



HAL
open science

Surrogate endpoints for overall survival in immune-checkpoint inhibitors–treated patients with advanced cancer

Sébastien Branchoux

► **To cite this version:**

Sébastien Branchoux. Surrogate endpoints for overall survival in immune-checkpoint inhibitors–treated patients with advanced cancer. Human health and pathology. Université de Bordeaux, 2020. English. NNT : 2020BORD0253 . tel-03505899

HAL Id: tel-03505899

<https://theses.hal.science/tel-03505899v1>

Submitted on 1 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse
présentée pour obtenir le grade de
Docteur de l'Université de Bordeaux

Ecole Doctorale Sociétés, Politique, Santé Publique
Spécialité Santé Publique, option Epidémiologie

Par Sébastien BRANCHOUX

Critères de substitution de la survie globale chez les patients atteints de cancer métastatique traités par inhibiteurs de points de contrôle immunologiques

Surrogate endpoints for overall survival in immune-checkpoint inhibitors–treated patients with advanced cancer

Sous la direction de Virginie RONDEAU & Carine BELLERA

Soutenue le 11 décembre 2020 devant les membres du jury :

M. MICHIELS Stefan	Dr, INSERM U1018, Paris	Rapporteur
M. PENEL Nicolas	Pr, Centre Oscar Lambret, Lille	Rapporteur
Mme MATHOULIN-PELISSIER Simone	Pr, INSERM U1219, Bordeaux	Président
Mme MOLLEVI Caroline	Dr, INSERM U1194, Montpellier	Examinatrice
Mme GAUDIN Anne-Françoise	Dr, Bristol Myers Squibb, Rueil-Malmaison	Invitée
Mme BELLERA Carine	Dr, INSERM U1219, Bordeaux	Directrice de thèse
Mme RONDEAU Virginie	Dr, INSERM U1219, Bordeaux	Directrice de thèse

Abstract

Advanced cancer treatment has been recently revolutionized by the development of the immune-checkpoint inhibitors (ICI). These immunomodulatory monoclonal antibodies are designed to either elicit a novel anti-tumoral immune response or revitalize an existing one to fight against cancer. Patients with cancer are living longer due to these improved therapies. Powering a study for overall survival (OS), the gold standard primary endpoint in randomized controlled trial (RCT) of anticancer drugs is becoming increasingly challenging. Therefore, it is of importance to identify and validate novel surrogate endpoints (SE) for OS in ICI-treated patients for expediting patients' access to innovative and potentially life extending medicines. We first systematically reviewed published studies reporting on an association between alternative endpoints and OS in ICI-treated patients. Then, based on the learnings from this systematic literature review and from the specificity of the mechanism of action of ICIs, we evaluated the surrogacy properties of an emerging intermediate endpoint in solid tumors, namely *time to next treatment* (TNT), in ICI-treated patients with advanced melanoma and renal cell carcinoma (aRCC), through recent innovative statistical models for the validation of SE. Based on the results of these surrogacy analyses, TNT seems a promising SE for OS in RCTs of ICI-treated patients with advanced melanoma and aRCC. We encourage sponsors of RCTs of ICI to carefully collect the date of subsequent systemic treatment, so that surrogacy analyses could consequently be performed with a larger number of RCTs in order to confirm our findings.

Résumé

La prise en charge du cancer au stade avancé ou métastatique a été profondément modifiée avec l'arrivée des inhibiteurs des points de contrôle immunologiques (*immune-checkpoint inhibitors (ICI)*). Ces anticorps monoclonaux immuno-modulateurs ont été développés pour soit déclencher une nouvelle réponse immunitaire anti-tumorale, soit réactiver une réponse existante pour lutter contre le cancer. L'espérance de vie des patients traités par ce type de thérapie est plus longue par rapport à ceux traités par les thérapies usuelles. Par conséquent, la puissance statistique requise dans un essai clinique randomisé (ECR) ayant pour objectif principal d'estimer l'effet relatif du traitement sur la survie globale (SG), critère de référence en oncologie, peut être difficile à atteindre. Dans ce contexte, il est important d'identifier et de valider des critères de substitution de la SG chez les patients traités par ICI afin notamment de permettre un accès précoce à ces traitements innovants. Nous avons tout d'abord effectué une revue systématique de la littérature des différents critères cliniques intermédiaires associés à la SG chez les patients traités par ICI. Puis, à partir des conclusions de cette revue et de la connaissance de la spécificité du mécanisme d'action des ICI, nous avons évalué les propriétés de substitution d'un nouveau critère, le « temps jusqu'à l'initiation d'un traitement systémique ultérieur » (*time to next treatment (TNT)*), chez les patients atteints d'un mélanome avancé ou d'un carcinome à cellules rénales avancé, à partir de modèles statistiques récemment développés pour la validation de critères de substitution. D'après les résultats de ces analyses, le TNT semble être un critère de substitution prometteur dans ces 2 populations. Nous encourageons les promoteurs d'ECR d'ICI à recueillir la date d'initiation du traitement systémique ultérieur afin de pouvoir réaliser des analyses similaires de plus grande ampleur et de confirmer ainsi nos résultats.

Acknowledgments

Au Pr Simone Mathoulin-Pelissier,

Au Dr Stefan Michiels,

Au Dr Caroline Mollevi,

Au Pr Nicolas Penel,

Pour avoir accepté de faire partie de mon jury de thèse.

A mes directrices de thèse, Dr Carine Bellera et Dr Virginie Rondeau,

Pour votre encadrement et votre soutien durant ces 3 années. C'était une réelle chance d'être encadré par vous deux. J'espère sincèrement que nous continuerons à collaborer dans un cadre différent.

Au Dr Anne-Françoise Gaudin,

Merci Anne-Françoise de m'avoir proposé une seconde fois de réaliser un projet doctoral dans le cadre de mon développement. Avec plus de maturité, on comprend mieux l'intérêt de mener ce type de projet. Je te remercie également pour tes qualités managériales et de m'avoir dégagé du temps pour réaliser cette thèse dans les délais impartis.

Au Laboratoire Bristol Myers Squibb et à l'équipe des ressources humaines,

Pour avoir accepté de financer cette thèse, et pour avoir mis à disposition les données cliniques nécessaires à la réalisation de cette thèse.

Au Dr Casimir Sofeu,

Pour ta disponibilité sur toute la période de ce projet, pour tes multiples explications des modèles statistiques que tu as développés. Tu as énormément contribué aux résultats de ces travaux de thèse. Merci.

Au Pr Antoine Italiano,

Pour votre disponibilité et pour nos riches échanges qui nous ont permis d'orienter les travaux de recherche dans la bonne direction.

Scientific valorization

Thesis-related communications

Publications in peer-reviewed journals

i. **Branchoux S**, Bellera C, Italiano A, Rustand D, Gaudin A-F, 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. (Impact Factor: 5.012)

ii. **Branchoux S**, Sofeu C, Gaudin A-F, Kurt M, Moshyk A, Italiano A, Bellera C, Rondeau V. Time to next treatment 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. *Submitted to European Journal of Cancer*.

iii. **Branchoux S**, Sofeu C, Kurt M, Gaudin A-F, May JR, Italiano A, Bellera C, Rondeau V. Investigating time to next treatment as a surrogate endpoint for overall survival in advanced renal cell carcinoma patients treated with immune-checkpoint inhibitors: results from the phase III CheckMate 214 trial. *Under review by the co-authors*.

Oral communication

Branchoux S, Sofeu C, Kurt M, Gaudin A-F, May J, Italiano A, Rondeau V, Bellera C. Investigating Time to Next Treatment As a Surrogate Endpoint for Overall Survival in Previously Untreated Intermediate- to Poor-Risk Advanced Renal Cell Carcinoma Patients: An Insight from the Phase III CheckMate-214. **ISPOR-EU 2020 virtual congress**.

Invited seminars

Branchoux S, Bellera C, Italiano A, Rustand D, Gaudin A-F, Rondeau V. Immune-checkpoint inhibitors and candidate surrogate endpoints for overall survival across tumour types: a systematic literature review. **8^{èmes} Journées Annuelles du Club Statisticiens et Mathématiques Appliquées à la Cancérologie (SMAC)**, January 2019, Bordeaux, France.

Posters

i. **Branchoux S**, Bellera C, Italiano A, Rustand D, Gaudin A-F, Rondeau V. Immune-checkpoint inhibitors and candidate surrogate endpoints for overall survival across tumour types: a systematic literature review. **EPICLIN 2019 / 26^{èmes} journées des Statisticiens de Centre de Lutte Contre le Cancer - Revue d'Épidémiologie et de Santé Publique 67, S143**.

ii. **Branchoux S**, Sofeu C, Gaudin A-F, Kurt M, Moshyk A, Italiano A, Bellera C, Rondeau V. Time to next treatment as a candidate surrogate endpoint for overall survival in advanced melanoma patients treated with immune-checkpoint inhibitors. **ISPOR-EU 2020 virtual congress**.

Other communications during the PhD program

Publication in peer-reviewed journals

Grumberg V, Roze S, Chevalier J, Gaudin A-F, Borrill J, **Branchoux S**. Comparison of overall survival (OS) extrapolations of immune-checkpoint inhibitors (ICI) in efficiency opinions of French 'Haute Autorité de Santé' (HAS) with updated OS data: were the projections accurate? *Under review by the co-authors*.

Oral communication

Grumberg V, Roze S, Chevalier J, Gaudin A-F, **Branchoux S**. Comparison of overall survival (OS) extrapolations of immune-checkpoint inhibitors (ICI) in efficiency opinions of French 'Haute Autorité de Santé' (HAS) with updated OS data: were the projections accurate? **ISPOR-EU 2020 virtual congress**.

Posters

i. **Branchoux S**, Cotte FE. Incorporation of anti-cancer drugs adverse events in Health Economics opinions of the French national authority for Health (HAS). **ISPOR-EU 2018 congress** (*Value in Health* 2018; **21**: S375).

ii. **Branchoux S**, Négrier S, de Peretti C, Malcolm B, May JR, Marié L, Gaudin A-F, Klijn SL, Ignacio TJ. Cost-effectiveness analysis of nivolumab in combination with ipilimumab versus sunitinib for the first-line treatment of intermediate- to poor-risk advanced renal cell carcinoma in France. **ISPOR-EU 2019 congress** (*Value in Health* 2019; **22**(3): S477).

iii. Brice P, **Branchoux S**, Prévost L, Jolivel R, Jouaneton B, Moreau-Mallet V, Gaudin AF, Cotté FE. Prise en charge en France des patients avec un lymphome de Hodgkin en rechute après une greffe de cellules souches autologue et après réponse au brentuximab vedotin: analyse à partir des données du PMSI. **Congress 2020 of the société française d'hématologie**.

iv. **Branchoux S**, Grevinga M, May JR, Malcolm B, Kroep S. Systematic identification of potential treatment effect modifiers in first line advanced renal cell carcinoma. **ISPOR-EU 2020 virtual congress**.

v. Kraan CW, Nientker KE, May JR, Malcolm B, Ejzykowicz F, Kurt M, Chun D, **Branchoux S**, Ho S, Spoorendonk JA. Efficacy & safety in previously untreated advanced/metastatic renal cell carcinoma – a systematic literature review update. **ISPOR-EU 2020 virtual congress**.

vi. Trip AM, May J, Malcolm B, Ejzykowicz F, Kurt M, Chun DS, Ho S, **Branchoux S**, Spoorendonk J. Real-World Effectiveness in Previously Untreated, Advanced/Metastatic Renal Cell Carcinoma – a Systematic Literature Review Update. **ISPOR-EU 2020 virtual congress**.

Résumé substantiel en français

Introduction

Le terme général de « cancer » s'applique à un groupe de maladies pouvant toucher n'importe quelle partie de l'organisme. L'un des traits caractéristiques du cancer est la prolifération rapide de cellules anormales qui, au-delà de leur délimitation habituelle, peuvent envahir des parties adjacentes de l'organisme, puis essaimer dans d'autres organes. On parle alors de métastases, celles-ci étant la principale cause de décès par cancer.¹ Le processus de cancérogenèse, qui repose sur la survenue d'aberrations génétiques successives perturbant de façon permanente la prolifération cellulaire, est un processus complexe faisant intervenir plusieurs acteurs liés à la cellule mais aussi à son environnement.

Le traitement des patients atteints de tumeur solide repose sur trois armes thérapeutiques principales, à savoir, la chirurgie, la radiothérapie, et les médicaments. Lorsque la maladie est à un stade avancé et que les patients présentent des métastases, les thérapies systémiques, telles que la chimiothérapie et les thérapies ciblées, sont utilisées à visée palliative dans une large majorité de cancers.

Au début des années 2010, la prise en charge du cancer métastatique a été profondément modifiée avec l'arrivée de nouveaux médicaments d'immunothérapie spécifique, les inhibiteurs de points de contrôle immunologiques (*immune-checkpoint inhibitors*, ICI). Ces médicaments permettent de bloquer les « freins de l'immunité » (PD-1, PD-L1, CTLA-4) et donc de réactiver le système immunitaire afin que celui-ci lutte plus efficacement contre les cellules tumorales.² L'arrivée sur le marché des ICI a entraîné la réévaluation des stratégies thérapeutiques de certains cancers, notamment le mélanome avancé (non résecable ou métastatique, stade III inopérable/IV), le cancer bronchique non à petites cellules métastatique et le carcinome à cellules rénales avancé (CCR). Pour ces traitements, le bénéfice en survie globale (SG), définie comme la durée entre la date de randomisation et le décès, peut être considéré comme considérable par rapport aux traitements de référence au regard des résultats des essais cliniques.^{3,4,5}

Le développement clinique d'un médicament est particulièrement réglementé. Généralement, l'efficacité et la tolérance d'un traitement expérimental sont étudiées au travers de trois phases cliniques (phases I à III) à l'issue desquelles une autorisation de mise sur le marché (AMM) peut être délivrée par les agences de régulation des médicaments (par exemple : *European Medicine*

Agency (EMA), Food and Drug Administration (FDA)). Au stade avancé ou métastatique du cancer, dans la dernière phase clinique avant l'obtention de l'AMM (phase III comparative), le critère de jugement principal est fréquemment la SG. Ce critère présente l'avantage d'être objectif et d'être une mesure pertinente du bénéfice pour le patient. Cependant, ce critère présente des inconvénients, comme la nécessité d'un suivi relativement long et d'un nombre important de patients afin de démontrer une différence cliniquement et statistiquement significative en comparaison au traitement de référence. De plus, lorsque des traitements innovants sont étudiés à des stades précoces de la maladie comme en première ligne de traitement systémique ou en situation adjuvante, l'effet du traitement sur la SG peut être influencé par les thérapies ultérieures, en particulier dans les cancers présentant un bon pronostic. A ces stades précoces de la maladie, le temps nécessaire pour obtenir des données de SG matures (c'est-à-dire une médiane de SG atteinte) peut être également relativement long. Dans ce contexte, la validation de critères de substitution (*surrogate endpoint*, SE), définis comme un biomarqueur ou un critère clinique intermédiaire dont l'objectif est de se substituer au critère clinique d'intérêt⁶, est particulièrement importante pour accélérer la mise à disposition des traitements innovants.

Selon l'*International Conference on Harmonization Guidelines on Statistical Principles for Clinical Trials*,⁷ un critère de substitution doit satisfaire trois conditions : (i) être biologiquement pertinent, (ii) au niveau du patient, être associé au critère de jugement principal : le critère doit permettre de prédire l'évolution de la maladie, et (iii) au niveau de l'essai, l'effet du traitement sur le critère de substitution doit être associé à celui sur le critère de jugement principal et permettre de prédire l'effet du traitement. Une simple corrélation entre 2 critères n'est donc pas suffisante pour qualifier un critère clinique intermédiaire de critère de substitution. Par ailleurs, les critères de substitution sont spécifiques d'une maladie ou d'un type de cancer, d'un stade de la maladie et du mécanisme d'action du traitement.

Plusieurs approches statistiques ont été développées pour apprécier la validité d'un critère de substitution. L'approche méta-analytique (c'est-à-dire considérant plusieurs essais cliniques randomisés) sur données individuelles est la référence car elle est la plus robuste.⁸ En effet, cette méthode est la seule permettant de mesurer la capacité du critère de substitution à prédire l'effet du traitement sur le critère de jugement principal. Parmi les approches méta-analytiques pour des critères de jugement de type « temps jusqu'à un événement », critères les plus utilisés en oncologie, celle développée par Burzykowski et coll.⁹ est aujourd'hui la plus utilisée. Dans cette approche, l'estimation de l'association s'effectue en deux étapes. Tout d'abord, un modèle à copule est utilisé pour estimer l'association au niveau individuel entre le critère de substitution et

le critère final ainsi que les effets du traitement sur le critère de substitution et sur le critère de jugement principal pour chaque essai. Dans un second temps, le coefficient de détermination (R^2_{trial}) est calculé à partir d'un modèle à effets aléatoires intégrant les effets traitement estimés au cours de la première étape. Le R^2_{trial} permet de mesurer l'association au niveau essai entre les effets du traitement sur le critère de substitution et sur le critère de jugement principal. R^2_{trial} est le plus souvent ajusté sur les erreurs d'estimation et est appelé $R^2_{\text{trial,adj}}$.

Cependant, des problèmes de convergence sont fréquemment rencontrés avec ce modèle et l'estimation du $R^2_{\text{trial,adj}}$ n'est pas toujours disponible.^{10,11,12} Ces problèmes numériques sont notamment influencés par le niveau d'association entre les deux critères, le nombre et la taille des essais considérés.¹² A partir de ces observations, Sofeu et coll. ont développé et validé deux nouveaux modèles statistiques pour la validation de critères de substitution. Le premier modèle est basé sur un modèle conjoint à effets aléatoires partagés, nommé le « *one-step joint surrogate model* ». ¹² Dans ce modèle, un effet aléatoire commun au critère de substitution et au critère de jugement principal est introduit pour prendre en compte l'hétérogénéité dans les temps de survie au niveau individuel. La variance de cet effet aléatoire permet de mesurer l'association entre les deux critères au niveau individuel. Au niveau essai, deux effets aléatoires corrélés en interaction avec le traitement sont considérés. Leurs paramètres de variance sont utilisés pour mesurer l'association entre les deux critères au niveau essai.

Compte tenu du temps de calcul nécessaire pour obtenir des estimations de tous les paramètres constituant le *one-step joint surrogate model*, Sofeu et coll. ont tout récemment développé une autre approche, nommée le « *one-step joint frailty-copula model* ». ¹³ Dans ce modèle conjoint, au lieu d'avoir des effets aléatoires au niveau individuel comme dans le *one-step joint surrogate model*, la dépendance entre le critère de substitution et le critère de jugement principal est prise en compte par une fonction de copule. La mesure de l'association entre les deux critères au niveau essai est effectuée suivant la même approche que dans le *one-step joint surrogate model*. Avec ces deux approches, une réduction considérable des problèmes de convergence a été observée par rapport à l'approche de Burzykowski et coll. ^{12,13}

Dans ces modèles, l'association au niveau individuel entre le critère de substitution et le critère de jugement principal est mesurée par le coefficient de corrélation τ de Kendall. Le τ de Kendall est compris entre -1 et +1, une valeur proche de 0 signifiant une indépendance entre les deux variables. Une valeur supérieure à 0,6 peut être considérée comme suffisante pour valider un critère de substitution au niveau individuel, des valeurs supérieures à 0,7 étant rarement observées avec ce type de modèle. ¹⁴ Au niveau essai, l'association est mesurée par le R^2_{trial} . Un R^2_{trial} proche de 1 est nécessaire pour qualifier un critère de substitution de valide au niveau essai. ^{9,12} Cette méthode propose également d'estimer l'effet seuil du critère de substitution

(*surrogate threshold effect*, STE), qui permet d'avoir une estimation de l'effet du traitement minimal à observer sur le critère de substitution afin de prédire un effet significatif du traitement sur le critère de jugement principal.¹⁵

Ces approches méta-analytiques supposent, par définition, la disponibilité de plusieurs essais cliniques randomisés (ECR). Lorsque peu d'ECR sont disponibles, il est suggéré de considérer le pays d'inclusion ou des régions du monde comme regroupement au sein du (ou des) ECR disponible(s) pour appliquer ces approches.¹⁶

Objectifs

L'identification de critères cliniques intermédiaires qui peuvent potentiellement mettre en évidence le bénéfice d'un traitement de manière précoce est un sujet d'intérêt majeur en oncologie. La validation d'un critère intermédiaire comme critère de substitution permettrait d'observer l'efficacité d'un traitement de manière plus précoce qu'en évaluant la SG. Le développement clinique de traitements innovants serait alors accéléré.

Le développement des ICI étant relativement récent, une synthèse exhaustive de l'information disponible sur les critères intermédiaires associés à la SG chez les patients atteints de cancer traités par ICI n'a pas été identifiée.

Aussi, le premier objectif de cette thèse était d'effectuer une revue systématique de la littérature afin de synthétiser toute l'information disponible sur les critères intermédiaires associés à la SG et d'identifier les critères de substitution validés chez les patients traités par ICI. D'après les résultats de cette revue, aucun critère de substitution validé n'a été identifié chez les patients atteints de cancer avancé ou métastatique traités par ICI. Cependant, deux critères composites développés spécifiquement pour les ICI, nommés le *durable response rate* et l'*intermediate response endpoint*, semblaient être mieux corrélés à la SG que les critères intermédiaires traditionnels comme la survie sans progression ou le taux de réponse objective. L'analyse des caractéristiques de ces deux critères a révélé qu'ils ne constituaient pas des critères de substitution potentiels pour des ECR menés auprès de patients traités par ICI.

En hématologie et dans les sarcomes des tissus mous, un nouveau critère intermédiaire, nommé « temps entre la date de randomisation et la date d'initiation d'un traitement systémique ultérieur ou la date de décès, selon ce qui survient en premier » (*time to next treatment*, TNT), a émergé récemment et s'avère être un critère cliniquement pertinent.¹⁷ Compte tenu de la spécificité des ICI et de leur action prolongée même après l'arrêt du traitement chez un nombre important de

patients, ce critère semble approprié pour évaluer l'efficacité des ICI chez patients atteints de tumeur solide.

Le second objectif de cette thèse était alors d'estimer le TNT et d'apprécier ses propriétés de substitution chez les patients atteints de mélanome avancé traités par ICI. Nous avons appliqué le *one-step joint surrogate model* sur les données de l'étude CheckMate 067.

Les propriétés de substitution du TNT étant prometteuses chez les patients atteints de mélanome avancé, **nous avons souhaité étudier si le TNT était un critère de substitution potentiel également chez les patients atteints d'un carcinome à cellules rénales avancé traités par ICI.** Suite à la publication au cours de cette thèse de la deuxième approche statistique développée par Sofeu et coll. (c'est-à-dire le *one-step joint frailty-copula model*), nous avons appliqué les deux types de modèles conjoints sur les données de l'étude CheckMate 214.

ARTICLE 1 : Inhibiteurs de checkpoints immunologiques et critères de substitution à la survie globale selon le type de cancer : résultats d'une revue systématique de la littérature

Contexte et objectif

L'arrivée des ICI a modifié profondément le traitement du cancer. Leur mécanisme d'action diffère de celui des autres thérapies en restaurant l'activité des cellules T de deux façons : soit en empêchant leur inactivation, soit en les réactivant afin de déclencher une réaction immunitaire contre les cellules tumorales. Le développement de critères cliniques intermédiaires qui peuvent potentiellement mettre en évidence le bénéfice d'un traitement de manière précoce est un sujet d'intérêt majeur en oncologie. Si un tel critère est cliniquement et statistiquement validé comme associé à la SG, il peut être utilisé comme critère de substitution. Les critères de substitution sont cependant spécifiques d'une maladie ou d'un type de cancer, d'un stade de la maladie et du mécanisme d'action du traitement. L'objectif de cette étude était de synthétiser toute l'information disponible sur les critères alternatifs associés à la SG chez les patients atteints de cancer traités par ICI et d'identifier ceux validés comme critère de substitution.

