treated with ICI.

METHODS

Data

Individual patient-level data (IPD) from the CheckMate 067 trial (ClinicalTrials.gov identifier: NCT01844505) were available for this *post hoc* study. Design, population characteristics and outcomes have been described in detail previously.^{2,24} Briefly, CheckMate 067 is a phase III, randomised, double-blind trial of nivolumab monotherapy or nivolumab plus ipilimumab combination versus ipilimumab monotherapy in subjects with previously untreated unresectable or metastatic melanoma. Randomisation was stratified according to tumour programmed death-ligand 1 status (positive versus negative or indeterminate), *BRAF*

mutation status (V600 mutation-positive versus wild-type) and American Joint Committee on Cancer metastasis stage (M0, M1a or M1b versus M1c).

Definition of efficacy outcomes

OS was defined as the time from the date of randomisation to the date of death (any cause). TNT-D was defined as the time between the date of randomisation and the date of subsequent systemic treatment initiation, or the date of death (any cause), whichever occurred first. Both outcomes were censored on the last date a subject was known to be alive.

Statistical analyses

OS and TNT-D distributions were generated using the Kaplan–Meier estimator. Median follow-up was estimated by the reverse Kaplan–Meier method. To describe these efficacy outcomes, hazard ratios (HRs) were estimated using a Cox proportional-hazards model stratified according to the factors used in the randomisation process. For the SE analysis, HRs were estimated with a joint frailty-copula model.

Robust assessment and validation of an SE require evaluation at both the individual and trial levels (meta-analytic approach).⁹ For two time-to-event endpoints, several meta-analytic surrogacy validation approaches are available, including 1-step and 2-step validation approaches.²⁵⁻²⁹ We considered the SE 1-step validation statistical method based on a joint frailty-copula model (Gumbel copula) as it has been demonstrated to reduce substantially numerical problems encountered with the 2-step method.²⁹ In this approach, Kendall's τ is used to estimate the strength of association between the candidate SE and the final endpoint at the individual level. The Kendall's τ is estimated as a function of the copula parameter, where values above an informal threshold of 0.6 are regarded as sufficient for the validity of the SE at the individual level, while it is uncommon to observe a value higher than 0.7.²⁹ The cluster-level association is estimated using the coefficient of determination (R^2_{trial}). The 95% confidence interval (CI) is estimated using the parametric bootstrap method for τ and the delta method for R^2_{trial} , which can result in confidence limits violating the (0-1) interval.²⁵ An R^2_{trial} sufficiently close to 1 is necessary for validating an SE at the cluster level.⁹

The meta-analytical approach for validation of SEs supposes the availability of data from multiple RCTs. When the data are limited to only a single or a few RCTs, one can consider the geographic location of centres as the cluster of analysis, which is common when evaluating potential SEs.³⁰⁻³² We clustered the patients based on country of enrolment resulting in 21 clusters (33% had the minimum 10 patients in each treatment arm). When the number of patients was lower than four per treatment arm per country, we pooled the countries in a region by similarities in their healthcare management (e.g. Finland, Sweden, Norway and Denmark were pooled as the Nordics region).²⁶ In the case of a high

level of surrogacy between TNT-D and OS at both the individual and trial levels, we calculated the surrogate threshold effect (STE), defined as the minimum treatment effect on the SE necessary to predict a significant treatment effect on the final endpoint.³³ As a sensitivity analysis, we carried out a leave-one-out cross-validation analysis and considered an alternative clustering to assess the robustness of predictive performance and the outputs of the model (e.g. Kendall's τ , R^2_{trial} and the STE).