Matériel et méthodes

Les bases de données PubMed et Embase ont été consultées afin de collecter toutes les communications scientifiques rapportant une association entre un critère clinique et la SG chez les patients traités par ICI et publiées entre le 01/01/2003 et le 31/03/2018. Deux chercheurs ont

collecté de manière indépendante les informations suivantes : type de publication, type de cancer, stade de la maladie, méthodes statistiques utilisées, critères alternatifs étudiés et résultats.

Résultats

Parmi les 6 335 références identifiées, 24 ont été retenues. Vingt références étaient des recherches originales, publiées majoritairement sous forme d'article (N=14), et 4 références étaient des discussions. La quasi-totalité des recherches ont été menées chez des patients atteints d'un cancer métastatique ou avancé (95%). Parmi les recherches originales, 30% d'entre elles ont restreint leurs analyses aux patients atteints de mélanome (N=6) et 30% aux patients atteints de cancer du poumon non à petites cellules (N=6). Dans 8 recherches, tous les types d'ICI (anti-PD(L)1, anti-CTLA4) ont été agrégés pour l'analyse et 6 présentaient des résultats par type d'ICI. Une importante hétérogénéité dans les approches statistiques utilisées a été mise en évidence. L'approche standard de validation d'un critère de substitution en 2 niveaux (analyse au niveau individuel et au niveau essai) a été réalisée et publiée dans seulement 3 articles. Au stade métastatique, les principaux critères alternatifs traditionnels, tels que la survie sans progression (N=10) et le taux de réponse objective (N=8), ont été testés. De nouveaux critères alternatifs composites qui intègrent une notion de durée de réponse, développés spécifiquement pour l'immunothérapie (nommés « *durable response rate* » et « *intermediate response endpoint* »), semblaient être mieux corrélés à la SG pour les types de cancer analysés.

Conclusion

Peu d'études ont analysé la corrélation entre des critères cliniques alternatifs et la SG chez les patients traités par ICI. D'après les données disponibles, il n'existe pas de critère de substitution à la SG validé dans cette population. Au stade métastatique, des analyses de la corrélation entre des critères composites et la SG, selon une approche statistique appropriée et par type de cancer, méritent d'être menées.

Ce travail est publié dans *Critical Reviews in Oncology and Hematology*.

Discussion supplémentaire

Après avoir réalisé une analyse approfondie des caractéristiques de ces deux critères, il s'est avéré qu'ils ne constituaient pas des critères de substitution potentiels pour les patients traités par ICI. Par ailleurs, cette revue systématique nous a permis d'observer que les critères intermédiaires basés sur la réponse tumorale étaient faiblement corrélés à la SG dans cette population de patients.

ARTICLE 2 : Temps jusqu'à l'initiation d'un traitement systémique ultérieur comme critère de substitution potentiel de la survie globale chez les patients atteints de mélanome avancé et traités par ICI

Contexte et objectif

Une des spécificités des ICI est l'effet prolongé du traitement. En effet, chez certains patients, notamment ceux atteints de mélanome, le traitement par ICI peut être arrêté (pour cause d'événement indésirable par exemple), sans que ces patients ne bénéficient immédiatement d'une ligne de traitement systémique ultérieure.¹⁸ Une période sans traitement, plus ou moins longue, est observée. Les critères intermédiaires basés sur la réponse tumorale n'étant pas prometteurs, nous nous sommes intéressés aux propriétés de substitution du « temps entre la date de randomisation et la date d'initiation d'un traitement systémique ultérieur ou la date de décès, selon ce qui survient en premier » (*time to next treatment*, TNT) chez les patients atteints de mélanome avancé, le suivi nécessaire pour obtenir des données de SG matures chez ces patients pouvant être particulièrement long. Ce critère capte, de par sa définition l'effet prolongé des ICI observés chez les patients atteints de mélanome.¹⁸ Aussi, l'objectif de cette étude était d'estimer le TNT et d'évaluer ses propriétés de substitution à la SG chez les patients atteints de mélanome avancé traités par ICI en 1^{ère} ligne de traitement systémique.

Matériel et méthodes

Les données individuelles de l'étude CheckMate 067 ont été utilisées pour cette étude. L'étude CheckMate 067 est un ECR de phase III, mené en double-aveugle, comparant l'efficacité et la tolérance de nivolumab associé à ipilimumab, et de nivolumab en monothérapie *versus* ipilimumab en monothérapie chez des patients atteints de mélanome avancé naïfs de traitement systémique. L'approche du *one-step joint surrogate model* a été utilisée pour évaluer les propriétés de substitution du TNT, en retenant le pays d'inclusion comme regroupement. Le coefficient de corrélation τ de Kendall et le coefficient de détermination (R^2_{trial}) ont été estimés pour la mesure de l'association au niveau individuel et au niveau essai. Le STE, l'effet du traitement minimal à observer sur le TNT afin de prédire un effet significatif du traitement sur la SG, a également été estimé. Une analyse de type « *leave-one-out cross-validation* » a été réalisée afin d'évaluer l'exactitude du modèle de prédiction.

Résultats

A partir des 945 patients inclus dans l'étude, 15 regroupements ont été considérés. Pour le groupe « nivolumab associé à ipilimumab » et pour le groupe « nivolumab en monothérapie », l'association au niveau individuel était acceptable (τ de Kendall > 0.6) et forte au niveau essai : le R^2_{trial} était proche de 1 avec des intervalles de confiance étroits. L'analyse de validation croisée a montré que les mesures d'association étaient peu sensibles à l'exclusion de regroupement de manière itérative. Cependant, la précision des prédictions était modérée.

Conclusion

Les résultats de cette étude suggèrent que le TNT peut être un critère de substitution pertinent chez les patients atteints de mélanome avancé traités par ICI. Nous encourageons les promoteurs d'ECR d'ICI à recueillir la date d'initiation du traitement systémique ultérieur afin de réaliser des analyses similaires en considérant plusieurs ECR et de confirmer ainsi les résultats de cette étude.

Ce travail a été soumis à la revue *European Journal of Cancer*.

ARTICLE 3 : Temps jusqu'à l'initiation d'un traitement systémique ultérieur comme critère de substitution potentiel de la survie globale chez les patients atteints de carcinome à cellules rénales avancé et traités par ICI

Contexte et objectif

D'après les résultats de l'étude menée auprès de patients atteints de mélanome avancé traités par ICI, le TNT semble être un critère de substitution approprié dans cette population. Comme chez les patients atteints d'un mélanome avancé, chez les patients atteints de CCR :

- Une proportion importante de patients arrête le traitement par ICI sans bénéficier directement d'une nouvelle ligne de traitement systémique;¹⁹
- Le suivi nécessaire pour obtenir des données de SG matures peut être long.⁴

Aussi, l'objectif de cette étude était d'évaluer les propriétés de substitution du TNT chez les patients atteints de CCR traités par ICI en 1^{ère} ligne de traitement systémique, à partir de deux approches statistiques, le *one-step joint surrogate model* et le *one-step joint frailty-copula model*, cette dernière ayant été publiée au cours de cette thèse.

Matériel et méthodes

Les données individuelles de l'étude CheckMate 214 ont été utilisées pour cette étude. L'étude CheckMate 214 est un ECR de phase III, mené en ouvert, comparant l'efficacité et la tolérance de nivolumab associé à ipilimumab *versus* sunitinib chez des patients atteints de CCR de pronostic intermédiaire ou défavorable naïfs de traitement systémique. Les approches du *one-step joint surrogate model* et du *one-step joint frailty-copula model* ont été utilisées pour évaluer les propriétés de substitution du TNT, en retenant le pays d'inclusion comme regroupement. Le coefficient de corrélation τ de Kendall et le coefficient de détermination (R^2_{trial}) ont été estimés pour la mesure de l'association au niveau individuel et au niveau essai. Le STE, c'est-à-dire l'effet du traitement minimal à observer sur le TNT afin de prédire un effet significatif du traitement sur la SG, a également été estimé. Une analyse de type « *leave-one-out cross-validation* » a été réalisée afin d'évaluer l'exactitude et la précision des prédictions.

Résultats

Dans cette étude, 847 patients ont été inclus et 22 regroupements ont été considérés. Les résultats des deux approches statistiques étaient similaires : l'association au niveau individuel était acceptable (τ de Kendall ≥ 0.59) et forte au niveau essai (les R^2_{trial} étaient proches de 1 avec des intervalles de confiance étroits). Le STE, exprimé en HR, était estimé à 0.70 pour le groupe « nivolumab associé à ipilimumab » *versus* le groupe « sunitinib » d'après le *one-step joint surrogate model* ; le *one-step joint frailty Clayton copula model* prédisait un effet significatif du traitement seulement lorsque le HR du TNT était compris entre 0,65 et 0,74. L'analyse de validation croisée a montré que les mesures d'association étaient peu sensibles à l'exclusion de regroupements de manière itérative. Cependant, la précision des prédictions était meilleure avec le *one-step joint frailty Clayton copula model*, bien que les intervalles de prédiction étaient relativement larges.

Conclusion

Les résultats de cette étude suggèrent que le TNT peut être un critère de substitution pertinent chez les patients atteints de CCR avancé traités par ICI. Des analyses similaires en considérant plusieurs ECR sont nécessaires pour confirmer les résultats de cette étude.

Ce travail est en cours de relecture par les auteurs.

Conclusion générale et perspectives

Des critères cliniques autres que la SG sont nécessaires pour l'évaluation réglementaire des traitements du cancer. Il est important d'identifier et de valider des critères de substitution pour les ICI afin de permettre un accès précoce aux traitements innovants.²⁰ L'espérance de vie des patients atteints de cancer et traités par ces nouvelles thérapies étant plus longue, la puissance statistique requise dans un ECR ayant pour objectif principal d'estimer l'effet relatif du traitement sur la SG peut être difficile à atteindre. Les critères de substitution sont par ailleurs particulièrement pertinents lorsque le temps nécessaire pour obtenir des données de SG matures est long, comme dans le mélanome avancé et le CCR.

Dans la première partie de ce travail, nous avons synthétisé toute l'information publiée sur les critères intermédiaires associés à la SG chez les patients traités par ICI. D'après les résultats de cette revue, aucun critère de substitution validé n'a été identifié chez les patients atteints de cancer avancé ou métastatique traités par ICI. Cette revue nous a cependant permis de confirmer que les critères intermédiaires basés sur la tumeur, tels que la survie sans progression ou le taux de réponse objective n'étaient probablement pas les critères les plus appropriés pour les ICI au regard de leur mécanisme d'action qui diffère des autres traitements du cancer. De plus, nous avons mis en évidence que la durée de la réponse est une caractéristique importante à considérer pour un critère intermédiaire d'ECR réalisé chez des patients traités par ICI.

Dans la seconde partie de ce travail, nous avons étudié les propriétés de substitution du TNT chez les patients atteints de mélanome avancé, compte tenu des résultats de la revue systématique de la littérature et de l'effet prolongé des ICI même après l'arrêt du traitement. Il s'agit des premiers travaux concluants dans ce domaine, bien que le travail ait été réalisé à partir d'un seul ECR. Nous nous sommes enfin intéressés aux propriétés de substitution du TNT dans une autre population de patients traités par ICI. Deux approches statistiques ont été appliquées. D'après les résultats de cette analyse, le TNT semble également être un critère de substitution pertinent chez les patients atteints de CCR traités par ICI.

Les résultats de ces travaux de thèse sont prometteurs. Il est néanmoins nécessaire de réaliser des analyses en considérant plusieurs ECR pour valider de manière plus robuste ce critère de substitution. Cependant, les ECR d'ICI par type de tumeur sont peu nombreux, et il peut être difficile d'avoir accès aux données individuelles des patients afin d'appliquer l'approche méta-analytique. Des chercheurs académiques pourraient solliciter les firmes pharmaceutiques

développant ces traitements pour obtenir les données individuelles et réaliser ainsi des analyses de plus grande ampleur. Ces travaux de recherche pourraient également être étendus, afin d'évaluer si le TNT est un critère de substitution potentiel chez les patients atteints par d'autres types de cancer et traités par ICI.

Abbreviations

AE: Adverse event

aNSCLC: Advanced non-small cell lung cancer

aRCC: Advanced renal cell carcinoma

ASCO: American society of clinical oncology

CI: Confidence interval

CTLA4: Cytotoxic T-lymphocyte antigen-4

CRF: Case report form

DLT: Dose-limiting toxicity

DRR: Durable response rate

EMA: European medicine agency

FDA: Food and drug administration

HR: Hazard ratio

HRA: Health regulatory agency

ICH: International conference of harmonization

ICI: Immune-checkpoint inhibitors

IMDC: International metastatic renal cell carcinoma database consortium

IME: Intermediate response endpoint

IPD: Individual-level patient data

IPI: Ipilimumab

IrAE: Immune-related adverse event

LOOCV: Leave-one-out cross validation

Mabs: Monoclonal antibodies

MoA: Mechanism of action

MTD: Maximum tolerated dose

NIVO: Nivolumab

OS: Overall survival

ORR: Overall response rate

PASS: Post-authorization safety study

PD-1: Programmed death 1 receptor

PD-L1: Programmed death ligand 1

PD-L2: Programmed death ligand 2

PFS: Progression-free survival

RMST: Restricted mean survival time

RP2D: Recommended Phase II dose

RCT: Randomized controlled trial

SE: Surrogate endpoint

TKI: tyrosine kinase inhibitor

TNT: Time to next treatment

TTF: Time to treatment failure

STE: Surrogate threshold effect

Definitions

Overall Survival (OS): OS is defined as the time from the date of randomization to the date of death (any cause).

Treatment-free interval (TFI): TFI is defined as the time from the end of a therapy until the date of initiation of a subsequent treatment or death, whichever occurs first.

Time to next treatment (TNT): TNT is defined as the time between the date of randomization and the date of initiation of the subsequent systemic treatment or the date of death, whichever occurs first.

Time to treatment failure (TTF): TTF is defined as the time between the date of randomization and the date of discontinuation of the index therapy for any reason (including death, progression, and toxicity).

Table of Contents

1	Introduction.....	25
1.1	Carcinogenesis and cancer epidemiology	25
1.2	Therapeutic management of cancer	25
1.3	RECIST 1.1 and iRECIST response evaluation criteria.....	31
1.3.1	RECIST 1.1 response evaluation criteria	31
1.3.2	iRECIST response evaluation criteria	33
1.4	Clinical development of anticancer therapies.....	33
1.5	Overall survival as an endpoint in RCT of anticancer drugs.....	35
1.6	Context and statistical evaluation of surrogate endpoints	35
1.6.1	Context.....	35
1.6.2	Statistical evaluation of surrogate endpoints.....	36
1.6.2.1	The two-stage models for the validation of surrogate endpoints	37
1.6.2.2	One-step approach – the joint frailty surrogate model.....	37
1.6.2.3	Surrogate threshold effect (STE)	39
1.7	Synthesis and research objectives	39
2	Immune-checkpoint inhibitors and candidate surrogate endpoints for overall survival across tumor types: A systematic literature review	41
2.1	Introduction.....	41
2.2	Publication.....	41
2.3	Additional discussion of the results from the systematic literature review	51
2.3.1	Durable response rate endpoint (DRR).....	51
2.3.2	Intermediate response endpoint (IME).....	53
3	Time to next treatment as a candidate surrogate endpoint for overall survival in advanced melanoma patients treated with immune-checkpoint inhibitors	55
3.1	Introduction.....	55
3.2	Description of CheckMate 067 trial	56

3.3	Publication.....	58
4	Time to next treatment as a candidate surrogate endpoint for overall survival in advanced renal cell carcinoma patients treated with immune-checkpoint inhibitors.....	77
4.1	Introduction.....	77
4.2	Description of the one-step joint frailty-copula approach.....	78
4.3	Description of CheckMate 214 trial.....	79
4.4	Publication.....	80
4.5	Additional discussion.....	107
5	General Discussion.....	109
6	Conclusion.....	113
7	References.....	115

1 Introduction

1.1 Carcinogenesis and cancer epidemiology

Cancer is a common term for a large group of diseases that can develop in almost any organ or tissue of the body when abnormal cells grow uncontrollably, go beyond their usual boundaries to invade adjoining parts of the body and/or spread to other organs.¹ Carcinogenesis is a complex process involving several actors linked to the cell but also to its environment. Cancer is first and foremost a disease of the genome, characterized by DNA damage. These lesions can be caused by environmental factors (radiation, chemicals, and viruses) or more rarely be inherited. Carcinogenesis is based on the occurrence of successive genetic aberrations that permanently disrupt cell proliferation, resulting in a tumor.

In 2018, around 18 million people were affected by cancer worldwide.¹ Cancer is the second leading cause of death globally, accounting for an estimated 9.6 million deaths, or one in six deaths, in 2018. Lung, prostate, colorectal, stomach and liver cancer are the most common types of cancer in men, while breast, colorectal, lung, cervical and thyroid cancer are the most common among women.¹

1.2 Therapeutic management of cancer

The treatment of patients with solid tumors (cancers other than those of the hematopoietic system such as leukemia) is based on three main therapeutics:

- Surgery;
- Radiotherapy;
- Anticancer therapies.

These therapeutic modalities can be used alone or in combination for curative purposes (aiming to cure the patient) or for palliative purposes (the cancer is incurable and the goal is to extend patient survival by also maintaining his health-related quality of life). The choice of a therapeutic strategy is based on various factors, such as:

- Patient-related: age, general condition, comorbidities,
- Tumor-related: size, location, histological type, grade, stage of dissemination.

a) Surgery

Two types of surgery exist:

- The curative surgery: this type of surgery is the main therapeutic strategy of most solid tumors, and allows to cure patients in a large proportion of cases. It consists of an excision of the tumor and sometimes of the lymph nodes.

- The palliative surgery which is only intended to improve symptoms and sometimes to extend survival of the patient.

b) Radiotherapy

Radiotherapy is with surgery one of the oldest and most widely used treatment methods in fighting against cancer. Radiotherapy can be either curative or palliative. Curative radiotherapy is used in more than half of the cases.

c) Anticancer therapies

Access to drugs to treat cancer is relatively recent: the first clinical studies using "nitrogen mustards" date back to 1942, and since the 1960s, combinations of drugs have been shown to be effective to regress or even cure advanced cancers. Anticancer therapies are classified according to their mechanism of action (MoA).

- Chemotherapy

Chemotherapy aims to treat tumor cells. Initially used in metastatic solid tumors and in hematologic neoplasms (e.g. leukemia and lymphoma), chemotherapy is increasingly used early in combination with other locoregional cancer therapies to "preventively" treat a disease (called adjuvant chemotherapy) or act synergistically on the primary tumor with other therapies (neoadjuvant chemotherapy, concomitant radio chemotherapy). The principle of anticancer chemotherapy is based on systemic treatment destroying tumor tissue more sensitive to the "damage" induced by these drugs than normal cells. Thus, the chemotherapies chosen must offer positive efficacy/toxicity and benefits/risks ratios depending on the drugs used and the treated tumors.

Most cancer chemotherapy drugs act by inhibiting the progression of cells within the cell cycle phases: this growth inhibition is frequently associated with an induction of differentiation, or apoptosis. The multiples abnormalities responsible for tumor cell growth therefore quickly led to the use of poly-chemotherapy combining drugs with multiple molecular targets. Combination of these drugs with different MoA yields an addition or a synergistic action of inducing the cell death.

- Hormonotherapy

Hormonotherapy is a technique that involves modifying the hormonal environment in patients with tumors dependent on sex hormones (estrogen and progesterone for cancer of the breast, androgens for prostate cancer).

- Targeted therapies

Targeting therapies do not target DNA but molecules specific to tumor cells which are little or no expressed by normal cells (also called tumor-targeting therapies). These drugs have therefore a better specificity of action and consequently a better safety profile. The expression “tumor-targeting” refers to drugs that (1) specifically alter the signaling functions of receptors expressed on the surface of malignant cells; (2) bind to, and hence neutralize, trophic signals produced by malignant cells or by stromal components of neoplastic lesions; (3) selectively recognize cancer cells based on the expression of a “tumor-associated antigen”, i.e., an antigen specifically (or at least predominantly) expressed by transformed cells but not (or at least less so) by their non-malignant counterparts.²¹ Among these drugs, we can mention:

- Tyrosine kinase inhibitors (TKI): these drugs inhibit the activity of proteins present on the surface of tumor cells which are crucial for their proliferation and their survival.
- Monoclonal antibodies (Mabs): produced by genetic engineering, they target an antigen present on the surface of the malignant cells blocking its function and / or causing cell death. Mabs are part of the wide class of immunotherapy treatment, presented in the next section.

- Immunotherapy

Anticancer immunotherapies are defined as interventions that mediate antineoplastic effects by initiating a novel or boosting an existing immune response against neoplastic cells.²¹ Figure 1 illustrates all the anticancer treatment classified as immunotherapies.

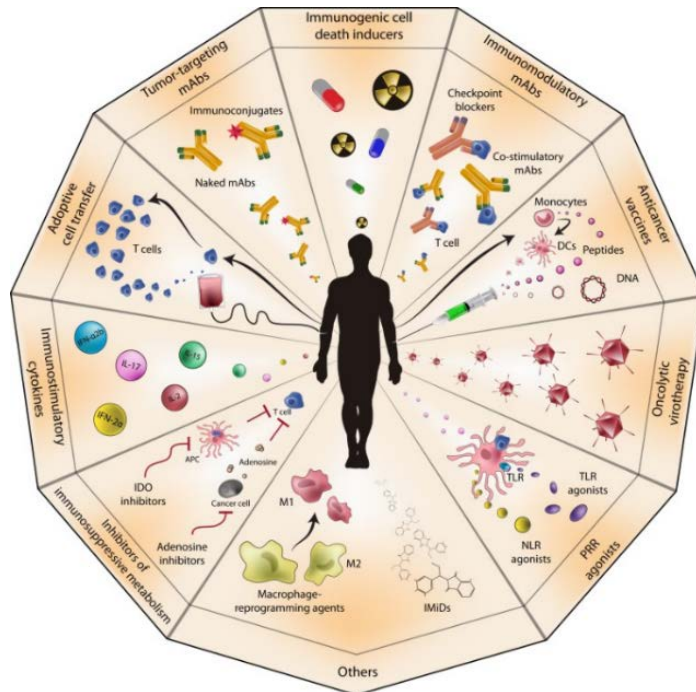


Figure 1. Illustration of the anticancer immunotherapies according to the mechanism of action (from Galluzzi et al²¹). Monoclonal antibodies (mAbs); dendritic cell (DC)-, pattern recognition receptor (PRR) agonists; 1MT, 1-methyltryptophan; APC, antigen-presenting cell; IDO, indoleamine 2,3-dioxygenase; IFN, interferon; IL, interleukin; IMiD, immunomodulatory drug; NLR, NOD-like receptor; TLR, Toll-like receptor.

Immunotherapies are generally classified as “passive” or “active” based on their ability to (re-) activate the host immune system against malignant cells.²¹ For instance, tumor-targeting mAbs are considered passive forms of immunotherapy; conversely, anticancer vaccines and checkpoint inhibitors exert anticancer effects only upon the engagement of the host immune system, constituting clear examples of active immunotherapy. The former type of therapy, called immune-checkpoint inhibitors (ICI), has revolutionized the cancer treatment (as also adoptive T-cell therapies in the meantime) and is considered as the third generation in this field after cytotoxic drugs and molecular agents acting at oncogene-related targets.²² In 2013, ICI and adoptive T-cell were granted the scientific breakthrough of the year. Moreover, the 2018 Nobel Prize in Physiology or Medicine was awarded to James P. Allison and Tasuku Honjo for their discovery of cancer therapy by inhibition of negative immune regulation.²³

The first developed and marketed ICI was ipilimumab in advanced melanoma in 2011. Ipilimumab is a Cytotoxic T-lymphocyte antigen-4 (CTLA-4), which is a key regulator of T-cell activity. Ipilimumab blocks T-cell inhibitory signals induced by the CTLA-4 pathway, increasing the number of reactive T-effector cells which mobilize to mount a direct T-cell immune attack against tumor cells. CTLA-4 blockade can also reduce T-regulatory cell function, which may contribute to an

anti-tumor immune response.²⁴

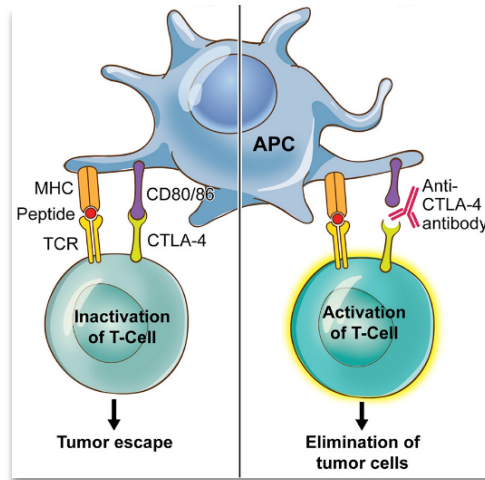


Figure 2. Illustration of the mechanism of action of ipilimumab. APC, antigen-presenting cell.

In 2014 and in 2015, two other types of ICI, namely pembrolizumab and nivolumab, have also been granted a marketing authorization by both FDA and EMA agencies for the treatment of advanced melanoma. Both therapies are humanized monoclonal anti-programmed cell death-1 (PD-1) antibody, which binds to the PD-1 receptor and blocks its interaction with PD-L1 and PD-L2.^{25,26} The PD-1 receptor is a negative regulator of T-cell activity that has been shown to be involved in the control of T-cell immune responses. Engagement of PD-1 with the ligands PD-L1 and PD-L2, which are expressed in antigen presenting cells and may be expressed by tumors or other cells in the tumor microenvironment, results in inhibition of T-cell proliferation and cytokine secretion.^{25,26} Nivolumab and pembrolizumab potentiates T-cell responses, including anti-tumor responses, through blockade of PD-1 binding to PD-L1 and PD-L2 ligands.^{25,26} Figure 3 presents the mechanism of action of anti-PD-1 agents.

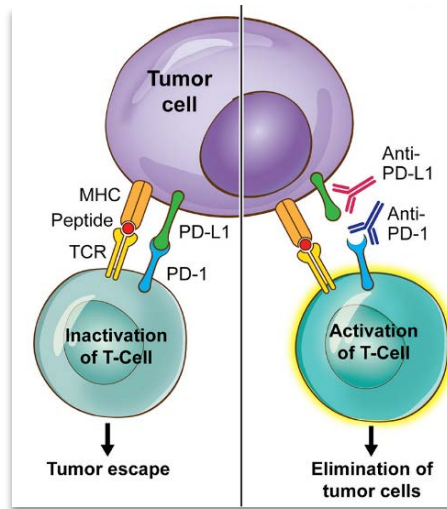


Figure 3. Illustration of the mechanism of action of anti-PD-1.

To date, seven ICIs (anti-CTLA4, anti-PD1, or anti-PDL-1) have received marketing authorization (ipilimumab, pembrolizumab, nivolumab, atezolizumab, avelumab, durvalumab and cemiplimab) in a variety of cancers, and more than 1,000 clinical trials involving the agents are under way.²⁷ ICI has led to the re-evaluation of the therapeutic strategies of some cancers. These treatments are indicated as monotherapies or in combination with other systemic treatments in more than a dozen types of cancer at different treatment settings, and at different stages such as locally advanced or metastatic. For some cancers in advanced stages, one of the current treatment strategies is to combine different ICIs targeting two distinct receptors, like nivolumab (anti-PD1) and ipilimumab (anti-CTLA4) in the treatment of advanced melanoma or renal cell carcinoma (aRCC) patients. ICI offer the possibility of new lines of treatment, and may potentially delay the use of palliative care.²

The new mechanism of action of these therapies involves some modifications in the clinical development or in the evaluation of the clinical benefit.²⁸ For instance, ICI agents typically can produce different patterns of response compared with standard systemic therapies. Being treated by ICI, some patients experience immune-related responses such as initial increase in the size of tumors or appearance of new lesions, before a subsequent and sustained reduction in tumor burden occur. This immunotherapy-specific phenomenon cannot be properly evaluated by means of the conventional response evaluation criteria in solid tumors (RECIST V.1.1) criteria, and has stimulated the development of immune-related response criteria, the iRECIST.²⁹ These response evaluation criteria in solid tumors are described in the subsequent section. There are also other efficacy specificities of ICI, such as a delayed onset of treatment effect, and subsets of patients

with prolonged and durable response. From a safety perspective, the adverse events (AEs) observed in ICI-treated patients are different from those observed in patients treated with standard therapies. These AEs, directly related to the MoA of this new therapeutic class, are named immune-related adverse events (IrAEs). They can develop with significant latency, becoming evident only with late follow-up or pharmacovigilance studies. Specific guidelines on the management of these types of toxicities were recently published.³⁰

1.3 RECIST 1.1 and iRECIST response evaluation criteria

1.3.1 RECIST 1.1 response evaluation criteria

The RECIST 1.1 criteria, revised in 2009, constitute a standardized measure for evaluating tumor response.³¹ The principle of RECIST 1.1 criteria follows these 3 steps:³²

- Estimation of the overall tumor burden at baseline (which will be used as a comparator for subsequent measurements);
- Identification of measurable and non-measurable lesions;
- Identification of target lesions and non-target lesions.

The definitions of the criteria used to determine objective tumor response for target lesions and non-target lesions is presented in Table 1 and Table 2, respectively.

Table 1. Criteria used to determine objective tumor response for target lesions³²

Complete Response (CR)	Disappearance of all target lesions. Any pathological lymph nodes (whether target or non-target) must have reduction in short axis to <10 mm.
Partial Response (PR)	At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters.
Progressive Disease (PD)	At least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm. (Note: the appearance of one or more new lesions is also considered progression).
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum diameters while on study.

Table 2. Criteria used to determine objective tumor response for non-target lesions³²

Complete Response (CR)	Disappearance of all non-target lesions and normalization of tumor marker level. All lymph nodes must be non-pathological in size (<10 mm short axis).
Non-CR/Non-PD	Persistence of one or more non-target lesion(s) and/or maintenance of tumor marker level above the normal limits.
Progressive Disease (PD)	Unequivocal progression of existing non-target lesions. (Note: the appearance of one or more new lesions is also considered progression).

Based on the assessment of the tumor presented in Table 1 and Table 2, the overall response status calculation at each time point for patients who have measurable disease at baseline (i.e. the majority of the patients) is presented in Table 3.

Table 3. Criteria used to determine objective tumor response at each time point³²

Target lesions	Non-target lesion	New lesions	Overall Response
CR	CR	No	CR
CR	Non-CR/non-PD	No	PR
CR	Not evaluated	No	PR
PR	Non-PD or not all evaluated	No	PR
SD	Non-PD or not all evaluated	No	SD
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD
Not all evaluated	Non-PD	No	Inevaluable

The RECIST 1.1 criteria were developed before the arrival of the immunotherapy agents and present some limitations for the response evaluation when the patients is treated with such drug with a different mechanism of action. The iRECIST criteria (“i” stands for “immune”) were consequently developed.

1.3.2 iRECIST response evaluation criteria

The mechanism of action of immune modulators should be taken into account when evaluating the response to immunotherapies. Two specific characteristics of immunotherapy have led to the development of the iRECIST consensus guideline:³³

- "pseudo-progression" phenomenon: increase in tumor size, secondarily followed by reduction, or the appearance of new lesions that disappear secondarily;
- The potential delay of the anti-tumor action, even without pseudo-progression (on the contrary of cytotoxic chemotherapy).

This guideline describes a standard approach to solid tumor measurements and definitions for objective change in tumor size for use in trials in which an immunotherapy is used.²⁹ The two main change in this consensus guideline is the management of new malignant lesions (target and non-target), and the need to confirm the progression of the disease for taking into account, identifying and characterizing atypical responses (e.g. delayed response after pseudo-progression). A new type of response was created, namely the "immune unconfirmed progressive disease (iUPD)". Progression should be confirmed during imaging examination following the iUPD (4-8 weeks later) for being classified as "immune confirmed progressive disease (iCPD)". Full explanation of the iRECIST consensus guideline is described by Seymour et al.²⁹

1.4 Clinical development of anticancer therapies

The drug development process can be broadly classified as pre-clinical and clinical. Pre-clinical refers to experimentation that occurs before it is given to human subjects, and clinical refers to experimentation in humans.

When authorized to be tested on humans, the therapeutic product goes through three main clinical phases (phase I to phase III) which will eventually lead to a marketing authorization. In oncology, trials include volunteer patients with cancer for whom validated treatments have failed.

Phase I trials aim to establish the recommended dose and/or schedule of new drugs or drug combinations for phase II trials.³⁴ Endpoints include toxicity endpoints and dose-limiting toxicity (DLT) defined as an AE that is serious enough to prevent an increase in dose or level. DLTs are determined *a priori*, in order to subsequently identify the maximum tolerated dose (MTD), defined as the highest dose that does not cause unacceptable rate of DLTs. The recommended phase II dose (RP2D) is defined based on the MTD, the overall safety profile of the drug, and its

pharmacokinetic profile. To this extent, increasing doses of the drug are tested on different cohorts (one to three patients) until the highest dose with acceptable DLT rate is found. The guiding principle for dose escalation is to avoid unnecessary exposure of patients to sub-therapeutic doses of the drug while preserving safety and maintaining rapid accrual. Phase I clinical trials in oncology are small (20 to 50 subjects), single-arm, open-label, sequential studies that usually include patients with a good performance status and for whom cancer has progressed despite standard treatments.

Phase II trials aim to assess preliminary signs of anti-tumor activity of a new drug, in order to screen out ineffective drugs and identify promising new drugs for further evaluation in phase III trials. Typical phase II endpoints include overall response rate (ORR) defined, for example, as per the RECIST criteria in the case of solid tumors presented in the previous section. The investigational drug is prescribed at the RP2D determined in preliminary phase I trials. For ethical reasons, studies of new agents in oncology usually are designed with two or more stages of accrual allowing early stopping due to inactivity of the agent. Randomization can also be employed in phase II trials. The primary aim in such case is not to formally compare the treatment arms as in subsequent phase III trials, but rather to collect efficacy data in a similar population treated with the standard strategy (in the absence of historical data), or to assess distinct administration schedules and/or routes of the drug. Sample size usually ranges from 30 to 60 patients, but can sometimes include more patients, specifically for randomized phase II trials.

Building on the data from phase I and II trials, phase III trials aim to provide confirmatory proof of the clinical benefit of a new treatment, by demonstrating that the new treatment is superior, non-inferior, or equivalent either to no treatment, placebo, or the best available therapy. Phase III trials are randomized to ensure that groups are alike in all important prognostic factors and only differ in the treatment each group receives, thus providing the basis for causality inference. The gold standard for the primary efficacy endpoint in phase III trial is overall survival (OS),^{35, 36} defined as the time between the date of randomization and the date of death from any cause. However, in some limited cases, progression-free survival (PFS) may be used as primary endpoint to measure efficacy of the drug, such as, for example, in trials when there is cross-over because of early demonstration of activity of the new treatment and lack of alternatives.³⁷

Finally, post-authorization safety studies (PASS) are carried out once the drug has been approved by HRA. The aim of PASS is to identify and evaluate the long-term effects of the new drug over a lengthy period for a greater number of patients than in phase III, usually thousands. New drugs can be tested continuously to uncover more information about long-term efficacy, safety and side effects after being approved for marketing.

1.5 Overall survival as an endpoint in RCT of anticancer drugs

In oncology, several clinical endpoints have been used to assess treatment efficacy.^{6,38} Overall survival (OS), defined as the time from randomization to the date of death (any cause), is frequently considered as the gold standard for assessing treatment efficacy in randomized controlled trials (RCT) of anticancer therapies from a regulatory perspective.^{35,36} Survival time is arguably the most objective metric for assessing the efficacy of anticancer treatment, and - along with quality of life - the most relevant measure of patient benefit.²⁰ However, this endpoint may require extensive follow-up and a substantial number of patients to demonstrate a clinically and statistically significant relevant difference when comparing therapeutic strategies, and include all cause-deaths. When innovative treatment is investigated at an early stage of the disease, OS might also present the disadvantage to be confounded by the effects from subsequent therapies especially in cancers with better prognosis where treatment landscape is rapidly evolving in all stages. Moreover, at this stage, it might take a significant period of time to demonstrate a significant OS benefit while there is an urgent unmet medical need. In this context, surrogate endpoints (SE) are therefore particularly critical to expedite innovative and potentially life extending medicines to patients.

1.6 Context and statistical evaluation of surrogate endpoints

1.6.1 Context

An SE is defined as ‘a biomarker or an intermediate endpoint intended to substitute for a clinical endpoint’.⁶ SEs have been extensively studied within the molecular targeted therapy era as they may reduce the need for large sample sizes, and thus decrease the duration and cost of trials.³⁹ As of today, there is some evidence for good surrogacy for some SEs, as highlighted in a recent systematic literature review by Savina et al.⁴⁰ For instance, progression-free survival, defined as time from randomization until a disease progression or death (any cause) whichever occurs first, may be an appropriate SE for OS in the context of advanced colorectal cancer treated with chemotherapy and/or radiotherapy, or in locally advanced lung cancer treated with chemotherapy. However, the validity of an SE depends both on the mechanism of action of the treatment and on the disease setting, and as such it is recommended to validate all potential SEs on a case-by-case basis.⁸ Consequently, **SE validated in the framework of chemotherapy or targeted therapy era might not be appropriate for ICI agents.** Finally, surrogacy assessment requires

both clinical validation, that is a strong biological rationale, and statistical validation, described in the next section.

1.6.2 Statistical evaluation of surrogate endpoints

Statistical validation of an SE requires assessing simultaneously³⁹:

- (i) The individual-level association, that is, one must ensure an association between the final endpoint and the candidate surrogate endpoint: the surrogate endpoint predicts the course of the disease in an individual patient;
- (ii) The trial-level association, that is, one must ensure that treatment effect on the candidate surrogate endpoint enables adequate prediction of the treatment effect on the final endpoint.

The International Conference on Harmonisation (ICH) guidelines on statistical principles for clinical trials reflects the distinction between these two levels of surrogacy: “in practice, the strength of the evidence for surrogacy depends upon (i) the biological plausibility of the relationship, (ii) the demonstration in epidemiological studies of the prognostic value of the surrogate for the clinical outcome, and (iii) evidence from clinical trials that treatment effects on the surrogate correspond to effects on the clinical outcome”.^{7,39}

Since 1989 and the definition of the operational criteria for surrogacy introduced by Prentice,⁴¹ several statistical methods have been developed for the assessment of surrogate endpoints.⁸ These methods are classified into two main categories: the single-trial and the meta-analytic approaches. Historically, the single-trial approach was developed first, and relied on data analysis from a single trial. As such, this method cannot provide an estimate of the trial-level association. The key motivation for identifying a valid surrogate endpoint is to be able to predict the treatment effect on the final endpoint based on the treatment effect observed on the surrogate. This prediction requires assessing the association between the treatment effects on the two endpoints (trial-level association). Data from several RCTs are needed for assessing the association. The meta-analytic approach was therefore subsequently developed in order to allow one to estimate both the individual- and trial-level associations.

The remainder of this section focuses on the meta-analytic methods, the most robust approaches for validating an SE, developed for time-to-event endpoints, one of the most common endpoints in phase III RCT of anticancer drugs.

1.6.2.1 The two-stage models for the validation of surrogate endpoints

Several meta-analytic approaches are available for the assessment of the surrogacy properties of time-to-event endpoints.^{9,10,11,12,13} Until recently, the copula model-based approach introduced by Burzykowski et al.⁹ was considered the reference method for validating an SE. Initially, the two-stage model was developed in the context of endpoints following a Gaussian distribution⁴², and was subsequently extended by Burzykowski et al to the case of survival endpoints. The estimation strategy relies on two steps:

- In the first step, a copula model is used to measure the individual-level surrogacy and to estimate the treatment effects on each endpoint (i.e. the surrogate and the final endpoints) and for each trial considered in the study. In the analysis of such bivariate survival data, the key element is an appropriate account for dependence between event times as both endpoints are assessed on the same patient. This is performed through the *copula*, which is a function to link two random variables together to form a joint distribution.⁴³

- In the second step, the coefficient of determination (R^2_{trial}) is computed from a mixed-effects model of the treatment effects estimated in the first step, accounting for the measurement error resulting from using estimated effects. Such R^2_{trial} is adjusted on the estimation errors and is called the adjusted coefficient of determination, or $R^2_{\text{trial, adj}}$.

However, $R^2_{\text{trial, adj}}$ is not always available, mainly due to convergence problems.¹² Renfro et al¹¹ reported that these convergence problems were most often encountered in the first stage model (at the individual level) of validation. However, even when the first stage provides estimates, the second stage (at the trial level) does not systematically leads to an estimate of the adjusted coefficient of determination due to further numerical problems. These concerns are frequently encountered and are influenced by the number and size of the groups (or trials).¹²

Based on these observations, Sofeu et al.¹² developed a new statistical approach, called the “joint frailty surrogate model”, summarized in the next section.

1.6.2.2 One-step approach – the joint frailty surrogate model

A frailty model is a random effects model for time variables, where the random effect (the frailty) has a multiplicative effect on the hazard. The joint surrogate model is based on two frailty survival models, which are extension of the Cox proportional-hazards model to account for unobservable heterogeneity of the survival times among individuals.⁴⁴ The joint surrogate model considers also a shared random effect associated with baseline risks accounting for heterogeneity between trials, as well as two correlated treatment-by-trial interaction.

Let S_{ij} and T_{ij} be the failure times associated respectively with the surrogate and the true endpoints, for subject j ($j = 1, \dots, n_i$) belonging to the trial i ($i = 1, \dots, G$). Let $\mathbf{Z}_{S,ij} = (Z_{Sij1}, \dots, Z_{Sijp})'$ and $\mathbf{Z}_{T,ij} = (Z_{Tij1}, \dots, Z_{Tijp})'$ be the covariates associated with S_{ij} and T_{ij} , respectively. These covariates include the treatment arm. The joint surrogate model is defined according to Sofeu et al.¹² as:

$$\begin{cases} \lambda_{S,ij}(t|\omega_{ij}, u_i, v_{Si}, \mathbf{Z}_{S,ij}) = \lambda_{0S}(t) \exp(\omega_{ij} + u_i + v_{Si}Z_{ij1} + \beta_S \mathbf{Z}_{S,ij}) \\ \lambda_{T,ij}(t|\omega_{ij}, u_i, v_{Ti}, \mathbf{Z}_{T,ij}) = \lambda_{0T}(t) \exp(\zeta\omega_{ij} + \alpha u_i + v_{Ti}Z_{ij1} + \beta_T \mathbf{Z}_{T,ij}) \end{cases}$$

with,

$$\omega_{ij} \sim N(0, \theta), \quad u_i \sim N(0, \gamma)$$

and

$$\begin{pmatrix} v_{Si} \\ v_{Ti} \end{pmatrix} \sim MVN(\mathbf{0}, \Sigma_v), \quad \Sigma_v = \begin{pmatrix} \sigma_{vS}^2 & \sigma_{vST} \\ \sigma_{vST} & \sigma_{vT}^2 \end{pmatrix}.$$

$\lambda_{S,ij}(\cdot)$, $\lambda_{0S}(t)$, and β_{Sk} ($k = 1, \dots, p$) are the hazard functions of failure time for the j^{th} patient in trial i , the baseline hazard function, and the fixed effects corresponding to the covariates X_{ijk} associated with the surrogate endpoint (S), respectively; $\lambda_{T,ij}(\cdot)$, $\lambda_{0T}(t)$, and β_{Tk} ($k = 1, \dots, p$) are defined similarly for the true endpoint (T). ω_{ij} is a shared individual-level random effect that takes into account heterogeneity at the individual levels; ω_{ij} considers the correlation between the surrogate and the true endpoints at the individual level. u_i is a shared random effect associated with the baseline hazard function that will serve to take into account the heterogeneity between trials of the baseline hazard function, associated with the fact that there are several trials. The power parameters ζ and α distinguish both individual and trial-level heterogeneities between the surrogate and the true endpoints. v_{Si} and v_{Ti} are two correlated random effects treatment-by-trial interactions (trial-level frailties in interaction with the treatment), which aim to assess the prediction of the treatment effect on the true endpoint.

- Assessment of the individual-level surrogacy

The Kendall's τ coefficient of correlation is used to estimate the strength of the association between the candidate SE and the final endpoint at the individual level. The Kendall's τ is the difference between the probability of concordance and the probability of discordance of two realizations of S_{ij} and T_{ij} . It belongs to the interval $[-1; 1]$ and assumes a zero value when S_{ij} and T_{ij} are independent. The Kendall's τ is adjusted on the individual-level random effects, 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. The 95% confidence interval is estimated using the parametric bootstrap method.

- Assessment of the trial-level surrogacy

The trial-level association is estimated using the coefficient of determination (R^2_{trial}) obtained from the covariance matrix Σ_v . The 95% confidence interval is estimated using the delta method, which can result in confidence limits violating the [0; 1] interval. An R^2_{trial} sufficiently close to 1 is necessary for validating an SE at the trial-level, although there is no consensus on the extent to the prediction indicating a valid endpoint surrogacy.⁴⁵

The meta-analytical approach for validating of SE supposes the availability of data from multiple RCTs. When the data is limited to only a single or a few RCTs (which is common for drugs presenting a new mechanism of action), one can consider the geographic location of the centres as the cluster of analysis, which is a common practice in evaluating potential SE.^{16, 46, 47}

1.6.2.3 Surrogate threshold effect (STE)

For perfect prediction of the treatment effect on the true endpoints, R^2_{trial} must be equal to 1. However, such a situation is likely impossible in practice. Therefore, for R^2_{trial} different from 1, it is not clear what threshold would be sufficient for a valid surrogate. In 2006, Burzykowski and Buyse introduced the surrogate threshold effect (STE), defined as “the minimum treatment effect on the surrogate necessary to predict a non-zero (i.e. significant) effect on the true endpoint”.¹⁵ A candidate SE is considered strongly valid if the STE can be reached in a RCT. One of the most interesting characteristics of STE is its natural interpretation from a clinical point of view. In the case of a high level of surrogacy between an alternative endpoint and the final endpoint at both the individual and the trial levels, it is relevant to assess the STE. This information might be useful for drawing early inferences on the significance of the final endpoint, based on the observed treatment effect on the SE.

1.7 Synthesis and research objectives

SE are attractive to measure as they allow the efficacy of promising treatments to be established earlier than it would be possible if OS had to be observed, especially with therapy prolonging substantially survival, like ICI.

As the development of ICIs is relatively recent, summarized information on SE for OS in ICI-treated patients is lacking. **The first objective of this doctoral thesis was therefore to summarize the current evidence on clinical alternative endpoints associated with OS in ICI-treated patients by cancer type and to identify validated SE for OS for this drug class,**

if any. The conducted systematic literature review was inconclusive over validated SE for OS in ICI-treated patients. However, two alternative composite endpoints statistically seemed to better correlate with OS in the cancer types analyzed, namely the durable response rate (DRR) and the intermediate response endpoint (IME). This work is presented in chapter 2, and was published in *Critical Reviews in Oncology and Hematology*.