We report results following the ReSEEM guidelines for reporting of SE evaluation.³⁴ All analyses were carried out using R software v3.6.1 and version 3.2.0 of the publicly available R package frailtypack.³⁵

RESULTS

Complete patient characteristics and efficacy results of CheckMate 067 have been reported.^{2,24} Fifteen clusters were generated for this study. Median follow-up for OS was 63.3 months. Table 1 summarises efficacy results for outcomes of interest. Median TNT-D was reached approximately three times earlier than median OS for nivolumab monotherapy and ipilimumab [nivolumab: TNT-D = 12.1 months (95% CI 8.9-18.0 months), OS = 36.9 months (28.3-58.7 months); ipilimumab: TNT-D = 6.2 months (5.4-7.4 months), OS = 19.9 months (16.9-24.6 months)] whereas for nivolumab plus ipilimumab combination therapy, median TNT-D was 24.2 months (95% CI 16.0-43.9 months), and median OS was not reached. Supplementary Table S1, available at <https://doi.org/10.1016/j.esmoop.2021.100340> shows the distribution of the events defined as TNT-D according to the treatment arm. Figure 1 presents the Kaplan–Meier curves for TNT-D. The Kaplan–Meier survival estimates of 5-year TNT-D in the nivolumab plus ipilimumab, nivolumab monotherapy and ipilimumab groups were 42.7%, 32.8% and 10.9%, respectively.

| | Experimental arm nivolumab (n = 316) | Experimental arm nivolumab + ipilimumab (n = 314) | Control arm ipilimumab (n = 315) |
|------------------------------------|--|--|--|
| Overall survival | | | |
| Median, months (95% CI) | 36.9 (28.2-58.7) | NR (38.2-NR) | 19.9 (16.8-24.6) |
| Events, n (%) | 176 (55.7) | 152 (48.4) | 230 (73.0) |
| Hazard ratio (95% CI) ^a | 0.63 (0.52-0.76) | 0.52 (0.42-0.64) | |
| Time to next treatment or death | | | |
| Median, months (95% CI) | 12.1 (8.9-18.0) | 24.2 (16.0-43.9) | 6.2 (5.4-7.4) |
| Events, n (%) | 210 (66.5) | 180 (57.3) | 275 (87.3) |
| Hazard ratio (95% CI) ^a | 0.55 (0.46-0.65) | 0.42 (0.34-0.50) | |

Overall survival outcomes from Larkin et al.²⁴

CI, confidence interval, NR, not reached.

^a Cox proportional-hazards model stratified according to tumour PD-L1 status, BRAF mutation status and American Joint Committee on Cancer metastasis stage.

Estimated parameters following surrogacy assessment are presented in Table 2. Individual association between TNT-D and OS was deemed acceptable in patients treated with nivolumab monotherapy and in those treated with nivolumab plus ipilimumab (Kendall's $\tau > 0.60$). At trial level, the coefficients of determination were close to 1 with narrow CIs, indicating a strong association between the treatment effects on TNT-D and OS. These were considered acceptable correlations at both the individual and trial levels, and the STEs were assessed (Table 2). The model predicted a significant HR for OS whereas HR for TNT-D was between 0.35 and 0.61, and between 0.12 and 0.49, for nivolumab monotherapy versus ipilimumab, and for nivolumab plus ipilimumab versus ipilimumab, respectively. Figure 2 shows the 95% prediction limits of the predicted treatment effect on OS for both treatments.

Sensitivity analyses

Supplementary Table S2, available at <https://doi.org/10.1016/j.esmoop.2021.100340>, shows results from the leave-one-out cross-validation analyses. Overall, data suggest that correlation measures were influenced minimally when the analyses were repeated with slightly smaller and balanced populations, when the models converged. As such, these sensitivity analyses confirm the robustness of the findings on the association between TNT-D and OS within the dataset available. For the STE, sensitivity analyses showed slight variations while one cluster was removed iteratively. Finally, the ability of the model to accurately predict the treatment effect on OS based on the observed treatment effect on TNT-D in each cluster was satisfactory, although some iterations did not converge (Supplementary Figure S1, available at <https://doi.org/10.1016/j.esmoop.2021.100340>). Results suggest an appropriate prediction of 75% (9/12) in the comparison of nivolumab monotherapy versus ipilimumab. For nivolumab plus ipilimumab versus ipilimumab, the quality of the prediction was slightly improved (8/10).