As presented at the end of this second chapter, the analysis of the characteristics of the two alternative endpoints identified through the literature review highlighted that these may not meet the expectations for being strong candidate SEs for OS in ICI-treated patients. On the other hand, time to next treatment (TNT), defined as the time between randomization and the date of initiation of the subsequent systemic treatment or the date of death whichever occurs first, has emerged as a relevant alternative clinical endpoint in hematologic malignancies and in metastatic soft-tissue sarcomas.¹⁷ Indeed, TNT captures durable response and the time free of systemic anticancer therapy, which may be achieved with ICI. **The second objective of this doctoral thesis was thus to formally assess the surrogate properties of TNT in previously untreated advanced melanoma patients treated with ICI.** We focused on previously untreated advanced melanoma patients as the maturity of survival data (i.e. median of OS) may take a significant period of time to be observed, and so identifying a SE for OS in this disease setting is particularly critical. We applied the SE one-step validation method based on a joint frailty model. This work is presented in chapter 3. The manuscript was submitted to *European Journal of Cancer*.

The surrogacy properties of TNT indicated that it may be an appropriate SE for OS in RCTs of advanced melanoma patients treated with ICI. In order to assess whether this SE may be also appropriate in another tumor type, **we formally assess the surrogate properties of TNT in previously untreated aRCC patients treated with ICI** (third objective). We relied on two statistical methodologies, namely the one-step joint surrogate model¹² as used in our analysis in advanced melanoma patients, as well as the one-step joint frailty-copula model¹³, which was developed during the course of this thesis, and which be even more robust than the former one. This work is presented in chapter 4. The manuscript is currently under review by the co-authors.

2 Immune-checkpoint inhibitors and candidate surrogate endpoints for overall survival across tumor types: A systematic literature review

2.1 Introduction

We searched the PubMed and Embase database for publications reporting on the association between a clinical endpoint and OS in ICI-treated populations from 01/01/2003 to 03/31/2018. We followed PRISMA guidelines for reporting of results.⁴⁸ Two independent reviewers collected information on tumor localization, disease stage, type of data (individual/aggregate), statistical methods, type of SE and results. Disagreements were resolved through discussions.

Out of 6,555 references retrieved, 24 were selected. Main reasons of exclusion were: non-specific to ICI or study related to PD-(L)1 expression. Among the selected references, 20 were research publications (6 non-small cell lung cancer, 6 melanoma, 1 urinary tract cancer, 7 multiple tumor locations) focusing on advanced or metastatic stage (N=19); 4 were general discussions. Patient-level data were considered in 10 references. Only 3 studies assessed surrogacy at both the patient and trial levels. The main traditional alternative endpoints included progression-free survival (N = 10) and objective response rate (N = 8). New alternative endpoints, such as durable response rate (DRR) (N = 1) and intermediate response endpoint (IME) (N = 1) statistically better correlate with OS in the cancer types analyzed.

Few studies investigated OS surrogacy for ICI. At the metastatic stage, there is no sufficient data to support validated SE for OS which should therefore remain the primary endpoint in RCT of ICI. Promising composite metrics such as DRR and IME were identified. Adequate surrogacy assessment based on validated statistical methods is still warranted to consider these endpoints as SE for OS in ICI-treated cancer patients. This work is published in *Critical Reviews in Oncology and Hematology*.

2.2 Publication

The publication is presented in the next pages.



Immune-checkpoint inhibitors and candidate surrogate endpoints for overall survival across tumour types: A systematic literature review

Sébastien Branchoux^{a,b,c,*}, Carine Bellera^{c,d,e}, Antoine Italiano^f, Denis Rustand^b, Anne-Françoise Gaudin^a, Virginie Rondeau^{b,c}

^a Department of Health Economics & Outcomes Research, Bristol-Myers Squibb, Rueil-Malmaison, France

^b Biostatistic Team, Bordeaux Population Health Center, ISPED, Centre INSERM U1219, INSERM, Bordeaux, France

^c Epicene Team (Cancer & Environnement), Bordeaux Population Health Center, ISPED, Centre INSERM U1219, INSERM, Bordeaux, France

^d Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Center, Bordeaux, France

^e Department of Clinical Epidemiology, INSERM CIC-EC 14.01, Bordeaux, France

^f Department of Early Phase Trial Unit, Institut Bergonié Comprehensive Cancer Centre, Bordeaux, France

ARTICLE INFO

Keywords:

Immune-checkpoint inhibitors
Surrogate endpoints
Overall survival
Literature review

ABSTRACT

Background: Surrogate endpoints (SEs) for overall survival (OS) are specific to therapeutic class. The objective of this review was to document all alternative endpoints studied for their association with OS in Immune-Checkpoint Inhibitors (ICI)-treated patients.

Methods: We searched PubMed and Embase for publications reporting the association between a clinical endpoint and OS in ICI-treated populations from 01/01/2003 to 03/31/2018.

Results: Out of 6,335 references identified, 24 were selected. Only 3 studies assessed surrogacy at both the patient and trial levels. The main traditional alternative endpoints included progression-free survival (N = 10) and objective response rate (N = 8). New alternative endpoints, such as durable response rate (N = 1) and intermediate response endpoint (N = 1) statistically better correlate with OS in the cancer types analysed.

Conclusion: Based on the published evidence, there is insufficient data to support validated SE for OS. Adequate surrogacy assessment of promising composite endpoints which consider a duration component is encouraged.

1. Introduction

Recently, cancer treatment has been revolutionized by the development of therapeutic agents targeting the immune system, considered as the third generation in this field after cytotoxic drugs and molecular agents acting at oncogene-related targets (Martin-Liberal et al., 2017). Among the different types of newly available immunotherapy treatments, immune-checkpoint inhibitors (ICI) constitute the most established and clinically promising therapy (Galluzzi et al., 2014). These immunomodulatory monoclonal antibodies are designed to elicit a novel or revitalise an existing antitumoral immune response.3 Antibodies which bind to the programmed cell death-1 (PD-1) receptor and block their interaction with ligands PD-L1 and PD-L2 are the most widely used therapies at the advanced stage of numerous cancers (Fujii et al., 2018; Rolfo et al., 2017). To date, six ICIs have received marketing authorization (ipilimumab, pembrolizumab, nivolumab, atezolizumab, avelumab and durvalumab) in several types of cancer. Some of them are still under investigation as monotherapy or in combination in

other types of cancer or disease settings.

In oncology, several clinical endpoints, defined as characteristics or variables that reflect how a patient feels, functions, or survives in response to a treatment, have been used to assess treatment efficacy (Fleming and Powers, 2012; Mathoulin-Pelissier et al., 2008). Overall survival (OS), defined as the time from randomization to the date of death (any cause), is considered as the gold standard for the primary efficacy endpoint in randomized clinical trials (RCT) (Pazdur, 2008). However, in some limited cases, progression-free survival (PFS) may be used as primary endpoint to measure efficacy of the drug, such as, for example, in trials when there is cross-over because of early demonstration of activity of the new treatment and lack of alternatives (Ocana and Tannock, 2001).

OS may require extensive follow-up to demonstrate significant and clinically relevant differences compared to standard care, and may thus delay patient access to promising new drugs. The development of alternative endpoints, such as recurrence-free survival (RFS - adjuvant setting) or PFS (metastatic setting), that could capture treatment benefit

* Corresponding author at: Bristol-Myers Squibb, 3, rue Joseph Monier, 92500, Rueil-Malmaison, France.

E-mail address: Sebastien.branchoux@bms.com (S. Branchoux). <https://doi.org/10.1016/j.critrevonc.2019.02.013>

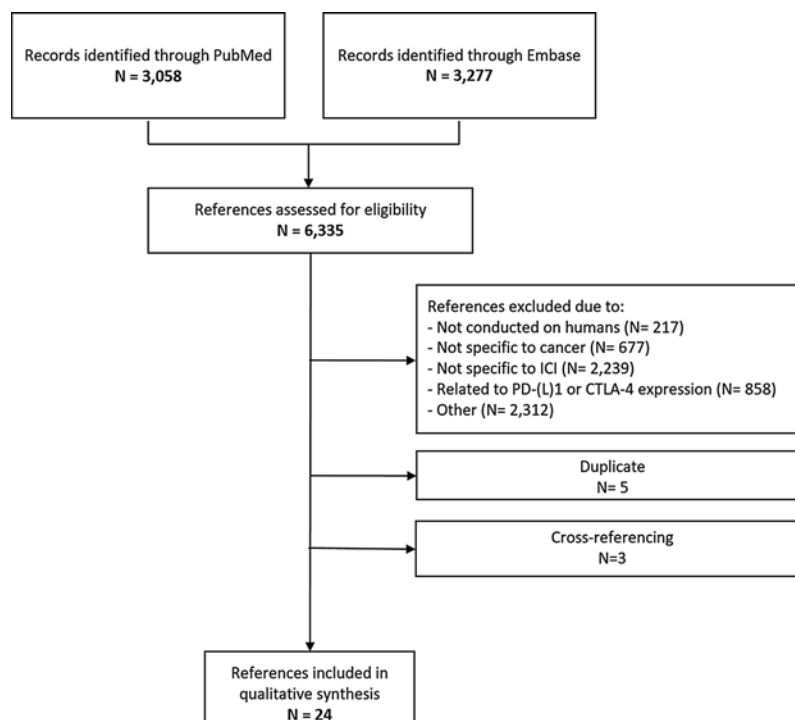


Fig. 1. Flow of information through the different phases of the systematic review of studies reporting on association between a clinical endpoint and overall survival in ICI-treated patients, as per PRISMA guidelines. (ICI = Immune-Checkpoint Inhibitors).

appropriately and be measurable earlier, has thus become central to clinical oncology. Such an endpoint, if clinically and statistically validated, could be considered as a surrogate endpoint (SE) for OS.

An SE is defined as ‘a biomarker intended to substitute for a clinical endpoint’ (Fleming and Powers, 2012). SEs have been extensively studied within the molecular targeted therapy era as they may reduce the need for large sample sizes, and thus decrease the duration and cost of trials (Alonso et al., 2017). Assessments of surrogacy should be performed at both the patient and trial levels, the most robust statistical method for validating a SE being the meta-analytic approach (Burzykowski et al., 2005). Since the validity of an SE depends both on the mechanism of action of the treatment and on the disease setting, it is recommended to validate all potential SEs on a case-by-case basis.

As of today, there is evidence for good surrogacy for certain SEs, as highlighted in a recent systematic literature review by Savina et al. (Savina et al., 2018) For example, PFS may be an appropriate SE for OS in the context of advanced colorectal cancer treated with chemotherapy and/or radiotherapy, or in locally advanced lung cancer treated with chemotherapy. As ICIs had not been developed extensively at the time of this review, information on SEs for OS in ICI-treated patients are still lacking.

In this context, we conducted a systematic literature review (i) to summarize current evidence on clinical alternative endpoints associated with OS in ICI-treated patients by cancer type and (ii) to identify validated SE for OS for this drug class, if any.

2. Methods

2.1. Search strategy and selection criteria

We searched PubMed and Embase for English language publications reporting the association between a clinical endpoint and OS in ICI-treated population from 01/01/2003 (year of the first clinical trial of the first ICI, ipilimumab) to 03/31/2018 (cut-off date). The computerized algorithm designed for extracting all references is available as Supplementary Material. Publications were first classified as eligible if

the abstract presented at least one of the three following characteristics: conducted in humans, specific to cancer and specific to ICI. The retained publications were then classified as related to ICI expression (such as prognostic role of PD-1 expression), which were excluded, or as related to an association between a clinical endpoint and OS, which were included in the qualitative synthesis. Other publications on subjects such as adverse events related to ICIs or practice guidelines were excluded. When studies led to multiple communications (i.e. conference abstract and then peer-review article), we only included the principal publication. We reported the results of the selection process following PRISMA guidelines for the reporting of systematic reviews. (Moher et al., 2009) This study is registered in the PROSPERO database (identification number: CRD42018097434).

2.2. Data analysis

Two reviewers (SB, DR) independently collected information on the format of the communication (abstract or article), type of publication (research study or discussion article), cancer type, treatment setting (adjuvant or metastatic), type of treatment, type of data (patient-level or aggregated data), number of trials, number of patients, alternative clinical endpoints analysed, statistical method and results. Disagreements between reviewers were resolved through discussions with senior authors (VR, CB).

3. Results

Out of 6,335 references retrieved through the algorithms, 24 were selected (Fig. 1). The most common reasons for exclusion were references not related to ICI (N = 2,239) and description of RCT results, practice guidelines or related to immune-related adverse events, main contributors of the “other” category (N = 2,312). PD-(L)1 or CTLA-4 expression was described in 15% of the publications (N = 858).

Of the references selected, twenty were primary research studies and four were discussion articles such as literature reviews (Table 1). Key characteristics of the research studies are summarized in Table 2.

Table 1
Characteristics of the references identified in the systematic literature review (N = 24).

Type of cancer	Treatment setting	Type of treatment	Authors	Type of publication	Type of data	Number of trials	Number of patients	Method to assess the association	Alternative clinical endpoints	Results
Melanoma	adjuvant	Interferon (main analysis) Anti-CTLA-4 (secondary predictive analysis)	(Suciu et al., 2018)	article	individual patient-level	13	6,708	2-stage approach: - ILS: Copula function with 2 Cox models (1 for RFS and 1 for OS), Spearman coefficient (r) - TLS: R ² from linear regression with adjustment for the treatment effect estimates weighted linear regressions	RFS	r=0.89 (95% CI [0.88–0.89]) R ² =0.91 (95% CI [0.81–1.01])
	metastatic	Mix of treatment (including ICI) Anti-CTLA-4	(Flaherty et al., 2013) (Lee et al., 2013)	abstract Abstract	Aggregated Individual patient-level	10 1	4,215 676	Principal component analysis & Cox model weighted linear regressions (Pearson coefficient)	PFS HRQoL (QLQC-30 questionnaire) median PFS OS milestone (1-y) OS milestone (2-y)	r=0.74 (95% CI [0.26–0.93]) r=0.90 (95% CI [0.66–0.97]) OS was associated with most of the HRQoL domain scores by treatment arm : r(mPFS)=0.45 (0.12–0.78) r(1y OS)=0.93 (0.84–0.96) r(2y OS)=0.79 (0.51–0.91) between treatment effects on SE and OS r(1-y OS)=–0.86 (0.3–0.97) r(2-y OS)=–0.83 (0.07–0.97)
		Mix of ICI	(Petrelli et al., 2016)	article	aggregated	13	3,373		OS milestone (1-y) OS milestone (2-y)	
		Oncolytic virus (talimogene laherparepvec)	(Kaufman et al., 2017)	article	individual patient-level	NA	436	Cox model	durable response rate	durable responders vs non-durable responders: at 9 months: HR=0.07; (95% CI [0.01–0.48]) at 12 months: HR=0.05 ; (95% CI [0.01–0.33]) at 18 months: HR=0.11; (95% CI [0.03–0.44])
		Anti-PD-1	(Nishino et al., 2017a)	article	individual patient-level	NA	96	Cox model	tumour burden increase from baseline	patients with <20% tumour burden increase during therapy had significantly reduced risk of death (HR = 0.19 95% CI [0.08–0.43])
		ICI ICI	(Chen, 2015) (Izar et al., 2017)	article article	NA NA	NA NA	NA NA	NA NA	OS milestone rate OS milestone rate PFS milestone rate Treatment-free interval Treatment-free survival	General discussion General discussion
Non-small cell lung cancer	metastatic / advanced	Mix of treatment (including ICI)	(Blumenthal et al., 2017)	article	aggregated	25	20,013	weighted linear regressions	12-month OS milestone ratio 9-month OS milestone ratio 9-month PFS milestone ratio 6-month ORR milestone ratio	R ² =0.80 (95% CI [0.63–0.91]) R ² =0.67 (95% CI [0.49–0.82]) R ² =0.19 (95% CI [0.03–0.49]) R ² =0.04 (95% CI [0.0002–0.28])
		Anti-PD(L)-1	(Shukuya et al., 2016)	article	aggregated	10	NR	Spearman correlation coefficient	median PFS Response rate	r = 0.473 r = 0.452
		Mix of treatment (including ICI)	(McCoach et al., 2017)	article	individual patient-level	4	660	Cox model	DepOR	For PD-1 treated patients: tumour shrinkage>50% is associated with OS improvement
		Mix of treatment (including ICI) Anti-PD-1	(Saad and Buyse, 2017) (Kazandjian et al., 2015)	article abstract	NA individual patient-level	NA 1	NA 117	NA Cox model	DepOR objective response rate	Discussion article on the article of McCoach et al. patients who achieved a best response of complete or partial response had the longest survival.
		Anti-PD-1	(Nishino et al., 2017b)	article	individual patient-level	NR	160	Cox model	tumour burden increase from baseline	patients with <20% tumour burden increase during therapy had significantly reduced hazards of death (HR = 0.24 ; p < 0.001)

(continued on next page)

Table 1 (continued)

Type of cancer	Treatment setting	Type of treatment	Authors	Type of publication	Type of data	Number of trials	Number of patients	Method to assess the association	Alternative clinical endpoints	Results	
		Anti-PD-1	(Nishino et al., 2017c)	abstract	individual patient-level	NR	134	Cox model	tumour burden increase from baseline	median OS when $\geq 50\%$ increase of tumour burden = 12.7 months [8.5–14.7] / median OS when $< 50\%$ = 4.5 [1.3–4.9] logrank test p-value = 0.0003 Tumour burden increase of 50% at 8 weeks of therapy was associated with significantly shorter OS	
Urinary tract cancer	metastatic	Anti-PD(L)-1	(Abdel-Rahman, 2018)	article	aggregated	13	2,792	Spearman or Pearson correlation coefficient	Overall response rate PFS OS milestone (1 y)	Uro. carcinoma: $r = -0.120$; $p = 0.758$; RCC: $r = -0.397$; $p = 0.440$ Uro. carcinoma: $r = -0.024$; $p = 0.955$; RCC: $r = 0.394$; $p = 0.440$ Uro. carcinoma: $r = 0.806$; $p = 0.016$; RCC: $r = 0.800$; $p = 0.104$	
Multi-tumours	Advanced	Anti-CTLA-4 Anti-PD(L)-1	(Ritchie et al., 2018)	article	aggregated	20	10,828	weighted linear regression, correlation coefficient	ORR PFS 6-month PFS	$r = 0.57$ (95% CI [0.23–0.89]) $r = 0.42$ (95% CI [0.04–0.81]) $r = 0.55$ (95% CI [0.14–0.92]) $R^2 = 0.47$ (95% CI [0.03–0.77])	
	Metastatic	Anti-CTLA4 Anti-PD(L)-1	(Roviello et al., 2017)	article	aggregated	17	8,994	weighted linear regression	Response rate	IME responders vs non responders: HR = 0.09 (95% CI [0.07–0.11]) $R^2 = 0.68$ (95% CI [0.02–0.91]) $r = NA$; $R^2 = 0.1277$ $r = 0.61$; $R^2 = 0.1303$ $r = 0.60$; $R^2 = 0.07$ to 0.10	
		Anti-PD(L)-1	(Gao et al., 2018)	article	individual patient-level	9	5,806	2-stage approach: - ILS: Cox model - TLS: weighted linear regression (R^2)	IME		
			Anti-PD(L)-1	(Mushti et al., 2018)	article	individual patient-level	13	6,722	2-stage approach: - ILS: Cox model, Spearman correlation coefficient (r) - TLS: weighted linear regression (R^2)	ORR PFS modified PFS	
			Anti-PD(L)-1	(Korn and Freidlin, 2018)	article	NA	NA	NA	NA	ORR PFS Modified PFS	Discussion article on the article of Mushti et al.
			Anti-PD(L)-1 Anti-CTLA-4	(Kaufmann et al., 2017)	abstract	aggregated	18	7,140	weighted regression model	ORR DCR PFS	anti-CTLA4: anti-PD(L)1: $R^2 = 0.016$ $R^2 = 0.066$ $R^2 = 0.160$ $R^2 = 0.038$ $R^2 = 0.000$ $R^2 = 0.432$
			Mix of treatment	(Tan et al., 2017)	article	aggregated	51	NR	Spearman correlation coefficient, weighted linear regression model	PFS	$r = 0.62$ $R^2 = 0.38$
		Anti-PD-1	(Cyawali et al., 2017)	abstract	aggregated	9	NR	correlation coefficient regression model	PFS	$r = 0.676$ $R^2 = 0.457$	

CI: confident interval; DepOR: depth of response; DCR: disease control rate; HR: hazard ratio; HRQoL: health-related quality of life; ICI: immune-checkpoint inhibitors; ILS: individual-level surrogacy; IME: intermediate response endpoint; NA: not applicable; NR : not reported ; ORR: objective response rate; OS: overall survival; PFS: progression-free survival; r: correlation coefficient; RCC : renal cell carcinoma; RFS: recurrence-free survival; R2: determination coefficient ; SE: surrogate endpoint ; TLS: trial-level surrogacy; Uro: urothelial Y: year.

Table 2

Summary of the key characteristics of the publications (N = 20 research studies) assessing association between a clinical endpoint and overall survival in ICI-treated patients.

Format of the communication, n (%)	
Article	14 (70)
Conference abstract	6 (30)
Type of cancer, n (%)	
Pooled cancer types	7 (35)
Non-small cell lung cancer	6 (30)
Melanoma	6 (30)
Urinary tract cancer	1 (5)
Type of treatment, n (%)	
anti-PD-1 / anti-PDL-1 / anti-CTLA-4	8 (40)
anti-PD-1 or anti-CTLA-4 alone	6 (30)
Mix of treatment types (including ICI)	4 (20)
Interferon	1 (5)
Oncolytic virus	1 (5)
Treatment setting, n (%)	
Advanced / metastatic	19 (95)
Adjuvant	1 (5)
Type of data, n (%)	
Aggregated data	10 (50)
Individual patient data	10 (50)
Statistical method[†], n (%)	
Weighted linear regression model	8 (40)
Spearman or Pearson correlation coefficient	5 (25)
Cox model	7 (35)
Two-stage approach (individual-level surrogacy and trial-level surrogacy)	3 (15)
Alternative clinical endpoints[†], n (%)	
<i>Time-to-event endpoints</i>	
Progression-free survival	10 (50)
OS milestone rate/ratio	3 (15)
Recurrence-free survival	1 (5)
<i>Categorical or continuous endpoints</i>	
Objective response rate	8 (40)
Disease control rate	1 (5)
Tumour burden increase from baseline	3 (15)
Depth of response	1 (5)
Durable response rate	1 (5)
Intermediate response endpoint	1 (5)
Health-related quality of life	1 (5)

[†] Multiple approaches or endpoints could be considered in the same study.

Most of the research studies were peer-reviewed articles (70%) focusing on a single cancer type (65%). Studies were primarily performed on pooled ICIs (40%) or on anti-PD-1 or anti-CTLA-4 only (30%), almost exclusively in the advanced/metastatic setting (95%). One study was performed on resected stage II–III melanoma interferon-treated patients and the predictive model was applied to an ICI-treated population, in the same disease setting. (Suciú et al., 2018)

For half of the publications, analyses on the association with OS were performed on individual patient data (N = 10). From a methodological perspective, multiple statistical approaches were considered, depending on the type of data available. Of the ten references presenting analyses performed on individual patient data, one third evaluated the association between the alternative endpoint and OS at both the patient and trial levels (meta-analytic approach). Only one of them used a two-stage approach with a bivariate joint distribution function. (Suciú et al., 2018)

Association between alternative endpoints and OS has been assessed for ten different types of clinical endpoints, classified as “time-to-event” or “categorical or continuous” endpoints. Definition of each alternative criterion is presented in Table 3. Half of the publications analysed the association between PFS and OS (N = 10). Correlation coefficients (r) between PFS and OS ranged from -0.024 to 0.90. Determination coefficients, while performed, were also low (R² ≤ 0.38). For OS

milestone rate, defined as the Kaplan-Meier survival probability at a time point defined *a priori* (milestone rate: ratio of them between treatment arms), a moderate to strong association was highlighted in three publications (r: from 0.79 to 0.93; R²: from 0.67 to 0.93) (Petrelli et al., 2016; Izar et al., 2017; Abdel-Rahman, 2018).