DISCUSSION

This is the first study to our knowledge to assess the surrogacy properties of TNT-D for OS in ICI-treated patients. We used a strong methodology, since we conducted the surrogacy assessment based on IPD, relied on a robust statistical model for the assessment of the patient- and trial-level associations and reported our findings following international guidelines.³⁴

Our findings indicate that TNT-D may be an appropriate SE for OS in double-blind RCTs of advanced melanoma patients treated with ICIs. These results were confirmed by our sensitivity analyses supporting the overall stability of the individual- and trial-level correlations. For both treatments, the Kendall's τ for individual-level associations was above 0.6, and the estimated R^2_{trial} was close to 1, with a narrow CI. The model predicted a significant HR for OS whereas HR for TNT-D was included in an interval due to the limited number of observations of HR for TNT-D close to 0. Usually,

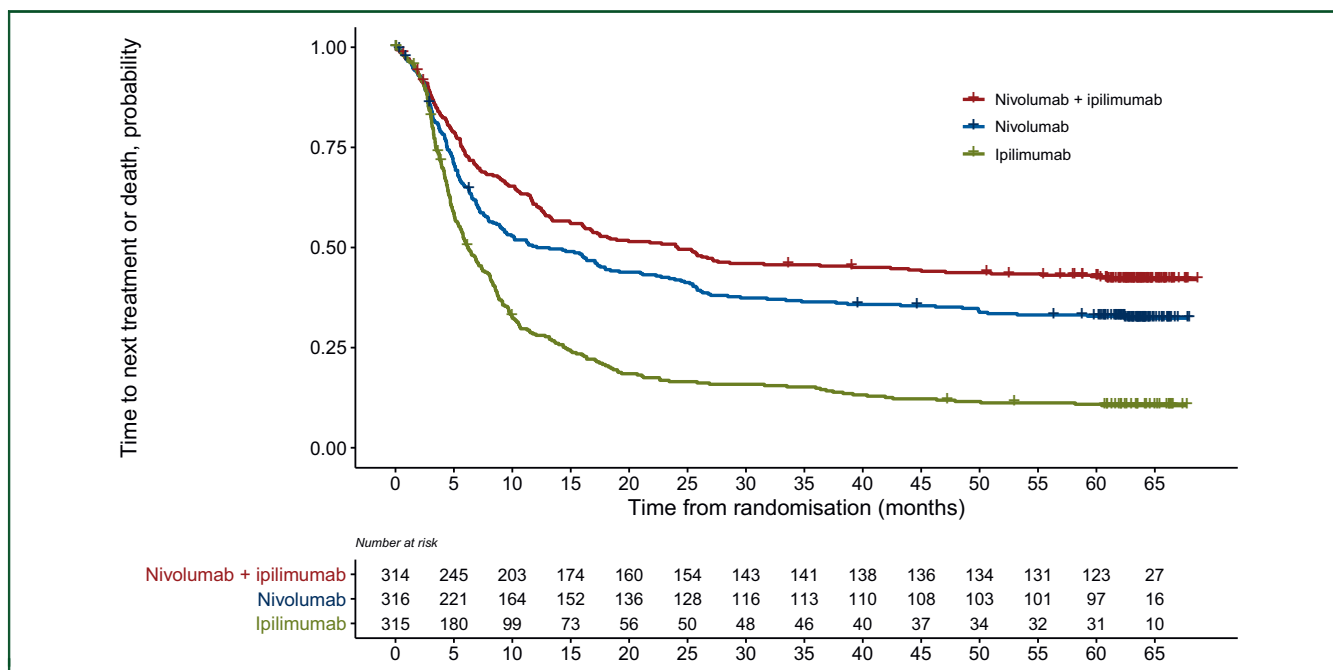


Figure 1. Kaplan–Meier survival curves for time to next treatment or death (TNT-D) for nivolumab plus ipilimumab, nivolumab monotherapy and ipilimumab monotherapy.

+ Signs on the Kaplan–Meier curves represent censorings.

a single value for the STE is provided. It might be relevant to focus only on the upper limit of STE provided, as more the HR for TNT-D will be close to 0, more the HR for OS will be significantly different from 1. Fitting a model based on limited data is known to be challenging, however, and therefore these STE estimates have to be considered cautiously for predicting a significant OS benefit in future RCTs based solely on our results.

TNT-D is a clinically relevant endpoint, and highly correlated with OS in ICI-treated patients with advanced melanoma. It is a pragmatic and often measurable endpoint in all randomised subjects, reaches maturity (i.e. median) earlier than the OS outcome and may provide greater statistical power at the time of analysis. TNT-D reflects the result of a therapeutic medical decision, a change in treatment usually occurring in response to a real

change in the patient status, by integrating the efficacy and toxicity components.^{16,36} Moreover, TNT-D is a comprehensive assessment of the unique outcomes observed with ICIs, as it may take into account the time between interruption of the treatment and the delay to the next systemic treatment, because the patients still benefit from the ICI therapy through the prolonged treatment effect. Nonetheless, we encourage collection of detailed information after discontinuation of index therapy in RCTs, such as the reason for not receiving subsequent therapy (e.g. patients’ functional status, patients’ refusal to receive a subsequent treatment). A perceived limitation of TNT-D is the potential influence of the prescribing patterns of individual physicians, especially when selecting the timing of switching over.³⁷ Differences in multidisciplinary care may also exist across geographic regions, and the availability of some therapies may drive treatment decisions.³⁷ Further exploration of TNT-D as an outcome of an RCT should be done, because TNT-D could complement PFS, which may suboptimally characterise the full impact of the novel mechanism of action of ICIs.³⁸ We focused in this study on TNT-D, because, by having the subsequent systemic treatment initiation or the death as endpoints, TNT-D has a greater overlap in the events shared with OS, since it captures all patients dying before progression just as PFS would, but also those who pseudo-progress but are never treated by a subsequent systemic treatment. This increased number of shared events is likely to yield a higher correlation. There have been some recent discussions around the importance of the depth of response (DepOR) as a relevant endpoint in ICI-treated patients.^{39,40} The meta-analytical approach for validation of SEs, however, was not applied in these research works.

| Criteria | Level of surrogacy | Point estimates (95% CI) |
|--|--------------------|--------------------------|
| Table 2. Estimated parameters for the assessment of the surrogate properties of time to next treatment or death for overall survival in previously untreated advanced melanoma patients treated with immune checkpoint inhibitors | | |
| Nivolumab monotherapy versus ipilimumab | | |
| Kendall’s τ | Individual | 0.63 (0.59-0.67) |
| R^2_{trial} | Trial | 1.00 (0.98-1.02) |
| Surrogate threshold effect (expressed in HR) | Trial | >0.35 and <0.61 |
| Nivolumab + ipilimumab versus ipilimumab | | |
| Kendall’s τ | Individual | 0.66 (0.62-0.70) |
| R^2_{trial} | Trial | 1.00 (0.99-1.01) |
| Surrogate threshold effect (expressed in HR) | Trial | >0.12 and <0.49 |

All values are rounded to 2 digits. CI, confidence interval; HR, hazard ratio; R^2_{trial} , coefficient of determination.