Among the categorical or continuous alternative endpoints, objective response rate (ORR) was the most frequently analysed (N = 8). Similarly to PFS, associations between ORR and OS were weak (r ≤ 0.57; R² ≤ 0.47). Disease control rate (DCR), defined as the sum of complete or partial response and stable disease, was tested in only one study (Kaufmann et al., 2017). No association between DCR and OS was highlighted (R² ≤ 0.16). Finally, two new composite endpoints, especially designed for immunotherapy treatment were identified. Durable response rate (DRR), a combination of standard response criteria and a prospective duration dimension of 6 months, was highly associated with OS (Kaufman et al., 2017). In melanoma patients, achieving a durable response was associated with a statistically significant improvement in OS in all the landmark analyses (Hazard ratio (HR) ranged from 0.05 to 0.11). Another more complex binary composite endpoint, called “Intermediate response endpoint” (IME) has also been described (Gao et al., 2018). The IME response is defined on the basis of non-target lesion progression, new lesion, and target lesion information determined by baseline tumour burden, tumour reduction depth, and tumour change dynamic within one year after randomisation (Table 3). At the patient-level, IME responders had improved OS compared with non-responders (HR = 0.09 (95% CI [0.07–0.11])). At the trial-level, association between OS and IME was moderate (R² = 0.68 between the

HR for OS and the odds ratio for IME).

In the four review articles selected, (Chen, 2015; Izar et al., 2017; Saad and Buyse, 2017; Korn and Freidlin, 2018) only two additional alternative endpoints were discussed for patients with advanced melanoma, namely treatment-free interval (TFI) and treatment-free survival (TFS). TFI, defined as the time since treatment interruption to disease progression and the need for further treatment, is used for the management of hematologic malignancies.

4. Discussion

To our knowledge, this is the first systematic literature review of alternative endpoints associated with OS in ICI-treated patients with cancer. Few references were identified (N = 24), even though the search algorithms were broad. This might be explained by the limited data available from RCT of ICIs due to the recent approval of these drugs. In light of the large number of publications identified on the prognostic role of PD-(L)1 or CTLA-4 expression (N = 858), research in this field has focused primarily on identification of subsets of responders rather than on surrogacy for OS. From our perspective, the surrogacy issue deserves as much interest, and could definitely be addressed by at least reporting specific data on OS and alternative endpoints.

Traditional alternative endpoints (PFS, ORR) as well as new promising composite ones (DRR, IME) were retrieved. PFS was the most studied alternative endpoint in ICI-treated patients with cancer. Historically, this has been demonstrated to be a valid SE for OS in some tumour types for other drug classes (Savina et al., 2018). Overall, we noticed a weak correlation between PFS and OS in ICI-treated patients. In a meta-analysis on the association between PFS and OS in the same population, no correlation between OS and PFS was found in terms of medians or gains in medians (Gyawali et al., 2018). Recently, Kaufman et al. concluded as well that PFS is an imperfect surrogate of OS according to the results of their meta-analysis performed on RCT of ICI (Kaufman et al., 2018). Pseudo-progression, even though relatively rare for certain tumour types (Wang et al., 2018), may partially explain this weak correlation between PFS and OS. Indeed, the effect of ICI as opposed to chemotherapy, is not on tumour cells, but on immune cells.

Table 3

Definition of the clinical endpoints identified in the systematic literature review.

Clinical alternative endpoint	Definition
<i>Time to event endpoints</i>	
Progression-free survival (PFS)	Time from randomization until the minimum between a disease progression or death (any cause)
OS Milestone rate	Kaplan-Meier survival probability at a time point defined a priori
OS Milestone ratio	Ratio of milestone rates between 2 treatment arms
Treatment-free interval (TFI)	Time since treatment interruption to disease progression and the need for re-initiation of the same treatment or initiation of another therapy
Treatment-free survival (TFS)	Time from end of therapy until need for next line treatment or death (the minimum)
Recurrence-free survival (RFS)	Time from randomization until recurrence of tumour or death from any cause (the minimum)
<i>Categorical or continuous endpoints</i>	
Objective response rate (ORR)	Proportion of patients with tumour size reduction of a predefined amount and for a minimum time period. ORR is composed as the sum of partial plus complete responses.
Disease control rate (DCR)	Proportion of patients with partial or complete responses to therapy or in stable disease and for a minimum time period
Tumour burden increase from baseline	Proportion of patients with < X% tumour burden increase from baseline, after a specific period of treatment and at a specific landmark time point
Depth of response (DepOR)	Percent tumour shrinkage at nadir, in comparison with baseline
Durable response rate (DRR)	Continuous response (complete response or partial response) beginning in the first 12 months of treatment and lasting 6 months or longer
Intermediate response endpoint (IME)	IME response is a binary endpoint (response or non-response) defined as satisfying all of the three following criteria: 1. The patient needed to be a target lesion responder, meaning the patient's target lesion score was less than an optimal cut-off value; 2. The patient had no unequivocal non-target lesion progression as determined per RECIST 1.1 criteria within 1 year; 3. The patient had no new unequivocal lesion as determined per RECIST 1.1 criteria within 1 year
Health-related quality of life (HRQoL)	Individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person's physical health, psychological state, personal beliefs, social relationships and their relationship to salient features of their environment (<i>World Health Organization definition</i>).

Being treated by ICI, some patients experience immune-related responses such as initial increase in the size of tumours or appearance of new lesions, before a subsequent and sustained reduction in tumour burden occurs. This immunotherapy-specific phenomenon has stimulated the development of immune-related response criteria, the iRECIST (Seymour et al., 2017). Another explanation might be the residual efficacy of ICI for a longer duration (delayed treatment effect), these drugs affecting OS more than PFS even after treatment discontinuation (Gyawali et al., 2018). The poor correlation between ORR or DCR and OS may be due to these ICI-specific phenomena as well. Lastly, potential information provided by other known metrics, such as tumour burden increase, depth of response, health-related quality of life or treatment-free interval merit further investigation in this population. To date, few studies have investigated these alternative endpoints.

This work also highlights new clinical criteria, such as OS milestone rate/ratio, DRR and IME, not assessed for previous generations of therapy. A moderate to strong correlation between 1-year OS milestone ratio and OS was observed in metastatic non-small cell lung cancer patients (Blumenthal et al., 2017). Unfortunately, several type of treatments were aggregated in this study due to the low number of RCT of individual ICIs, which precluded formal validation of this endpoint as an SE. Notably, Petrelli et al. focused on ICI-treated melanoma patients and described a strong correlation between 1-year OS milestone rate and median OS (Petrelli et al., 2016). However, we face methodological concerns such as the impossibility of assessing surrogacy at the patient-level with the milestone rate, which is an aggregate measure. For this reason, this metric cannot be considered an adequate candidate SE. Moreover, even though milestone rates may present some advantages, another major limitation lies in the challenge of selecting the optimal milestone time point which may differ between cancer types and tumour stage (Hoss et al., 2013).

Based on the current published evidence, new composite endpoints, such as DRR and IME, statistically better correlate with OS in the metastatic cancer types considered. It seems to be more appropriate for capturing the unique pattern of antitumor response and survival with ICI treatments in the advanced setting. The main advantage of the suggested exploratory metrics is the inclusion of a duration component, which allows the persistence of the response to be taken into account, even though pre-specification of the time dimension is challenging. It also necessitates the use of specific statistical methods such as landmark

analysis. However, duration of response should not be considered as the sole criterion for surrogacy. Emens et al. have argued that a composite endpoint which includes ORR and duration of response might best predict the effect of immunotherapy on long-term survival (Emans et al., 2017). Recently, Pfeiffer et al. considered both ORR and duration of response in their surrogacy assessment for OS in advanced non-small cell lung cancer patients (Peiffer et al., 2018). Combination of ORR and duration of response performed better as a surrogate for OS than duration of response alone.

Considerable heterogeneity in the statistical approaches used to assess surrogacy was noted in the studies evaluated. Only three research studies applied a statistical method for evaluating both patient-level and trial-level surrogacy. This two-stage approach is the only technique to validate an SE adequately, and is most robust when performed with a joint distribution function, as in the study of Suciú et al. (Shi et al., 2011; Suciú et al., 2018). This method ensures that correlation between endpoints measured on the same patient is taken into account. Only half of the studies identified in this review considered individual patient data. Difficulty in gaining access to individual patient data might be the main reason for not considering this type of analysis. Another major limitation of the available research is the pooling of tumour types, since validation of an SE should be performed within a given disease setting and for a specific drug class with a shared mechanism of action (Burzykowski et al., 2005). Thus, based on the literature published and from a methodological point of view, no validated SEs for OS are available today for the study of ICIs in clinical trials.

This study presents some limitations inherent to any systematic literature review. Publication bias stressed by Moher et al. may occur in this type of review as non-significant associations are less likely to be reported (Moher et al., 2009). Unfortunately, the impact of such bias could not be assessed. Two databases were consulted and one can assume that some references might not be included in these databases. We selected these databases as they are the two largest medical ones. Embase has the additional attractive characteristic to include major oncology conference proceedings (such as ESMO and ASCO). Keeping in mind these limitations, this work however is a valuable step forward in the context of assessing surrogacy in ICI-treated patients. To our knowledge, we have provided the first comprehensive list of alternative endpoints analysed for their association with OS in ICI-treated

populations, relying on a strong methodological approach. Results were reported according to international guidelines (Moher et al., 2009), and we did not restrict our investigations to primary research studies, but also considered general discussions of alternative endpoints. Repeating this study in upcoming years would be of interest, when more clinical studies of ICIs have been performed in the same or other cancer types and published.

As previously noted, endpoints other than OS are essential for regulatory approval of anticancer agents and it is of importance to identify novel surrogate for efficacy for ICI (Kaufman et al., 2018; Saad and Buyse, 2016). However, based on the current literature published, there is no sufficient data to support validated SE for OS. In ICI-treated patients at the metastatic stage, adequate surrogacy assessment of promising composite endpoints which take into account a duration component is encouraged.

Funding

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

Declaration of interests

SB and AFG are employees of Bristol-Myers Squibb.

CB declares no competing interests.

AI has advisory board consulting with Epizyme, Lilly, Merck Sharp & Dohme, Novartis, Pharmamar and Roche. He has received research grants from Astra-Zeneca, Bayer, Bristol-Myers Squibb, Chugai, Merck Sharp & Dohme, Novartis, Pharmamar, Pfizer and Roche.

DR declares activity as independent methodological expert for Servier Laboratories.

VR declares activities as independent methodological expert for advisory boards for AstraZeneca, GlaxoSmithKline, Novartis and Medtronic.

Acknowledgments

We wish to sincerely thank Adam Doble from Foxymed company for having reviewed this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.critrevonc.2019.02.013>.

References

Abdel-Rahman, O., 2018. Surrogate end points for overall survival in trials of PD-(L)1 inhibitors for urinary cancers: a systematic review. *Immunotherapy* 10 (2), 139–148.

Alonso, A., Bigirimurame, T., Burzykowski, T., et al., 2017. Applied Surrogate Endpoint Evaluation Method with SAS and R. CRC Press, Boca Raton.

Blumenthal, G.M., Zhang, L., Zhang, H., et al., 2017. Milestone analyses of immune checkpoint inhibitors, targeted therapy, and conventional therapy in metastatic non-small cell lung cancer trials: a meta-analysis. *JAMA Oncol.* 3 (8), e171029. <https://doi.org/10.1001/jamaoncol.2017.1029>.

Burzykowski, T., Molenberghs, G., Buyse, M., 2005. *The Evaluation of Surrogate Endpoints*. Springer-Verlag, New-York.

Chen, T.-T., 2015. Milestone survival: a potential intermediate endpoint for immune checkpoint inhibitors. *J. Natl. Cancer Inst.* 107 (9). <https://doi.org/10.1093/jnci/djv156>.

Emans, L.A., Ascierto, P.A., Darcy, P.K., et al., 2017. Cancer immunotherapy: opportunities and challenges in the rapidly evolving clinical landscape. *Eur. J. Cancer* 81, 116–129.

Flaherty, K.T., Lee, S.J., Dummer, R., et al., 2013. A meta-analysis of randomized, controlled trials in metastatic melanoma establishes progression-free survival as a surrogate for overall survival. *Eur. J. Cancer* 49 (2), S856.

Fleming, T.R., Powers, J.H., 2012. Biomarkers and surrogate endpoints in clinical trials. *Stat. Med.* 31 (25), 2973–2984.

Fujii, T., Naing, A., Rolfo, C., Hajjar, J., 2018. Biomarkers of response to immune

checkpoint blockade in cancer treatment. *Crit. Rev. Oncol. Hematol.* 130, 108–120.

Galluzzi, L., Vacchelli, E., Bravo-San Pedro, J.-M., et al., 2014. Classification of current anticancer immunotherapies. *Oncotarget* 5 (24), 12472–12508.

Gao, X., Zhang, L., Sridhara, R., 2018. Exploration of a novel intermediate response endpoint in immunotherapy clinical studies. *Clin. Cancer Res.* 24 (10), 2262–2267.

Gyawali, B., Shimokata, T., Ando, Y., 2017. Correlation and differences in effect sizes between progression free survival (PFS) and overall survival (OS) among PD-1 inhibitors. *Ann. Oncol.* 28 (55), v403–v427.

Gyawali, B., Hey, S.P., Kesselheim, A.S., 2018. 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 Network Open* 1 (2), e180416.

Hoss, A., Topalian, S., Chen, T.-T., et al., 2013. Facilitating the development of immunotherapies: intermediate endpoints for immune checkpoint modulators. Conference on Clinical Research. Available from: https://www.pharma-medtechbi.com/?/media/Supporting%20Documents/The%20Pink%20Sheet/75/46/FOCR_immunotherapy_issue_brief.pdf.

Izar, B., Regan, M.M., McDermott, D.F., 2017. Clinical trial design and endpoints for stage IV melanoma in the modern era. *Cancer J.* 23 (1), 63–67.

Kaufman, H.L., Andtbacka, R., Collichio, F.A., et al., 2017. Durable response rate as an endpoint in cancer immunotherapy: insights from oncolytic virus clinical trials. *J. Immunother. Cancer* 5 (72). <https://doi.org/10.1186/s40425-017-0276-8>.

Kaufman, H., Schwartz, L., William Jr, W., et al., 2018. Evaluation of classical endpoints as surrogates for overall survival in patients treated with immune checkpoint blockers: a systematic review and meta-analysis. *J. Cancer Res. Clin. Oncol.* <https://doi.org/10.1007/s00432-018-2738-x>.

Kaufmann, H., Schwartz, L.H., William, W.N., et al., 2017. Evaluation of clinical endpoints as surrogates for overall survival in patients treated with immunotherapies. *J. Clin. Oncol.* 35 (15s), e14557.

Kazandjian, D., Blumenthal, G., Khozin, S., et al., 2015. An exploratory responder analysis of best RECIST response and survival in patients with metastatic squamous NSCLC treated with nivolumab. *J. Thorac. Oncol.* 10 (9), S234.

Korn, E.L., Freidlin, B., 2018. Surrogate and intermediate endpoints in randomized trials: What's the goal? *Clin. Cancer Res.* 24 (10), 2239–2240.

Lee, D., Harvey, B., Gaudin, A.-F., et al., 2013. An exploratory analysis of the association between EORT C-QLQ C30 domains and progression free/overall survival in advanced melanoma after 12 weeks of treatment on ipilimumab compared to GP100 in a Phase III clinical trial. *Value Health* 16 (7), A422.

Martin-Liberal, J., Hierro, C., Ochoa de Olza, M., Rodon, J., 2017. Immuno-oncology: the third paradigm in early drug development. *Target. Oncol.* 12 (2), 125–138.

Mathoulin-Pelissier, S., Gourgou-Bourgade, S., Bonnetain, F., Kramar, A., 2008. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J. Clin. Oncol.* 26 (22), 3721–3726.

McCoach, C.E., Blumenthal, G.M., Zhang, L., et al., 2017. 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.* 28 (11), 2707–2714.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLoS Med.* 6 (7), e1000097.

Mushti, S.L., Mulkey, F., Sridhara, R., 2018. Evaluation of overall response rate and progression-free survival as potential surrogate endpoints for overall survival in immunotherapy trials. *Clin. Cancer Res.* 24 (10), 2268–2275.

Nishino, M., Giobbie-Hurder, A., Manos, M.P., et al., 2017a. Immune-related tumor response dynamics in melanoma patients treated with pembrolizumab: identifying markers for clinical outcome and treatment decisions. *Clin. Cancer Res.* 23 (16), 4671–4679.

Nishino, M., Dahlberg, S.E., Adeni, A.E., et al., 2017b. Tumor response dynamics of advanced non-small cell lung cancer patients treated with PD-1 inhibitors: imaging markers for treatment outcome. *Clin. Cancer Res.* 23 (19), 5737–5744.

Nishino, M., Adeni, A., Lydon, C., et al., 2017c. Spectrum of early progression in advanced NSCLC patients treated with PD-1 inhibitors: identifying markers for poor outcome. *J. Thorac. Oncol.* 12 (11), S1915–16.

Ocana, A., Tannock, I.F., 2001. When are “positive” clinical trials in oncology truly positive? *JNCI* 103, 16–20.

Pazdur, R., 2008. Endpoints for assessing drug activity in clinical trials. *Oncologist* 13 (S2), 19–21.

Peiffer, B.M., Kulakova, M., Hashim, M., Heeg, B., 2018. Can duration of response be used as a surrogate endpoint for overall survival in advanced non-small cell lung cancer? *J. Clin. Oncol.* 36 (15), S9082.

Petrelli, F., Coinu, A., Cabiddu, A., et al., 2016. Early analysis of surrogate endpoints for metastatic melanoma in immune checkpoint inhibitor trials. *Medicine* 95 (26), e3997.

Ritchie, G., Gasper, H., Man, J., et al., 2018. Defining the most appropriate primary end point in phase 2 trials of immune checkpoint inhibitors for advanced solid cancers. A systematic review and meta-analysis. *JAMA Oncol.* 4 (4), 522–528.

Rolfo, C., Caglevic, C., Santaripa, M., et al., 2017. Immunotherapy in NSCLC: a promising and revolutionary weapon. *Adv. Exp. Med. Biol.* 995, 97–125.

Roviello, G., Andre, F., Venturini, S., et al., 2017. Response rate as a potential surrogate for survival and efficacy in patients treated with novel immune checkpoint inhibitors: a meta-regression of randomised prospective studies. *Eur. J. Cancer* 86, 257–265.

Saad, E.D., Buyse, M., 2016. Statistical controversies in clinical research: end points other than overall survival are vital for regulatory approval of anticancer agents. *Ann. Oncol.* 27 (3), 373–378.

Saad, E.D., Buyse, M., 2017. 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.* 28 (11), 2629–2630.
- Savina, M., Gourgou, S., Italiano, A., et al., 2018. Meta-analyses evaluating surrogate endpoints for overall survival in cancer randomized trials: a critical review. *Crit. Rev. Oncol. Hematol.* 123, 21–41.
- Seymour, L., Bogaerts, J., Perrone, A., et al., 2017. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol.* 18 (3), e143–52.
- Shi, Q., Renfro, L.A., Bot, B.M., Burzykowski, T., Buyse, M., Sargent, D.J., 2011. Comparative assessment of trial-level surrogacy measures for candidate time-to-event surrogate endpoints in clinical trials. *Comput. Stat. Data Anal.* 55, 2748–2757.
- Shukuya, T., Mori, K., Amann, J.M., et al., 2016. Relationship between overall survival and response or progression-free survival in advanced non-small cell lung cancer patients treated with anti-PD-1/PD-L1 antibodies. *J. Thorac. Oncol.* 11 (11), 1927–1939.
- Suciu, S., Eggermont, A., Lorigan, P., et al., 2018. Relapse-Free Survival as a surrogate for overall survival in the evaluation of stage II–III melanoma adjuvant therapy. *J. Natl. Cancer Inst.*
- Tan, A., Porcher, R., Crequit, P., Ravaud, P., Dechartres, A., 2017. Differences in treatment effect size between overall survival and progression-free survival in immunotherapy trials: a meta-epidemiologic study of trials with results posted at ClinicalTrials.gov. *J. Clin. Oncol.* 35 (15), 1686–1694.
- Wang, Q., Gao, J., Wu, X., 2018. Pseudoprogression and hyperprogression after checkpoint blockade. *Int. Immunopharmacol.* 58, 125–135.

2.3 Additional discussion of the results from the systematic literature review

Based on the results from the systematic literature review, two promising composite metrics, namely DRR and IME were identified as interesting alternative endpoints. We discussed hereafter the pro and cons of these two new alternative endpoints.

2.3.1 Durable response rate endpoint (DRR)

Kaufman et al. studied the DRR endpoint in patients with metastatic melanoma treated with oncolytic virus.⁴⁹ In their conclusion, they suggested that this endpoint could be a relevant surrogate endpoint for patients treated with ICI. DRR is defined as the rate of patients with a continuous response (complete response or partial response as per RECIST criteria V1.1) beginning in the first 12 months of treatment and lasting 6 months or longer. Figure 3 illustrates the design of the endpoint.

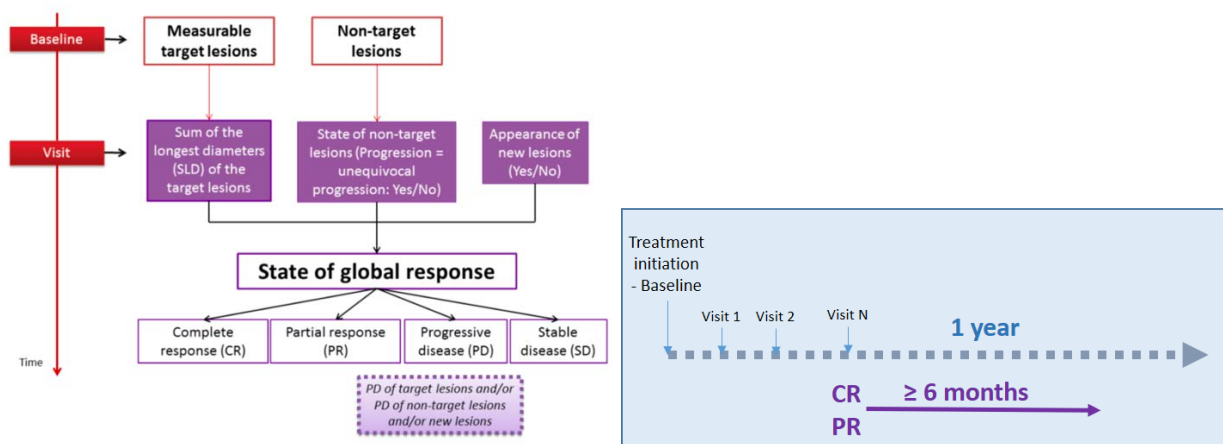


Figure 3. Illustration of the concept of durable response endpoint.

In the study by Kaufman et al., the evaluation of the surrogate properties of DRR was carried out by using a Cox model and a landmark analysis in order to take into account the lead-time bias. It is of importance to account for lead-time bias in the comparison of the time since randomization to an event between groups classified according to a characteristic appearing during follow-up (for example the response).⁵⁰ Figure 4 illustrates lead-time bias (called also guarantee time) in the frame of patients classified according to transplantation.

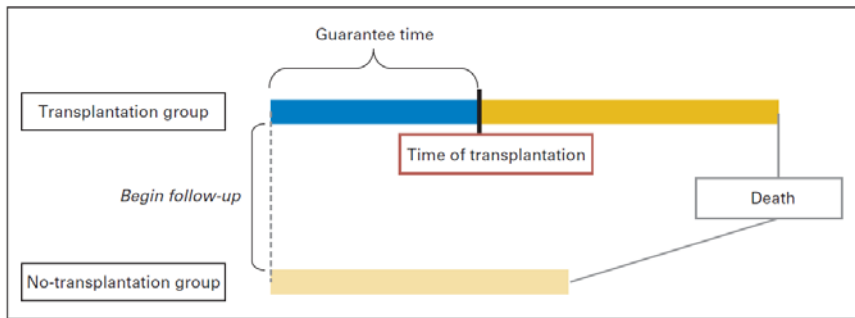


Fig 1. Impact of guarantee time on survival according to Gail³ transplantation model. Patients who undergo transplantation seem to have longer survival times. However, patients in the transplantation group must survive at least until donor is available (blue), and their survival reflects total time before and after transplantation (blue and gold). More favorable survival does not result from the effect of transplantation but rather from the inclusion of extra waiting time (guarantee time). By contrast, the no-transplantation group includes all patients who died before a suitable donor could be found.