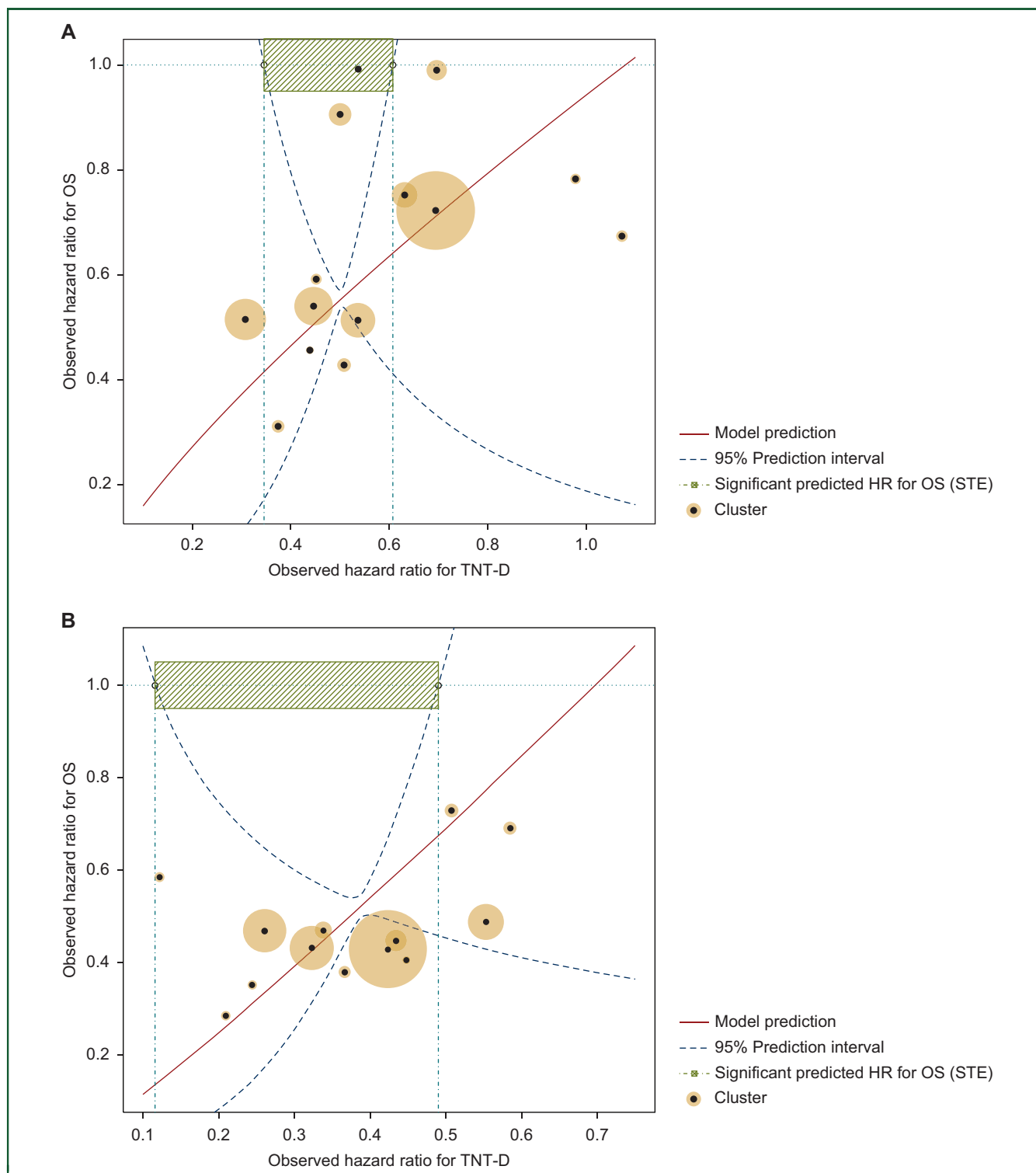


Figure 2. Relation between treatment effect on overall survival (OS) and the treatment effect on time to next treatment or death (TNT-D) for nivolumab monotherapy (A) and nivolumab plus ipilimumab (B) versus ipilimumab, with the surrogate threshold effect (STE).

The horizontal cyan line corresponds to a null hazard ratio (HR) for OS, i.e. HR = 1. The vertical line crosses the upper boundary of the 95% prediction limit where the HR for OS is 1. Between the vertical lines, the model predicts a significant treatment effect on OS. Circles are proportional to the number of patients in each cluster.

In order to assess the surrogacy properties of DepOR, new surrogate approaches need to be developed, since the statistical methods currently available cannot be applied when the dataset contains multiple zero values (i.e. patients with a DepOR of 0).

For two time-to-event endpoints, the copula model-based approach introduced by Burzykowski et al.²⁶ is a commonly used statistical method for validating an SE; however, model convergence issues and large standard error of the R^2_{trial} are frequently encountered with this 2-step

approach.^{29,41} The novel 1-step joint-frailty copula model was considered in this study because it reduces convergence and numerical problems and was found to be more robust than the traditional 2-step surrogacy approaches for two time-to-event end-points.²⁹

The limited amount of data is the main limitation of this study. IPD from RCTs of other therapies that share a mechanism of action and are tested in advanced melanoma patients, however, are limited. Moreover, access to such data, required for applying an appropriate surrogacy approach, is challenging. The estimated R^2_{trial} were close to 1 and the CIs were narrow. These values, although high, are reported frequently.^{25,27,29} The point estimates and their CIs must be interpreted with caution, though, because data from only one RCT were available for this study. The cluster-level definition was therefore the country of enrolment instead of RCT as advised in a meta-analytic approach. The heterogeneity in treatment effects between clusters in this RCT might be limited compared with that with a traditional meta-analytic approach; this may yield an overestimation of the correlation at the cluster level and may have artificially narrowed the reported CIs. Finally, due to the limited sample size, we were not able to carry out any subgroup analysis to assess whether TNT-D-OS association is similar within distinct subpopulations as, for example, according to the *BRAF* mutation status.

In conclusion, TNT-D appears to be a promising SE of OS in RCTs of ICI-treated patients with advanced melanoma. As a potential SE for OS, TNT-D may be crucial to present in the frame of regulatory and healthcare decision-making surrounding ICIs, especially when OS data are immature. We encourage sponsors of RCTs to carefully record the date of subsequent systemic treatment, as it has been reported to be captured heterogeneously.⁴² Surrogacy analyses could consequently be carried out with a larger number of RCTs to confirm these findings. Conducting similar surrogacy analyses in ICI-treated patients with other cancer types such as aRCC, or for ICI in combination with vascular endothelial growth factor inhibitors or chemotherapy would definitely be relevant for evaluating whether TNT-D is also a candidate SE for OS in these different settings.

ACKNOWLEDGEMENTS

We thank Professor Brigitte Dreno, head of the Dermatology Department, CHU of Nantes, France, for her review and her comments of the manuscript.

FUNDING

The study was funded by Bristol Myers Squibb. The funder contributed to the study design, collection, analysis and interpretation of the data in collaboration with the authors of this manuscript.

DISCLOSURE

SB, A-FG, MK and AM are employed by Bristol Myers Squibb (BMS). CLS and VR declare no known competing financial interests or personal relationships that could have appeared

to influence the work reported in this paper. AI declares advisory board consulting with Epizyme, Lilly, Merck Sharp & Dohme, Novartis, Pharmamar and Roche; and research grants from AstraZeneca, Bayer, BMS, Chugai, Merck Sharp & Dohme, Novartis, Pharmamar, Pfizer and Roche. CB declares personal fees from BMS outside of the submitted work.