Figure 4. Illustration of the concept of lead-time bias (guarantee time).

The analysis at a given time t_x after the randomization t_0 (landmark analysis) consists in the analysis of the patients alive at t_x only (patients who died between t_0 and t_x are not included in the analysis). According to the results of the study by Kaufman et al., patients with a durable response had a significantly higher probability of overall survival compared to those without a durable response (HR = 0.07; 95% CI [0.01; 0.48] at the 9-month landmark).

However, this statistical approach is not the recommended one for evaluating the surrogate properties of a clinical endpoint, the recommended approach being the meta-analytical approach.⁸ After the systematic literature review, it was therefore planned to study this criterion using this statistical method. Before performing the appropriate statistical analyses, the properties of DRR were studied.

In order to use an SE in a RCT, it is necessary that this endpoint can be analyzed for the intention-to-treat (ITT) population. However, in order to take into account the lead-time bias due to the evaluation of the tumor response after randomization, a landmark analysis must be performed for the DRR. This analysis does not take into account the ITT population, and would provide information only for a subgroup of patients, and may significantly reduce the sample size.

Moreover, according to Kaufman et al., the duration of the response must be at least 6 months to be considered in durable response. The vast majority of patients who respond to ICI, especially those with advanced melanoma, have a duration of response lasting more than 6 months.⁵¹ The DRR would then be very close to the ORR, a clinical endpoint which is weakly correlated with OS according to the published evidence of the surrogate properties of ORR in ICI-treated patients, especially in advanced melanoma.⁵²

Therefore, although this criterion appeared promising during the systematic literature review, **it does not seem to fulfill the requested properties for being a candidate SE in this specific population.**

2.3.2 Intermediate response endpoint (IME)

The IME is a composite clinical endpoint developed by FDA team members. IME is defined on the basis of non-target lesion progression, new lesion, and target lesion information determined by baseline tumor burden, tumor reduction depth, and tumor change dynamic within one year after randomization (Figure 5).

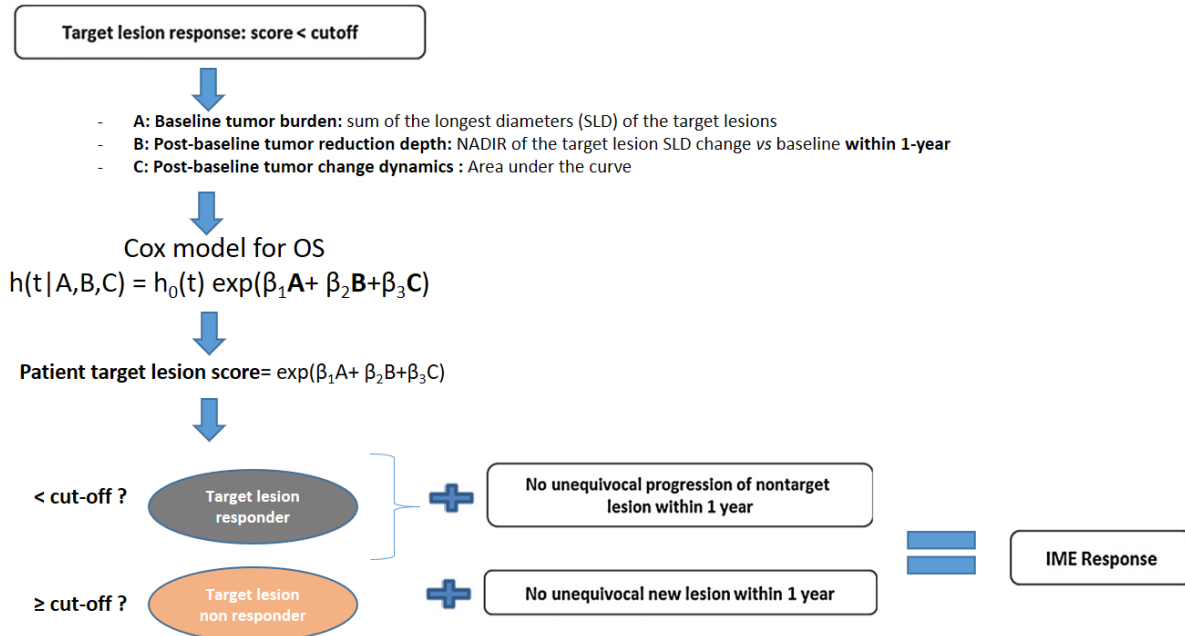


Figure 5. Illustration of the concept of intermediate response endpoint.

Based on the publication of Gao et al., the IME appeared to be a promising SE given the observed association between IME and OS. However, it cannot be described as valid SE based on this work only. Indeed, the validation of a SE should be performed by cancer type while the population of analysis in the work of Gao et al. consisted of patients with different cancer types (melanoma, lung cancer, etc.). At one stage of this doctoral thesis, it was therefore planned to study the surrogate properties of this criterion by cancer type. However, we identified an issue related to the statistical approach used in the construction of the IME.

A Cox model is applied in the construction of the IME. In this model, data related to target lesions are used as explanatory variables. These endogenous variables (i.e. variables associated with a change in the risk of occurrence of the event) can bias the estimate of the treatment effect when they are integrated into a Cox model.⁵³

This methodological issue, as well as the complexity of IME while an SE must be easily measurable⁶, led to the exclusion of IME from the list of relevant candidate SE.

3 Time to next treatment as a candidate surrogate endpoint for overall survival in advanced melanoma patients treated with immune-checkpoint inhibitors

3.1 Introduction

In hematologic malignancies and in metastatic soft-tissue sarcomas, time to next treatment (TNT) has emerged as a relevant alternative clinical endpoint.¹⁷ TNT has also recently been used in breast, colon, and prostate cancer.^{54,55,56} **TNT is defined as the time between baseline (i.e. randomization, inclusion or treatment initiation) and the date of initiation of the subsequent systemic treatment or the date of death whichever occurs first.** One attractive characteristic of TNT is its ability to capture the treatment-free interval (TFI), defined as the time from end of a therapy until the date of initiation of a subsequent line of treatment or death, whichever occurs first.⁵⁷ Figure 6 illustrates the concepts of TNT and TFI.

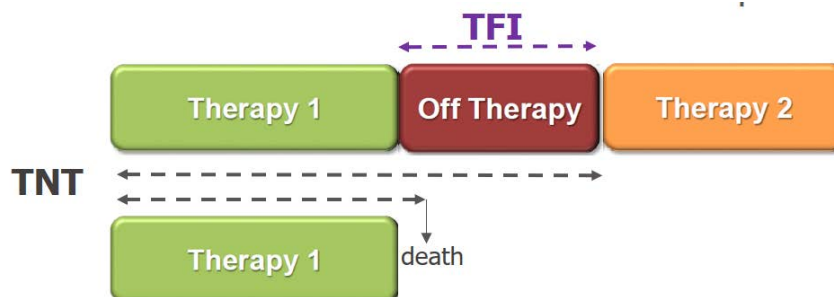


Figure 6. Illustration of the concepts of time to next treatment (TNT) and treatment-free interval (TFI).

Recent studies have found that ICI perform better than non-immunotherapies during the TFI,^{18,19} suggesting a lingering benefit to ICI-treated patients. As such, TNT may better capture the post-treatment benefits with respect to OS among ICI-treated patients. Moreover, by having the initiation of subsequent-line of treatment as an endpoint, TNT will have a greater overlap in the events shared with OS because it will capture all patients dying prior to progression/subsequent-line of anti-cancer treatment (just as PFS would) as well as those who progress but are never treated by a subsequent-line of treatment. This increased number of shared events is likely to lead to a higher correlation. Even though TNT appears to be a relevant clinical endpoint for ICI-treated patients, our systematic literature review highlighted that data on surrogacy analyses based on this alternative endpoint are lacking.⁵⁸ The objective of this work was therefore to formally assess the surrogate properties of TNT in previously untreated advanced melanoma patients treated with ICI. We focused on previously

untreated advanced melanoma patients as the maturity of survival data (i.e. median of OS) may take a significant period of time to be observed in this disease setting, and so identifying an SE for OS is particularly critical. Bristol Myers Squibb granted access to CheckMate 067 RCT individual-level patient data (IPD), briefly described in the following section.

IPD from the 60-month results of the CheckMate 067 RCT were used. Analyses were performed for nivolumab monotherapy (NIVO) or nivolumab with ipilimumab (NIVO+IPI) versus ipilimumab monotherapy (IPI). The SE one-step validation method based on a joint frailty model was used where the country of enrollment was applied to define synthetic clusters. Kendall's τ coefficient of correlation and the coefficient of determination (R^2_{trial}) were estimated for respective measurements of association at the individual and cluster levels. Surrogate threshold effect (STE), the maximum threshold hazard ratio (HR) for TNT that would translate into OS benefit, was estimated. A leave-one-out cross-validation analysis was performed to evaluate model robustness and predictive accuracy.

Fifteen clusters were considered from 945 patients. For both nivolumab-containing arms, the association between TNT 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 STEs were 0.58 for NIVO versus IPI and 0.39 for NIVO+IPI versus IPI. Cross-validation results showed minimum variation of the correlation measures and modest predictive accuracy for the model.

Results suggest that TNT may be a valuable SE in previously untreated advanced melanoma patients treated with ICI. Surrogacy analyses considering multiple RCTs of ICI-treated melanoma patients are warranted for confirming our findings once additional trial will become available. This work was submitted to *European Journal of Cancer*.

3.2 Description of CheckMate 067 trial

Study Design

CheckMate 067 is a phase III, randomized, double-blind 3-arm trial of nivolumab monotherapy or nivolumab combined with ipilimumab versus ipilimumab monotherapy in subjects with previously untreated unresectable or metastatic melanoma. Patients were randomly assigned according to a 1:1:1 ratio to receive one of the following regimens: 3 mg of nivolumab per kilogram of body weight every 2 weeks (plus ipilimumab-matched placebo); 1 mg of nivolumab per kilogram every 3 weeks plus 3 mg of ipilimumab per kilogram every 3 weeks for 4 doses, followed by 3 mg of nivolumab per kilogram every 2 weeks for cycle 3 and beyond; or 3 mg of ipilimumab per kilogram every 3 weeks for 4 doses (plus nivolumab-matched placebo). Randomization was stratified according to tumor PD-

L1 status (positive vs. negative or indeterminate), BRAF mutation status (V600 mutation–positive vs. wild-type), and American Joint Committee on Cancer metastasis stage (M0, M1a, or M1b vs. M1c). Treatment continued until disease progression (as per RECIST, version 1.1), development of unacceptable toxic events, or withdrawal of consent. Patients could be treated after progression, provided that they had a clinical benefit and did not have substantial adverse effects, as assessed by the investigator. Enrollment period was from June 2013 to March 2014. At database lock, the minimum follow-up from the date on which the last patient underwent randomization was 60 months.

OS and TNT outcomes

Full patient characteristics and efficacy results were reported in publications from Larkin et al.^{3,59} As it is interesting to observe the joint evolution of the TNT and OS curves, Kaplan-Meier survival curves are displayed for both outcomes in Figure 7 and Figure 8.

Out of the 945 patients included in the study, 70.4% had initiated subsequent systemic anticancer therapy or died at the end of the follow-up. Median TNT was reached approximately three times earlier than median OS for nivolumab monotherapy and ipilimumab (nivolumab: TNT = 12.1 months, 95% CI [8.9; 18.0], OS=36.9 months 95% CI [28.3; 58.7] ; ipilimumab: TNT= 6.2 months 95% CI [5.4; 7.4], OS= 19.9 months 95% CI [16.9; 24.6]); median TNT for nivolumab combined with ipilimumab was 24.2 months (95% CI [16.0; 43.9]), and median OS was not reached. The Kaplan-Meier survival estimates of 5-year TNT in the nivolumab combined with ipilimumab, nivolumab monotherapy, and ipilimumab groups were 42.7%, 32.8%, and 10.9%, respectively.

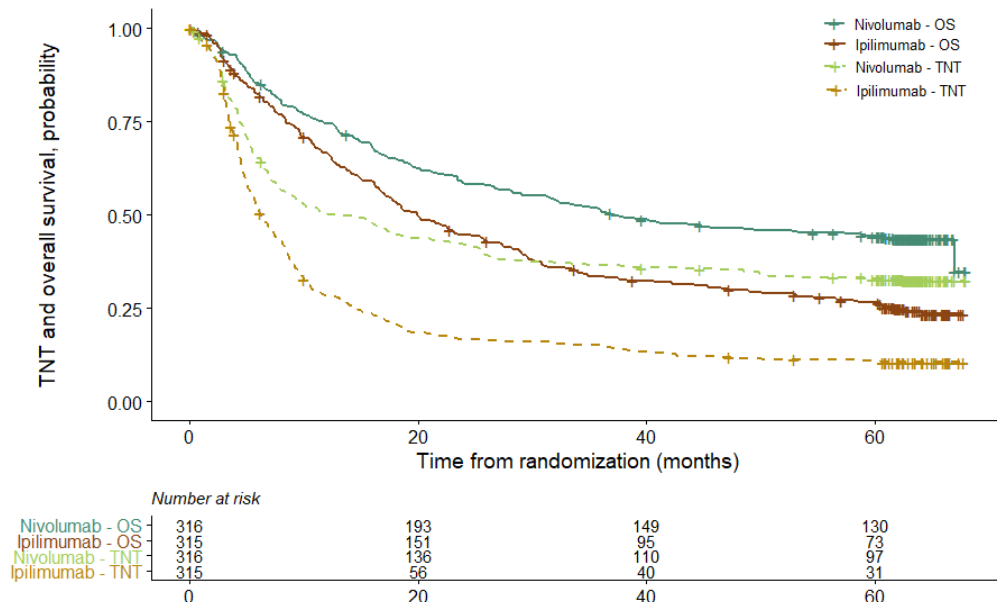


Figure 7. Kaplan-Meier survival curves for OS (solid line) and TNT (dotted line) for nivolumab monotherapy vs. ipilimumab monotherapy. OS: overall survival; TNT: time to next treatment; '+' signs on the Kaplan-Meier curves represent censorings.

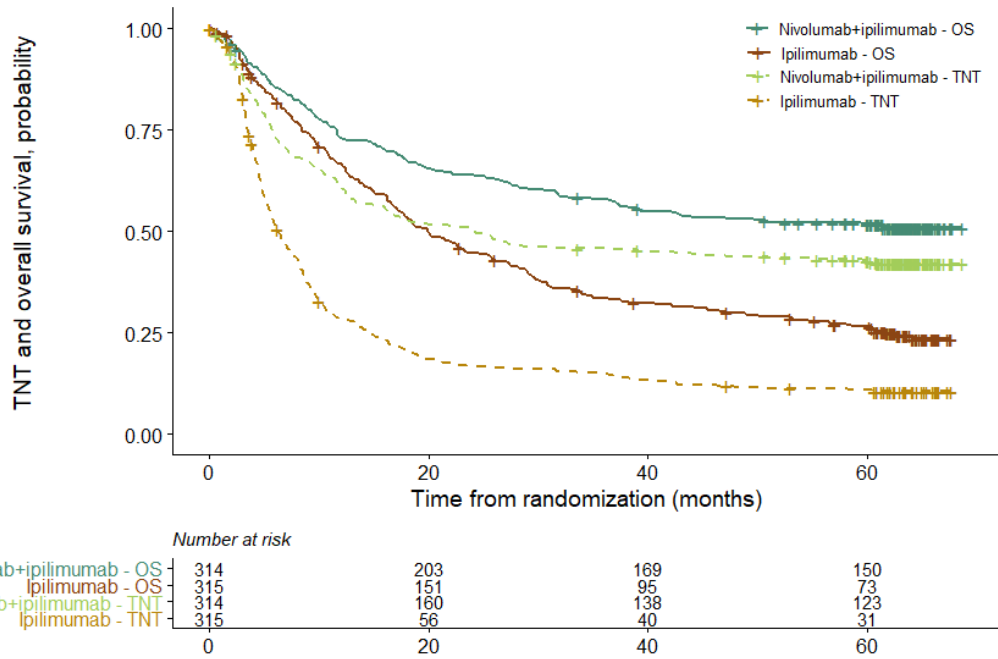


Figure 8. Kaplan-Meier survival curves for OS (solid line) and TNT (dotted line) for nivolumab in combination with ipilimumab vs. ipilimumab monotherapy. OS: overall survival; TNT: time to next treatment; '+' signs on the Kaplan-Meier curves represent censorings.

3.3 Publication

The manuscript is presented in the next pages.

**TIME TO NEXT TREATMENT 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**

Sébastien Branchoux^{1,2,3}, Casimir Ledoux Sofeu², Anne-Françoise Gaudin¹, Murat Kurt⁴, Andriy Moshyk⁴, Antoine Italiano⁵, Carine Bellera^{3,6,7}, Virginie Rondeau^{2,3}

1. Health Economics & Outcomes Research, Bristol Myers Squibb, Rueil-Malmaison, France
2. Biostatistic Team, Bordeaux Population Health Center, ISPED, Centre INSERM U1219, INSERM, Bordeaux, France
3. Epicene Team (Cancer & Environment), Bordeaux Population Health Center, ISPED, Centre INSERM U1219, INSERM, Bordeaux, France
4. Health Economics & Outcomes Research, Bristol Myers Squibb, Lawrenceville, New Jersey, USA
5. Department of Early phase trial unit, Institut Bergonié Comprehensive Cancer Centre, Bordeaux, France
6. Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Center, Bordeaux, France
7. Department of Clinical Epidemiology, INSERM CIC-EC 14.01, Bordeaux, France

Corresponding author

Sébastien Branchoux
Laboratoire Bristol Myers Squibb
3, rue Joseph Monier
92500 Rueil-Malmaison
France
+33 1 58 83 61 11
Sebastien.branchoux@bms.com

Running head: TNT as a surrogate endpoint for OS in advanced melanoma

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

ABSTRACT

PURPOSE: Time to next treatment (TNT) may be a patient-relevant endpoint by capturing durable response and the time free of systemic anticancer therapy, which may be achieved with immune-checkpoint inhibitors (ICI). This study investigated TNT 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 randomized controlled trial (RCT) (NCT01844505) were used. Analyses were performed for nivolumab monotherapy (NIVO) or nivolumab with ipilimumab (NIVO+IPI) versus ipilimumab monotherapy (IPI). The SE one-step validation method based on a joint frailty model was used where the country of enrollment was applied to define synthetic 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. Surrogate threshold effect (STE), the maximum threshold hazard ratio (HR) for TNT that would translate into OS benefit, was estimated. A leave-one-out cross-validation analysis was performed to evaluate model robustness and predictive accuracy.

RESULTS: Fifteen clusters of data were generated from 945 patients. For both nivolumab-containing arms, the association between TNT 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 STEs were 0.58 for NIVO versus IPI and 0.39 for NIVO+IPI versus IPI. Cross-validation results showed minimum variation of the correlation measures and modest predictive accuracy for the model.

CONCLUSION: Results suggest that TNT may be a valuable SE in previously untreated advanced melanoma patients treated with ICI. Surrogacy analyses considering multiple RCTs of ICI-treated melanoma patients are warranted for confirming these findings.

Introduction

Immune-checkpoint inhibitors (ICI) represent one of the most established and clinically promising therapies in oncology¹. These treatments are indicated as monotherapies or in combination with other systemic treatments in more than a dozen types of cancer at different stages such as locally advanced or metastatic, at different treatment settings (adjuvant, palliative), while their clinical development is still expanding. For some advanced stage cancers, such as melanoma and renal cell carcinoma (aRCC), one of the current treatment strategies is to combine different ICIs targeting two distinct receptors like nivolumab (anti-PD1) and ipilimumab (anti-CTLA4).^{2,3}

In oncology, several clinical endpoints have been used to assess treatment efficacy.^{4,5} Overall survival (OS), defined as the time from randomization to the date of death (any cause), is frequently considered as the gold standard for assessing treatment efficacy in randomized controlled trials (RCT) of anticancer therapies from a regulatory perspective.^{6,7,8} This endpoint may require more extensive follow-up and a higher number of patients to demonstrate a clinically and statistically significant relevant difference when comparing therapeutic strategies which in turn imply higher budget requirements due to increased cost of surveillance. In this context, evaluation and validation of surrogate endpoints (SE) are therefore particularly important to expedite patients' access to innovative and potentially life extending medicines. To formalize, an SE is defined as 'a biomarker or an intermediate endpoint intended to substitute for a clinical endpoint'⁴; its validity depends both on the mechanism of action of the treatment and on the disease setting.⁹

A recent systematic literature review summarized the current evidence base on clinical alternative endpoints associated with OS in ICI-treated patients.¹⁰ Current published evidence is inconclusive over validated SE for OS in ICI-treated patients (N=24). Traditional alternative endpoints, such as progression-free survival (PFS) and overall response rate (ORR) based on conventional response criteria (RECIST 1.1) were weakly correlated with OS in this specific population. To name a few, possible reasons behind such a finding were delayed treatment effect, pseudo-progression, and the

medical imaging technique used to monitor disease progression.^{11,12,13,14,15}

In hematologic malignancies and in metastatic soft-tissue sarcomas, time to next treatment (TNT) has emerged as a relevant alternative clinical endpoint.¹⁶ TNT is defined as the time between baseline (i.e. randomization, inclusion or treatment initiation) and the date of next subsequent systemic treatment initiation or the date of death whichever occurs first. Only a handful of studies reported TNT in ICI-treated patients, and the tumor areas were advanced melanoma, aRCC, and non-small cell lung cancer in these studies.^{17,18,19,20} One attractive characteristic of TNT is its ability to capture the treatment-free interval (TFI) which is defined as the time from the end of index therapy until the date of initiation of a subsequent line of treatment or death whichever occurs first.²¹ TFI has been recently described in ICI-treated advanced melanoma^{21,22} and is considered as an integral component of patient's net health benefit in the American Society of Clinical Oncology value framework.²³ Even though TNT appears to be a relevant clinical endpoint for ICI-treated patients, the systematic literature review highlighted that data on surrogacy analyses based on this alternative endpoint are yet to be studied.¹⁰ The objective herein was thus to estimate TNT and formally assess the surrogate properties of TNT in previously untreated advanced melanoma patients treated with ICI.

Methods

Data

Individual patient-level data (IPD) from CheckMate 067 trial (ClinicalTrials.gov identifier: NCT01844505) were available for this post hoc study. Design, population characteristics and outcomes of this RCT has been described in detail previously.^{2,24} In brief, CheckMate 067 is a phase 3, randomized, double-blind trial of nivolumab monotherapy or nivolumab plus ipilimumab combination versus ipilimumab monotherapy in subjects with previously untreated unresectable or metastatic melanoma. Randomization was stratified according to tumor PD-L1 status (positive vs. negative or indeterminate), BRAF mutation status (V600 mutation–positive vs. wild-type), and

American Joint Committee on Cancer metastasis stage (M0, M1a, or M1b vs. M1c). At database lock, the minimum follow-up from the date on which the last patient underwent randomization was 60 months.