REFERENCES

- Galluzzi L, Vacchelli E, Bravo-San Pedro JM, et al. Classification of current anticancer immunotherapies. *Oncotarget*. 2014;5(24):12472-12508.
- Larkin J, Chiarion-Sileni V, Gonzalez R, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med*. 2015;373(1):23-34.
- Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med*. 2018;378(14):1277-1290.
- Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med*. 2012;31(25):2973-2984.
- Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J Clin Oncol*. 2008;26(22):3721-3726.
- Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist*. 2008;13(suppl 2):19-21.
- U.S. Department of Health and Human Services Food and Drug Administration Oncology Center of Excellence Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). Clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry. 2018. Available at <https://www.fda.gov/media/71195/download>. Accessed October 5, 2020.
- European Medicines Agency. Guideline on the evaluation of anticancer medicinal products in man. Available at https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man-revision-5_en.pdf. Accessed October 5, 2020.
- Burzykowski T, Molenberghs G, Buyse M. *The evaluation of surrogate endpoints*. New York: Springer-Verlag; 2005.
- Branchoux S, Bellera C, Italiano A, Rustand D, Gaudin AF, Rondeau V. Immune-checkpoint inhibitors and candidate surrogate endpoints for overall survival across tumour types: a systematic literature review. *Crit Rev Oncol Hematol*. 2019;137:35-42.
- Mushti SL, Mulkey F, Sridhara R. Evaluation of overall response rate and progression-free survival as potential surrogate endpoints for overall survival in immunotherapy trials. *Clin Cancer Res*. 2018;24(10):2268-2275.
- Gyawali B, Hey SP, Kesselheim AS. A comparison of response patterns for progression-free survival and overall survival following treatment for cancer with PD-1 inhibitors: a meta-analysis of correlation and differences in effect sizes. *JAMA Netw Open*. 2018;1(2):e180416.
- Kaufman HL, Schwartz LH, William WN Jr, et al. Evaluation of classical clinical endpoints as surrogates for overall survival in patients treated with immune checkpoint blockers: a systematic review and meta-analysis. *J Cancer Res Clin Oncol*. 2018;144(11):2245-2261.
- Buyse M, Saad ED, Burzykowski T. Assessing treatment benefit in immuno-oncology. *Stat Biosci*. 2020;12:83-103.
- Mesnard C, Bodet-Milin C, Eugene T, Nguyen JM, Khammari A, Dreno B. Predictive value of FDG-PET imaging for relapse in metastatic melanoma patients treated with immunotherapy. *J Eur Acad Dermatol Venereol*. 2020;34(10):2261-2277.
- Savina M, Le Cesne A, Blay JY, et al. Patterns of care and outcomes of patients with METastatic soft tissue SARcoma in a real-life setting: the METASARC observational study. *BMC Med*. 2017;15(1):78.
- Luke JJ, Ghate SR, Kish J, et al. Real-world time to next treatment (TTNT) for first-line (1L) targeted and immuno-oncology therapies for BRAF-mutated metastatic melanoma (MM) by lactate dehydrogenase (LDH) level. *J Clin Oncol*. 2019;8(suppl):141.