Definition of the efficacy outcomes

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

Statistical analyses

OS and TNT distributions were generated using the Kaplan-Meier estimator. For the description of these efficacy outcomes, hazard ratios (HR) were estimated with the use of a Cox proportional-hazards model stratified according to the factors used in the randomization process. For the SE analysis, HR were estimated with a joint frailty 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 one-step and two-step validation approaches.^{25,26,27,28}

We considered the SE one-step validation statistical method based on a joint frailty model as it has been demonstrated to reduce substantially numerical problems encountered with the two-stage method.²⁵ In this approach, the 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 adjusted on the individual-level random effects, 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 validating of SE supposes the availability of data from multiple RCTs. When the data is limited to only a single or a few RCTs, one can consider the geographic location of centres as the cluster of analysis, which is a common practice in evaluating potential SE.^{30,31,32} We clustered the patients with respect to the country of enrollment resulting in twenty-one clusters (33% of the clusters had 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 region with respect to similarities in their healthcare management (e.g. Finland, Sweden, Norway and Denmark pooled as Nordics region).²⁶ In the case of a high level of surrogacy between TNT and OS at both the individual and the 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 performed 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 reported results following the ReSEEM guidelines for the reporting of surrogate endpoint evaluation.³⁴ All analyses were performed using R software v3.6.1 and the version 3.2.0 of the publicly available R package Frailtypack.³⁵

Results

Full patient characteristics and efficacy results of CheckMate 067 trial were reported in previous articles.^{2,24} Fifteen clusters were generated for this study. Table 1 summarizes efficacy results for outcomes of interest for this study. Median TNT was reached approximately three times sooner than median OS for nivolumab monotherapy and ipilimumab (nivolumab: TNT = 12.1 months, 95% CI [8.9; 18.0], OS=36.9 months 95% CI [28.3; 58.7] ; ipilimumab: TNT= 6.2 months 95% CI [5.4; 7.4], OS= 19.9 months 95% CI [16.9; 24.6]) whereas for nivolumab plus ipilimumab combination therapy

median TNT was 24.2 months (95% CI [16.0; 43.9]), and median OS was not reached yet. Figure 1 presents the Kaplan-Meier curves for TNT. The Kaplan-Meier survival estimates of 5-year TNT in the nivolumab plus ipilimumab, nivolumab monotherapy, and ipilimumab groups were 42.7%, 32.8%, and 10.9%, respectively.

Table 1. Summary of the efficacy results from CheckMate 067. Overall survival outcomes from Larkin et al.²⁴

	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]*	0.63 [0.52; 0.76]	0.52 [0.42; 0.64]	
Time to next treatment			
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]*	0.55 [0.46; 0.65]	0.42 [0.34; 0.50]	

* Cox proportional-hazards model stratified according to tumor PD-L1 status, BRAF mutation status, and AJCC metastasis stage; CI: confidence interval; NR: Not Reached.

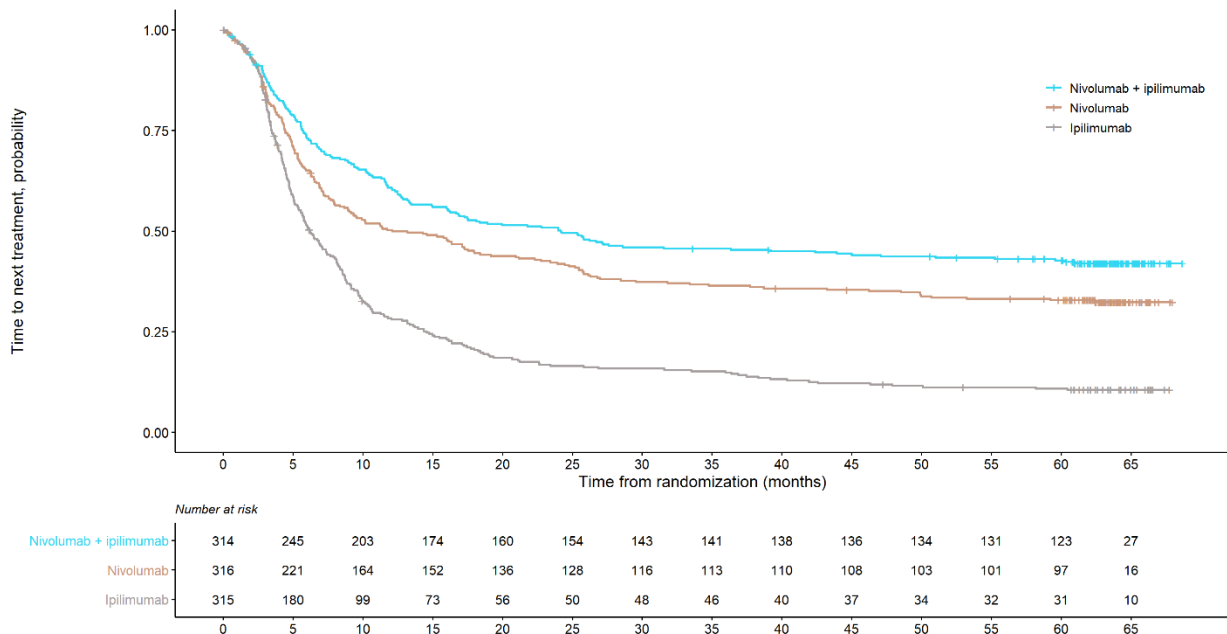


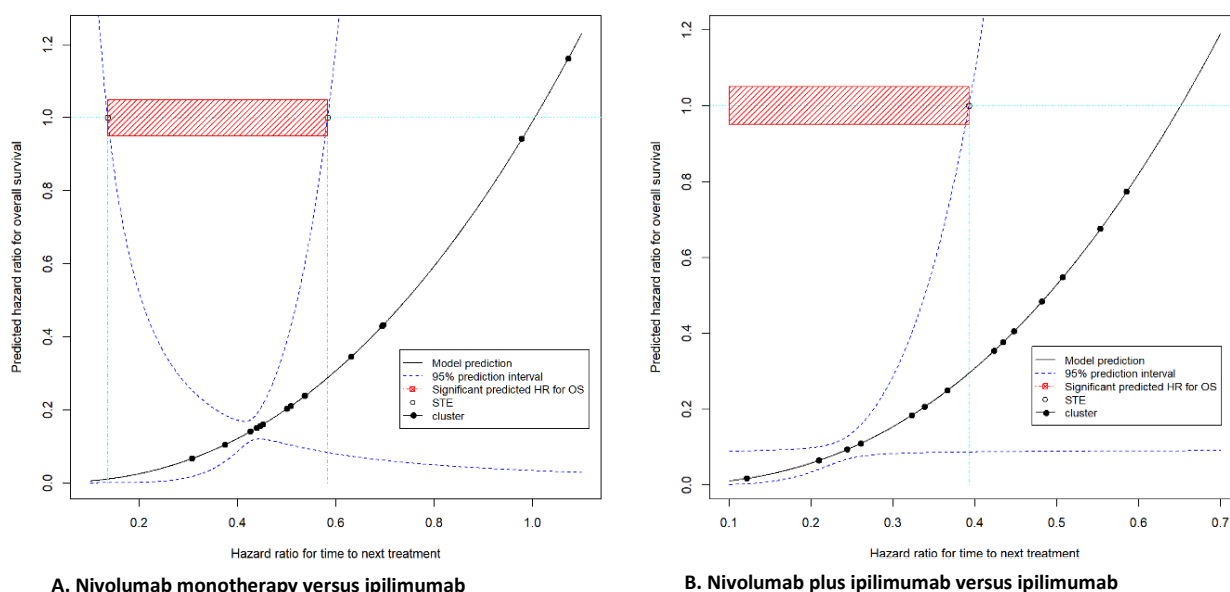
Figure 1. Kaplan-Meier survival curves for TNT for nivolumab plus ipilimumab, nivolumab monotherapy, and ipilimumab monotherapy. TNT: time to next treatment; '+' signs on the Kaplan-Meier curves represent censorings. Estimated parameters following surrogacy assessment are presented in Table 2.

Table 2. Estimated parameters for the assessment of the surrogate properties of TNT for OS in previously untreated advanced melanoma patients treated with ICI.

Criteria	Level of surrogacy	Point estimates [95% CI]
Nivolumab monotherapy vs ipilimumab		
Kendall's τ	Individual	0.61 [0.58; 0.64]
R²_{trial}	Trial	1.00 [0.98; 1.02]
STE (expressed in HR)	Trial	>0.14 and <0.58
Nivolumab + ipilimumab vs ipilimumab		
Kendall's τ	Individual	0.65 [0.62; 0.67]
R²_{trial}	Trial	1.00 [0.99; 1.01]
STE (expressed in HR)	Trial	0.39

All values are rounded to 2 digits; CI: confidence interval; HR: Hazard ratio; STE: Surrogate threshold effect.

Individual association between TNT and OS was deemed acceptable in patients treated with nivolumab monotherapy and in those treated with nivolumab in combination with ipilimumab (Kendall's $\tau > 0.60$). At trial-level, the coefficients of determination were close to 1 with narrow confidence intervals, indicating a strong association between the treatment effects on TNT and OS. Considering these acceptable correlations at both the individual and trial levels, the STE were assessed (Table 2). For nivolumab monotherapy versus ipilimumab, the model predicted a significant HR for OS while HR for TNT was between 0.14 and 0.58. The STE was equal to 0.39 for nivolumab plus ipilimumab. The interval provided for nivolumab monotherapy versus ipilimumab is uncommon and is due to the limited available observations with a HR for TNT below 0.3 in the dataset, resulting in an extremely wide upper prediction limit. Figure 2 presents the 95% prediction limits of the predicted treatment effect on OS for both treatments.



A. Nivolumab monotherapy versus ipilimumab

B. Nivolumab plus ipilimumab versus ipilimumab

Figure 2. Predictions of the treatment effect on OS based on the treatment effect on TNT for nivolumab monotherapy (A, left) and nivolumab plus ipilimumab (B, right) versus ipilimumab.

STE: surrogate threshold effect. For each cluster, the model predicts HR for OS based on the estimated HR for TNT. The horizontal cyan line corresponds to a null hazard ratio (HR) for overall survival (OS), i.e., HR = 1. The vertical line crosses the upper boundary of the 95% prediction limit where the HR for OS is 1. The corresponding value on the x-axis corresponds to the STE. In (A), the model predicted a significant treatment on OS when HR for TNT between 0.14 and 0.58, as no cluster with a HR for TNT <0.3 was observed in the dataset.

Sensitivity analyses

In supplementary Table S1, we reported results from the leave-one-out cross-validation analyses. Overall, except in the scenario excluding cluster 1 for which the model did not converge, data suggest that correlation measures were minimally influenced when the analyses were repeated with slightly smaller and balanced populations, as well as when a different clustering was considered (Table S2). As such, these sensitivity analyses confirm the robustness of the findings on the association between TNT and OS within the dataset available. For the STE, the sensitivity analyses showed some 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 in each cluster was limited (Figure S1). Results suggested a quite high error of prediction of 64% (8/14) with optimistic predicted HR for OS in the comparison of nivolumab monotherapy versus ipilimumab. For nivolumab plus ipilimumab versus ipilimumab, the quality of the prediction was better, even though the error of prediction was 33% (5/15).

Discussion

To our knowledge, this is the first study assessing the surrogacy properties of TNT for OS in ICI-treated patients. We used a strong methodology, since we conducted 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 may be an appropriate SE for OS in double-blind RCTs of advanced melanoma patients treated with ICI. These results were confirmed by our sensitivity analyses supporting the overall stability of the individual- and trial-level correlations. For both treatments, the Kendall' τ for individual-level associations were above 0.6, and the estimated R^2_{trial} were close to 1 with narrow confidence intervals. The goodness of fit of the model was moderate according to the leave-one-out cross-validation results. We observed a relatively high variability in the estimates of some parameters, which yields a large prediction variance for the treatment effects on OS and therefore low values for STE. Fitting a model based on a limited amount of data is known to be challenging and, therefore, these STE estimates could not be considered for predicting a significant OS benefit in future RCTs based solely on our results.

TNT 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 randomized subjects, reaches maturity (i.e. median) earlier than the OS outcome, and may provide greater statistical power at the time of analysis. TNT 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 is a comprehensive assessment of the unique outcomes observed with ICI, as it may take into account the time between interruption of the treatment and the delay to the next systemic treatment, as the patients still benefit from the ICI therapy through the prolonged treatment effect. However, we encourage collecting detailed information after discontinuation of index therapy in RCT such as the reason of not receiving subsequent therapy (e.g.

patients' functional status, patients' refusal to receive a subsequent treatment). Perceived weaknesses of TNT are the potential influences 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; the availability of some therapies may drive treatment decisions.³⁷ Also, further explorations of TNT as an outcome of RCT should be done, since TNT could complement PFS, which may sub optimally characterize the full impact of the novel mechanism of action of ICIs.³⁸ For two time-to-event endpoints, the copula model-based approach introduced by Burzykowski et al. is a commonly used statistical method for validating an SE. However, model convergence issues and large standard error of the coefficient of determination are frequently encountered with this two-step approach.^{25,39} The novel one-step validation joint model was considered in this study since it reduces convergence and numerical problems and it was found to be more robust than the traditional two-step surrogacy approaches for two time-to-event endpoints.²⁵

The limited amount of data is the main limitation of this study. However, IPD from RCTs of other therapies sharing the same mechanism of action and tested in advanced melanoma patients 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 confidence intervals were narrow. These values, although high, are frequently reported.^{25,27} The point estimates and their confidence intervals need however to be interpreted with caution since data from only one RCT was available for this study. The cluster-level definition was therefore the country of enrollment instead of RCT as advised in a meta-analytic approach. The heterogeneity in the treatment effects between clusters in this RCT might be limited compared to that of a traditional meta-analytic approach. This may yield an overestimation of the correlation at the cluster-level, and may have artificially narrowed the reported confidence intervals. Finally, due to the limited sample size, we were not able to perform any subgroup analysis in order to assess whether TNT-OS association is similar within distinct subpopulation, as for example according to the BRAF mutation status.

TNT seems a promising SE of OS in RCTs of ICI-treated patients with advanced melanoma. As a potential SE for OS, TNT may be crucial to present in the frame of regulatory and health care decision making of ICIs, especially when OS data are immature. We encourage sponsors of RCTs to carefully collect the date of subsequent systemic treatment, as it has been reported to be heterogeneously captured.⁴⁰ Surrogacy analyses could consequently be performed with a larger number of RCTs in order to confirm these findings, and for improving the quality of the predictions. 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 (VEGF) inhibitors or chemotherapy would definitely be relevant for evaluating whether TNT is also a candidate SE for OS in these other settings.

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.

Acknowledgment

We thank Pr Brigitte Dreno, head of dermatology department, CHU of Nantes, France, for her review and her comments of the manuscript.

Table S1. Sensitivity analyses in the Kendall's τ and R^2_{trial} values by removing one cluster iteratively. All values are rounded to 2 digits; CI: confidence interval; HR: hazard ratio; STE: surrogate threshold effect

Cluster excluded	Nivolumab monotherapy versus ipilimumab				Nivolumab + ipilimumab versus ipilimumab			
	N of patients in the cluster	Kendall's τ Point estimates [95% CI]	R^2_{trial} Point estimates [95% CI]	STE (expressed in HR)	N of patients in the cluster	Kendall's τ Point estimates [95% CI]	R^2_{trial} Point estimates [95% CI]	STE (expressed in HR)
#1	75	<i>Model did not converge</i>			77	0.62 [0.59; 0.64]	1.00 [0.99; 1.01]	0.41
#2	19	0.61 [0.58; 0.64]	1.00 [0.95; 1.05]	>0.19 and <0.57	22	0.65 [0.62; 0.67]	1.00 [0.99; 1.01]	0.40
#3	27	0.60 [0.57; 0.63]	1.00 [1.00; 1.00]	>0.09 and <0.59	26	0.62 [0.59; 0.65]	1.000 [1.00; 1.00]	0.43
#4	12	0.61 [0.58; 0.63]	1.00 [0.98; 1.02]	>0.14 and <0.59	16	0.64 [0.61; 0.66]	1.00 [0.99; 1.01]	0.40
#5	14	0.61 [0.58; 0.64]	1.00 [0.99; 1.01]	>0.11 and <0.60	14	0.65 [0.62; 0.67]	1.00 [0.99; 1.01]	0.40
#6	41	0.62 [0.59; 0.64]	1.00 [1.00; 1.00]	>0.13 and <0.56	38	0.63 [0.59; 0.65]	1.00 [0.85; 1.14]	>0.17 and <0.26
#7	25	0.61 [0.58; 0.64]	1.00 [0.99; 1.01]	>0.10 and <0.61	32	0.62 [0.59; 0.64]	1.00 [0.99; 1.02]	0.43
#8	21	0.61 [0.58; 0.63]	1.00 [0.96; 1.03]	>0.20 and <0.53	25	0.64 [0.61; 0.66]	1.00 [0.99; 1.01]	0.38
#9	47	0.62 [0.58; 0.64]	1.00 [0.95; 1.05]	>0.16 and <0.59	40	0.62 [0.59; 0.65]	1.00 [0.99; 1.01]	0.43
#10	63	0.61 [0.57; 0.64]	1.00 [1.00; 1.00]	>0.23 and <0.58	66	0.65 [0.61; 0.67]	1.00 [0.98; 1.02]	0.37
#11	20	0.62 [0.58; 0.64]	1.00 [0.99; 1.01]	>0.11 and <0.60	18	0.65 [0.62; 0.67]	1.00 [0.99; 1.01]	0.41
#12	70	0.63 [0.59; 0.66]	1.00 [1.00; 1.00]	>0.02 and <0.63	80	0.64 [0.61; 0.66]	1.00 [0.98; 1.02]	0.40
#13	23	0.61 [0.58; 0.63]	1.00 [0.98; 1.02]	>0.23 and <0.58	17	0.65 [0.62; 0.67]	1.00 [0.98; 1.02]	0.39
#14	21	0.62 [0.58; 0.64]	1.00 [0.97; 1.03]	>0.09 and <0.61	19	0.65 [0.62; 0.67]	1.00 [0.99; 1.01]	0.40
#15	143	0.62 [0.59; 0.65]	1.00 [0.99; 1.01]	0.58	139	0.63 [0.59; 0.65]	1.00 [0.98; 1.02]	0.40

Table S2. Sensitivity analysis considering 9 clusters instead of 15. All values are rounded to two digits; HR: hazard ratio; STE: surrogate threshold effect

Criteria	Level of surrogacy	Point estimates [95%CI]
Nivolumab monotherapy vs ipilimumab		
Kendall's τ	Individual	0.62 [0.58; 0.64]
R^2_{trial}	Trial	1.00 [0.99; 1.01]
STE (expressed in HR)	Trial	>0.04 and <0.61
Nivolumab plus ipilimumab vs ipilimumab		
Kendall's τ	Individual	0.62 [0.59; 0.64]
R^2_{trial}	Trial	1.00 [0.99; 1.01]
STE (expressed in HR)	Trial	0.39

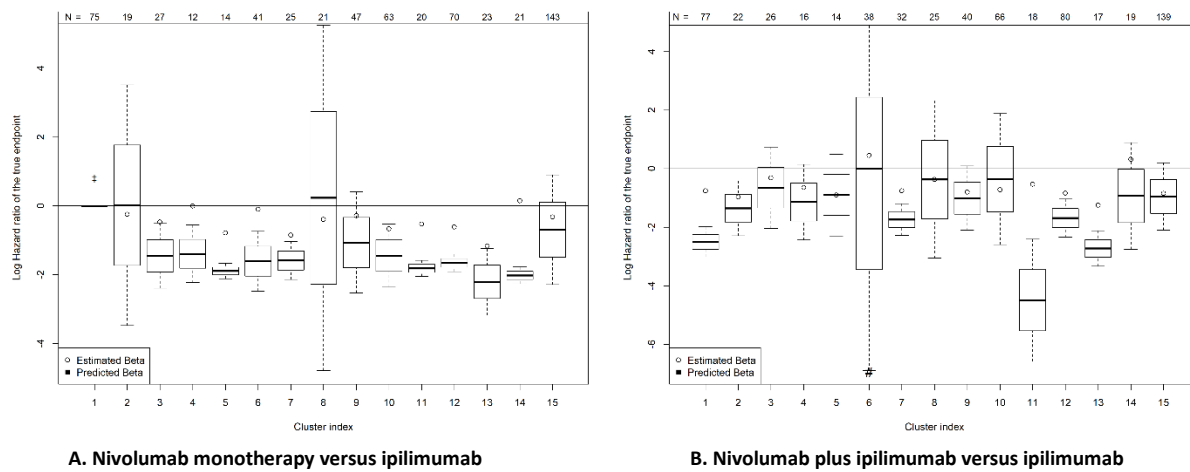


Figure S1. Prediction of the log (HR) on OS based the log (HR) of TNT vs the observed log (HR) in each cluster by leaving one cluster out at a time for nivolumab monotherapy (A, left) and nivolumab plus ipilimumab (B, right), versus ipilimumab. For each cluster, the one-step joint surrogate model was estimated by excluding the cluster of interest. Dotted bars are the 95% prediction intervals; “Estimated Beta”: log (HR) estimated by a simple Cox model; “Predicted Beta”: log (HR) estimated by a joint frailty model; #: prediction bounds restricted for illustration purpose; ‡: model did not converge;

References

- 1 Galluzzi L, Vacchelli E, Bravo-San Pedro J-M et al. Classification of current anticancer immunotherapies. *Oncotarget* 2014; **5**(24):12472–508.
- 2 Larkin J, Chiarion-Sileni V, Gonzalez R et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 2015; **373**:23-34.
- 3 Motzer RJ, Tannir NM, McDermott DF et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med* 2018; **378**:1277-90.
- 4 Fleming TR, Powers JH. Biomarkers and Surrogate Endpoints in Clinical Trials. *Stat Med* 2012; **31**(25): 2973–84.
- 5 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):3721–26.
- 6 Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist* 2008; **13**(S2):19–21.
- 7 U.S. Department of Health and Human Services, Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry; 2018. Available from: <https://www.fda.gov/media/71195/download>.
- 8 European Medicines Agency. Guideline on the evaluation of anticancer medicinal products in man, 2017. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man-revision-5_en.pdf
- 9 Burzykowski T, Molenberghs G, Buyse, M. The Evaluation of Surrogate Endpoints. New-York: Springer-Verlag, 2005.
- 10 Branchoux S, Bellera C, Italiano A et al. 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.
- 11 Mushti SL, Mulkey F and 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):2268–75.
- 12 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 Network Open* 2018; **1**(2):e180416.
- 13 Kaufman H, Schwartz L, William Jr W et al. Evaluation of classical 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):2245-61.
- 14 Buyse M, Saad ED, Burzykowski T. Assessing Treatment Benefit in Immuno-oncology. *Stat Biosci* 2020; **12**:83–103.
- 15 Mesnard C, Bodet-Milin C, Eugène T et al. Predictive value of FDG-PET imaging for relapse in metastatic melanoma patients treated with immunotherapy. *J Eur Acad Dermatol Venereol* 2020; doi: 10.1111/jdv.16358 [Epub ahead of print].
- 16 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 Medicine* 2017; **15**(1):78.
- 17 Luke JJ, Ghatge 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; **37**(8S): 141.

- 18 Dudani S, Graham J, Wells JC. 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):861-67.
- 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**(8S): viii303-viii331.
- 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): 288-93.
- 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):3350-58.
- 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): 2925-34.
- 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**:1535-46.
- 25 Sofeu 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):2928-42.
- 26 Burzykowski T, Molenberghs G, Buyse M et al. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *Appl Statist* 2001; **50**: 405-22.
- 27 Rotolo F, Paoletti X, Burzykowski T et al. A Poisson approach to the validation of failure time surrogate endpoints in individual patient data meta-analyses. *Stat Methods Med Res* 2019; **28**(1):170-83.
- 28 Renfro L, Shi Q, Sargent D, Carlin B. Bayesian adjusted R^2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials. *Stat Med* 2012; **31**(8):743-61.
- 29 Sofeu CL, Rondeau V. How to use frailtypack for validating failure-time surrogate endpoints using individual patient data from meta-analyses of randomized controlled trials. *Plos One* 2020; **15**(1):e0228098.
- 30 Alonso A, Bigirimurame T, Burzykowski T et al. Applied surrogate endpoint evaluation methods with SAS and R. Chapman & Hall/ CRC biostatistics series, 2016.
- 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):6139-48.
- 32 Coart E, Suci 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**:171-74.
- 33 Burzykowski T, Buyse M. Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceut Statist* 2006; **5**: 173–86.
- 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 Spectrum* 2019; **3**(1)pkz002.
- 35 Krol A, Mauguen A, Mazroui Y et al. Tutorial in Joint Modeling and Prediction: a Statistical Software for Correlated Longitudinal Outcomes, Recurrent Events and a Terminal Event. *J Stat Soft* 2017; **81**(3):1-52.
- 36 Marshall J, Schwartzberg LS, 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**(15S): 11521.