18. Dudani S, Graham J, Wells JC, et al. First-line immuno-oncology combination therapies in metastatic renal-cell carcinoma: results from the international metastatic renal-cell carcinoma database consortium. *Eur Urol.* 2019;76(6):861-867.
19. McDermott DF, Rini BI, Motzer RJ, et al. Treatment-free interval (TFI) following discontinuation of first-line nivolumab plus ipilimumab (N+I) or sunitinib (S) in patients (pts) with advanced renal cell carcinoma (aRCC): CheckMate-214 analysis. *Ann Oncol.* 2018;29:VIII309.
20. Gray JE, Villegas A, Daniel D, et al. Three-year overall survival with durvalumab after chemoradiotherapy in stage III NSCLC-update from PACIFIC. *J Thorac Oncol.* 2020;15(2):288-293.
21. Izar B, Regan MM, McDermott DF. Clinical trial design and endpoints for stage IV melanoma in the modern era. *Cancer J.* 2017;23(1):63-67.
22. Regan MM, Werner L, Rao S, et al. Treatment-free survival: a novel outcome measure of the effects of immune checkpoint inhibition – a pooled analysis of patients with advanced melanoma. *J Clin Oncol.* 2019;37(35):3350-3358.
23. Schnipper LE, Davidson NE, Wollins DS, et al. Updating the American Society of Clinical Oncology value framework: revisions and reflections in response to comments received. *J Clin Oncol.* 2016;34(24):2925-2934.
24. Larkin J, Chiarion-Sileni V, Gonzalez R, et al. Five-year survival with combined nivolumab and ipilimumab in advanced melanoma. *N Engl J Med.* 2019;381(16):1535-1546.
25. Sofe CL, Emura T, Rondeau V. One-step validation method for surrogate endpoints using data from multiple randomized cancer clinical trials with failure-time endpoints. *Stat Med.* 2019;38(16):2928-2942.
26. 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. *Appl Stat.* 2001;50(4):405-422.
27. Rotolo F, Paoletti X, Burzykowski T, Buyse M, Michiels S. A Poisson approach to the validation of failure time surrogate endpoints in individual patient data meta-analyses. *Stat Methods Med Res.* 2019;28(1):170-183.
28. Renfro LA, Shi Q, Sargent DJ, Carlin BP. Bayesian adjusted R^2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials. *Stat Med.* 2012;31(8):743-761.
29. Sofe CL, Emura T, Rondeau V. A joint frailty-copula model for meta-analytic validation of failure time surrogate endpoints in clinical trials. *BIOM J.* 2020. Epub ahead of print.
30. Alonso A, Bigirimurame T, Burzykowski T, et al. *Applied surrogate endpoint evaluation methods with SAS and R.* Boca Raton, USA: Chapman and Hall/CRC; 2017.
31. Collette L, Burzykowski T, Carroll KJ, et al. Is prostate-specific antigen a valid surrogate end point for survival in hormonally treated patients with metastatic prostate cancer? Joint research of the European Organisation for Research and Treatment of Cancer, the Limburgs Universitair Centrum, and AstraZeneca Pharmaceuticals. *J Clin Oncol.* 2005;23(25):6139-6148.
32. Coart E, Suciú 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.
33. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat.* 2006;5(3):173-186.
34. Xie W, Halabi S, Tierney JF, et al. A systematic review and recommendation for reporting of surrogate endpoint evaluation using meta-analyses. *JNCI Cancer Spectr.* 2019;3(1):pkz002.
35. Król A, Mauguen A, Mazroui Y, Laurent A, Michiels S, Rondeau V. Tutorial in joint modeling and prediction: a statistical software for correlated longitudinal outcomes, recurrent events and a terminal event. *J Stat Software.* 2017;81:1-52.
36. Marshall J, Schwartzberg L, Bepler G, et al. Novel panomic validation of time to next treatment (TNT) as an effective surrogate outcome measure in 4,729 patients. *J Clin Oncol.* 2016;34:11521.
37. Campbell BA, Scarisbrick JJ, Kim YH, Wilcox RA, McCormack C, Prince HM. Time to next treatment as a meaningful endpoint for trials of primary cutaneous lymphoma. *Cancers (Basel).* 2020;12(8):2311.
38. Regan MM, Atkins MB, Powles T, et al. Treatment-free survival, with and without toxicity, as a novel outcome applied to immuno-oncology agents in advanced renal cell carcinoma. *Ann Oncol.* 2019;30:V393-V394.
39. Osgood C, Mulkey F, Mishra-Kalyani PS, et al. FDA analysis of depth of response (DpR) and survival across 10 randomized controlled trials in patients with previously untreated unresectable or metastatic melanoma (UMM) by therapy type. *J Clin Oncol.* 2019;37(15 suppl): 9508.
40. McCoach CE, Blumenthal GM, Zhang L, et al. Exploratory analysis of the association of depth of response and survival in patients with metastatic non-small-cell lung cancer treated with a targeted therapy or immunotherapy. *Ann Oncol.* 2017;28(11):2707-2714.
41. Belhechmi S, Michiels S, Paoletti X, Rotolo F. An alternative trial-level measure for evaluating failure-time surrogate endpoints based on prediction error. *Contemp Clin Trials Commun.* 2019;15: 100402.
42. Blumenthal GM, Gong Y, Kehl K, et al. Analysis of time-to-treatment discontinuation of targeted therapy, immunotherapy, and chemotherapy in clinical trials of patients with non-small-cell lung cancer. *Ann Oncol.* 2019;30(5):830-838.