- 37 Campbell BA et al. Time to next treatment as a meaningful endpoint for trials of primary cutaneous lymphoma. *Cancers* 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**(5S):v393.
- 39 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.
- 40 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):830-38.

4 Time to next treatment as a candidate surrogate endpoint for overall survival in advanced renal cell carcinoma patients treated with immune-checkpoint inhibitors

4.1 Introduction

Based on the preliminary evidence that TNT may be an appropriate SE for OS in RCTs of ICI-treated patients with advanced melanoma, we were interested in conducting similar surrogacy analyses in ICI-treated patients with another cancer type. Similarly to melanoma, in advanced renal cell carcinoma (aRCC):

- It may take a significant period of time to get mature survival data;
- A significant benefit in terms of TFI for ICI-treated patients was observed;¹⁹

It was therefore particularly relevant to assess whether TNT may be an appropriate SE for OS in this tumor type. Access to IPD from CheckMate 214 trial was granted for this *post-hoc* analysis. Full description of CheckMate 214 has been published by Motzer et al.^{4,60} and is briefly presented in section 3.

In August 2020, a new statistical approach for validating a SE was accepted for publication, namely the one-step joint frailty-copula model.¹³ This approach, fully described by Sofeu et al.¹³ is summarized in section 4.2. In this work, we applied both approaches.

We used patient-level data from CheckMate 214 RCT with 42 months' minimum follow-up. In this RCT, patients were randomly assigned to nivolumab combined with ipilimumab (NIVO+IPI) versus sunitinib. The one-step joint surrogate model and the one-step joint frailty-copula model (considering Clayton or Gumbel copula) were used where the country of enrollment was applied to define synthetic clusters (N=22). Kendall's τ and the coefficient of determination (R^2_{trial}) were estimated for respective measurements of association at the individual and cluster levels. Surrogate threshold effect (STE), the maximum threshold hazard ratio (HR) for TNT that would translate into OS benefit, was estimated. A leave-one-out cross-validation analysis was performed to evaluate model robustness and predictive accuracy.

Twenty-two clusters were considered for 847 patients. Results from both models showed an acceptable association between TNT and OS at the individual level (Kendall's τ ranged from 0.59

to 0.70) and strong at the cluster level (R^2_{trial} close to 1). The estimated STE was 0.70 for NIVO+IPI versus sunitinib based on the one-step joint surrogate model; the joint frailty Clayton copula model predicted a significant treatment effect on OS only when the HR for TNT was between 0.65 and 0.74 and the joint frailty Gumbel copula model did not provide STE. Cross-validation results showed association measures were minimally influenced when the analyses were repeated removing one cluster iteratively. The predictive accuracy of the joint frailty Clayton copula model was improved compared to the other approaches, although the confidence intervals were not narrow.

TNT may be a valuable SE in previously untreated aRCC patients treated with ICI. Further surrogacy analyses considering multiple RCTs of ICI-treated aRCC patients are warranted for confirming these findings. This work is currently under review by the co-authors.

4.2 Description of the one-step joint frailty-copula approach

The one-step joint frailty copula approach is an alternative approach for the validation of surrogate endpoints, based on an extension of the joint-frailty copula model of Emura et al.⁶¹ This novel statistical model includes two correlated random effects treatment-by-trial interaction and a shared random effect at trial level accounting for heterogeneity on baseline risks. At the individual level, the joint survivor function of failure time endpoints are linked using copula functions. Therefore, instead of assuming independence between the survival endpoints conditional on a shared frailty term as in the one-step joint surrogate model, a dependence using copula models is considered.¹³ Two different copula functions are available in this model: the Clayton and the Gumbel-Hougaard copulas. Both copulas induce positive association. In the Clayton copula, the strength of the association between the SE and the final endpoint decreases with decreasing value of the copula parameter, θ ($\theta > 1$), and reaches independence when $\theta = 1$. In the Gumbel-Hougaard copula, the strength of the association between the SE and the final endpoint decreases with increasing value of θ ($\theta \in [0; 1]$), and reaches independence when $\theta = 1$. Hence, the two copulas capture different types of dependence structure between S_{ij} and T_{ij} .⁶²

The one-step joint frailty-copula surrogate model is defined according to Sofeu et al.¹³ as:

$$\begin{aligned} & \bar{F}(s_{ij}, t_{ij} | Z_{S,ij}, Z_{T,ij}, v_i) \\ & = P(S_{ij} \geq s_{ij}, T_{ij} \geq t_{ij} | Z_{S,ij}, Z_{T,ij}, v_i) = C_{\theta}(\bar{F}(s_{ij} | Z_{S,ij}, u_i, v_{Si}), \bar{F}(t_{ij} | Z_{T,ij}, u_i, v_{Ti})) \end{aligned}$$

With the conditional survival functions associated with the endpoints as follows:

$$\bar{F}_{S_{ij}}(t_{S,ij}|Z_{S_{ij}}, u_i, v_{S_i}) = \exp \left\{ - \int_0^{t_{S,ij}} \lambda_{0S}(x) \exp \left(u_i + v_{S_i} Z_{ij1} + \beta_S Z_{S_{ij}} \right) dx \right\}$$

$$\bar{F}_{T_{ij}}(t_{T,ij}|Z_{T_{ij}}, u_i, v_{T_i}) = \exp \left\{ - \int_0^{t_{T,ij}} \lambda_{0T}(x) \exp \left(\alpha u_i + v_{T_i} Z_{ij1} + \beta_T Z_{T_{ij}} \right) dx \right\}.$$

C_θ is the copula function. $\lambda_0(x)$ and β are respectively the baseline hazard function and the fixed effects corresponding to the covariates Z associated with the failure time endpoint. u_i is a shared random effect associated with the baseline hazard function that will serve to take into account the heterogeneity between trials of the baseline hazard function, associated with the fact that there are several trials. The power parameter α distinguishes trial-level heterogeneities between the surrogate and the true endpoints. v_{S_i} and v_{T_i} are two correlated random effects treatment-by-trial interactions (trial-level frailties in interaction with the treatment), which aim to assess the prediction of the treatment effect on the true endpoint.

The individual level surrogacy is assessed using Kendall's τ , based on the copula parameter, as in the classical two-step copula approach.⁹ The trial-level association is estimated using the coefficient of determination (R^2_{trial}), based on the covariance matrix of two correlated trial-level random effects in interaction with the treatment, as in the one-step joint surrogate model.

4.3 Description of CheckMate 214 trial

Study design

CheckMate 214 is a randomized, open-label, phase III trial of nivolumab combined with ipilimumab followed by nivolumab monotherapy versus sunitinib monotherapy, conducted in previously untreated aRCC patients. The primary efficacy population was composed of intermediate- to poor-risk patients as per international metastatic renal cell carcinoma database consortium (IMDC). Randomization (in a 1:1 ratio) was performed with a block size of 4 with stratification according to IMDC risk score (0 vs. 1 or 2 vs 3 to 6) and geographic region (United States vs. Canada and Europe vs. the rest of the world). Treatment continued until disease progression (as per RECIST, version 1.1), development of unacceptable toxic events, or withdrawal of consent. Patients could be treated after progression, provided that they had a clinical benefit and did not have substantial adverse effects, as assessed by the investigator. After the trial was stopped when nivolumab combined with ipilimumab showed significant OS benefit in the primary efficacy population, an amendment to the study protocol permitted to the patients to discontinue after two years of study treatment even in the absence of disease

progression or unacceptable toxicity, and also to cross-over to nivolumab combined with ipilimumab arm from sunitinib arm. Enrollment period was from October 2014 through February 2016. At database lock, the minimum follow-up from the date on which the last patient underwent randomization was 42 months.

TNT and OS outcomes

Full patient characteristics and efficacy results were reported in publications from Motzer et al.^{4,60} Out of the 847 patients included in the study, 76.5% had initiated a subsequent systemic anticancer therapy or died at the end of the follow-up. Median TNT was reached approximately 3 times earlier than median OS in the nivolumab combined with ipilimumab arm (median TNT= 15.2 months, 95% CI [11.5; 19.4], median OS= 47.0 months 95% CI [35.6; Not estimable]), as well as in the sunitinib arm (median TNT = 8.5 months, 95% CI [7.3; 10.2], median OS=26.6 months 95% CI [22.1; 35.5]). Figure 9 presents the Kaplan-Meier curves for OS and TNT. The Kaplan-Meier survival estimates of 42-month TNT in the nivolumab combined with ipilimumab, and sunitinib groups were 31.6%, and 12.6%, respectively.

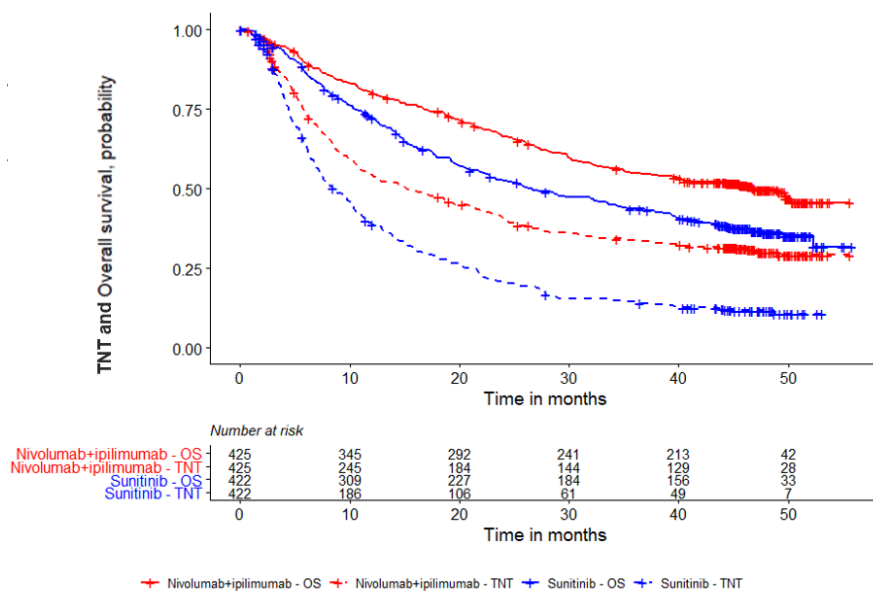


Figure 9. Kaplan-Meier curves for OS and TNT for nivolumab combined with ipilimumab and for sunitinib in previously untreated advanced intermediate- to poor-risk aRCC patients from CheckMate-214 trial. TNT: time to next treatment; '+' signs on the Kaplan-Meier curves represent censorings.

4.4 Publication

The manuscript is presented in the next pages.

1 **INVESTIGATING TIME TO NEXT TREATMENT AS A SURROGATE ENDPOINT FOR OVERALL**
2 **SURVIVAL IN PREVIOUSLY UNTREATED INTERMEDIATE- TO POOR-RISK ADVANCED RENAL CELL**
3 **CARCINOMA PATIENTS: AN INSIGHT FROM THE PHASE III CHECKMATE-214 TRIAL**

4 Sébastien Branchoux^{1,2,3}, Casimir Ledoux Sofeu², Anne-Françoise Gaudin¹, Murat Kurt⁴, Jessica May⁵,
5 Antoine Italiano⁶, Virginie Rondeau^{2,3}, Carine Bellera^{3,7,8}
6
7

8 1. Department of Health Economics & Outcomes Research, Bristol Myers Squibb, Rueil-Malmaison, France

9 2. Biostatistic Team, Bordeaux Population Health Center, ISPED, Centre INSERM U1219, INSERM, Bordeaux, France

10 3. Epicene Team (Cancer & Environment), Bordeaux Population Health Center, ISPED, Centre INSERM U1219, INSERM, Bordeaux, France

11
12 4. Bristol Myers Squibb, Lawrenceville, NJ 08648

13 5. Bristol Myers Squibb, , Uxbridge, Middlesex, United-Kingdom

14 6. Department of Early phase trial unit, Institut Bergonié Comprehensive Cancer Centre, Bordeaux, France

15 7. Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Center, Bordeaux, France

16 8. Department of Clinical Epidemiology, INSERM CIC-EC 14.01, Bordeaux, France
17
18
19
20

21 **Keyword:** immune-checkpoint inhibitors, nivolumab, ipilimumab, surrogate endpoint, overall survival,
22 time to next treatment, advanced renal cell carcinoma
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 **Abstract**

50
51 **PURPOSE:** Time to next treatment (TNT) may be a patient-relevant endpoint by capturing durable
52 response and the time free of systemic anticancer therapy, which may be achieved with immune-checkpoint
53 inhibitors (ICI). Preliminary evidence of the potential validity of TNT as an SE for OS in ICI-treated
54 patients with advanced melanoma was recently reported. This study investigated TNT as a surrogate
55 endpoint (SE) of overall survival (OS) in previously untreated advanced renal cell carcinoma (aRCC)
56 patients.

57 **METHODS:** Patient-level data from CheckMate-214 randomized clinical trial (RCT) (NCT02231749)
58 with 42 months' minimum follow-up were used. In this RCT, patients were randomly assigned to
59 nivolumab combined with ipilimumab (NIVO+IPI) versus sunitinib. The one-step joint surrogate model
60 and the one-step joint frailty-copula model were used where the country of enrollment was applied to define
61 synthetic clusters. Kendall's τ and the coefficient of determination (R^2_{trial}) were estimated for respective
62 measurements of association at the individual and cluster levels. Surrogate threshold effect (STE), the
63 maximum threshold hazard ratio (HR) for TNT that would translate into OS benefit, was estimated. A
64 leave-one-out cross-validation analysis was performed to evaluate model robustness and predictive
65 accuracy.

66 **RESULTS:** Twenty-two clusters were considered from 847 patients. Results from both models showed an
67 acceptable association between TNT and OS at the individual level (Kendall's τ ranged from 0.59 to 0.70)
68 and strong at the cluster level (R^2_{trial} close to 1). The estimated STE was 0.70 for NIVO+IPI versus sunitinib
69 based on the one-step joint surrogate model; the joint frailty Clayton copula model predicted a significant
70 treatment effect on OS only when the HR for TNT was between 0.65 and 0.74. Cross-validation results
71 showed association measures were minimally influenced when the analyses were repeated removing one
72 cluster iteratively. The predictive accuracy of the joint frailty copula model was improved compared to the
73 other approach, although the prediction intervals were not narrow.

74 **CONCLUSIONS:** TNT may be a valuable SE in previously untreated aRCC patients treated with ICI.
75 Further surrogacy analyses considering multiple RCTs of ICI-treated aRCC patients are warranted for
76 confirming these findings.

77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96

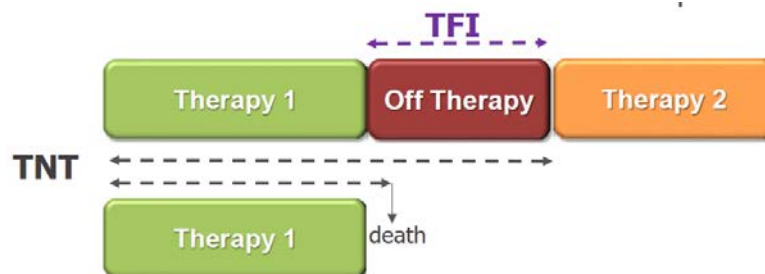
97 **Background**

98 Immune checkpoint inhibitors (ICI) are emerging as the new corner stone of cancer treatment due to their
99 ability to produce durable responses in patients with various cancers.¹ Seven ICI have now been approved
100 by health regulatory agencies to treat a variety of cancers, and more than 1,000 clinical trials involving
101 these agents are under way.² In oncology, overall survival (OS), defined as the time from randomization to
102 the date of death (any cause), is the most objective and compelling primary endpoint of randomized clinical
103 trial (RCT), especially in the advanced setting of the disease.^{3,4} This endpoint may however require
104 extensive follow-up and a substantial number of patients to demonstrate a clinically and statistically
105 significant relevant difference when comparing therapeutic strategies. OS might also present the
106 disadvantage to be confounded by the effects from subsequent therapies especially in cancers with better
107 prognosis where treatment landscape is rapidly evolving in all stages.⁵ In this context, surrogate endpoints
108 (SE), defined as ‘biomarkers or intermediate endpoints intended to substitute for a clinical endpoint’,⁶ are
109 therefore particularly critical to provide an early access to innovative treatment.

110 We recently reported preliminary evidence of the potential validity of time to next treatment (TNT) as an
111 SE for OS in ICI-treated patients with advanced melanoma.⁵ TNT is defined as the time between baseline
112 (i.e. randomization, inclusion or treatment initiation) and the date of next subsequent systemic treatment
113 initiation or the date of death, whichever occurs first. Besides its potential validity as an SE for OS, TNT
114 presents also the attractive characteristic to capture durable response and the time free of systemic
115 anticancer therapy between the ICI protocol therapy cessation until the date of initiation of a subsequent
116 treatment (treatment-free interval, TFI), which may be achieved with ICI. Figure 1 illustrates the concepts
117 of TNT and TFI. The validity of an SE depending on the tumor type⁷ and such TFI having been observed
118 in previously untreated advanced renal cell carcinoma (aRCC) patients treated with nivolumab combined

119 with ipilimumab,⁸ there is an interest to assess the surrogacy properties of TNT in ICI-treated patients with
120 aRCC.

121 Several statistical methods have been developed to validate an SE, the most robust ones being those
122 developed in the meta-analytic framework.⁷ For two time-to-event endpoints, the copula model-based
123 approach introduced by Burzykowski et al. is one of the commonly used statistical method.⁹ However,
124 model convergence issues and the non-guarantee of the positivity of the variance of some model parameters
125 can be encountered with this two-step approach.^{10,11} Based on these observations, Sofeu et al. recently
126 developed two alternative approaches, namely the one-step joint surrogate model,¹⁰ and the one-step joint
127 frailty-copula model.¹² Through simulation studies, the authors reported that both methods reduce
128 convergence and numerical problems when compared to the traditional two-step surrogacy approach.^{10,12}
129 The objective of this study was thus to formally assess the surrogate properties of TNT in ICI-treated aRCC
130 patients, through these two recently validated statistical approaches.



131
132 **Figure 1. Illustration of the definition of time to next treatment (TNT) and treatment-free interval (TFI).**

133

134 **Methods**

135 *Data*

136 Individual patient-level data (IPD) from CheckMate 214 RCT (ClinicalTrials.gov identifier:
137 NCT02231749) were used for this post hoc analysis. Design, population characteristics and outcomes of
138 this RCT have been described in detail previously.^{13,14} In brief, CheckMate 214 is a randomized, open-

139 label, phase 3 trial of nivolumab plus ipilimumab followed by nivolumab monotherapy versus sunitinib
140 monotherapy, conducted in previously untreated aRCC patients. The primary efficacy population was
141 composed of intermediate- to poor-risk patients as per international metastatic renal cell carcinoma database
142 consortium (IMDC); this population was thus considered for this study. Randomization (in a 1:1 ratio) was
143 performed with stratification according to IMDC risk score and geographic region (United States vs.
144 Canada and Europe vs. the rest of the world). Treatment continued until disease progression (as per
145 RECIST, version 1.1), development of unacceptable toxic events, or withdrawal of consent. Patients could
146 be treated after progression, provided that they had a clinical benefit and did not have substantial adverse
147 effects, as assessed by the investigator. After the trial was stopped when nivolumab combined with
148 ipilimumab showed significant OS benefit in the primary efficacy population, an amendment to the study
149 protocol permitted to the patients to discontinue after two years of study treatment even in the absence of
150 disease progression or unacceptable toxicity, and to also cross-over to nivolumab combined with
151 ipilimumab arm from sunitinib arm. The minimum follow-up from the date on which the last patient
152 underwent randomization was 42 months.

153

154 *Definition of the endpoints*

155 OS was defined as the time from the date of randomization to the date of death (any cause). TNT was
156 defined as the time between the date of randomization and the date of subsequent systemic treatment
157 initiation, or the date of death (any cause), whichever occurs first. Both outcomes were censored on the last
158 date a subject was known to be alive.

159

160

161 *Statistical analyses*

162 OS and TNT distributions were generated using the Kaplan-Meier estimator. Hazard ratios (HR) were
163 estimated with the use of a Cox proportional-hazards model stratified on the factors used in the
164 randomization process. For the SE analyses, HR were estimated with either the one-step joint surrogate
165 model¹⁰ or the one-step joint frailty-copula models.¹²

166 Both surrogacy methods considered in this study are meta-analytical approaches. The patient-level
167 surrogacy measures the individual association between the candidate SE and the clinical endpoint of
168 interest. The trial-level surrogacy assesses the association between the treatment effect on the candidate SE
169 and the treatment effect on the final endpoint. These two associations were estimated using both
170 aforementioned joint frailty models. These models were detailed in earlier publications,^{10,12} and are briefly
171 presented hereafter.

172 In the one-step joint surrogate model, individual-level random effects shared by the candidate SE and the
173 final endpoint account for the heterogeneity of the survival times at the individual levels.¹⁰ The Kendall's
174 τ coefficient of correlation, adjusted on the individual-level random effects, is used to estimate the strength
175 of the association between the candidate SE and the final endpoint at the individual level. The trial-level
176 association is estimated using the coefficient of determination (R^2_{trial}), based on the covariance matrix of
177 two correlated trial-level random effects in interaction with the treatment. The 95% confidence intervals
178 (CI) are estimated using the parametric bootstrap method for τ and with the delta method for R^2_{trial} , which
179 can result in confidence limits violating the [0; 1] interval.¹⁰

180 The one-step joint frailty-copula model does not account for a shared individual-level random effect for the
181 dependence between the endpoints at the individual level, as the previous model does. Instead, the joint
182 survivor functions for the candidate SE and the final endpoint are linked using a copula function, either the

183 Clayton or the Gumbel-Hougaard copula.¹² Both copulas induce positive association. In the Clayton copula,
184 the strength of the association between the SE and the final endpoint decreases with decreasing value of the
185 copula parameter, θ ($\theta > 1$), and reaches independence when $\theta = 1$. In the Gumbel-Hougaard copula, the
186 strength of the association between the SE and the final endpoint decreases with increasing value of θ ($\theta \in$
187 $[0; 1]$), and reaches independence when $\theta = 1$. In this model, dependence is measured by the copula
188 parameter, while in the one-step joint surrogate model, independence between the endpoints conditional on
189 the shared random effects is assumed. The Kendall's τ coefficient of correlation, estimated here as a
190 function of the copula parameter, is used to assess the association at the individual-level. The estimation of
191 R^2_{trial} in this model follows the same approach as in the one-step joint surrogate model. The 95% CI for
192 both τ and R^2_{trial} are estimated using the delta method.¹²

193 The meta-analytical approach for validating of an SE assumes the availability of multiple RCTs. When the
194 data is limited to single or a few RCTs, one can consider the geographic location of the centres as the cluster
195 of analysis, which is a common practice in evaluating SE.^{5,15,16,17} When the number of patients was lower
196 than 4 per treatment arm per country, we pooled the countries in region with respect to similarities in their
197 healthcare management (e.g. Finland, Sweden, Norway and Denmark pooled as Nordics region).⁹ For
198 considering an SE as valid, Kendall's τ coefficient of correlation should be above an informal threshold of
199 0.6, and the R^2_{trial} sufficiently close to 1.^{7,18} In such case, we estimated the surrogate threshold effect (STE),
200 defined as the minimum treatment effect on the SE necessary to predict a significant treatment effect on the
201 final endpoint.¹⁹ A candidate SE is considered strongly valid if the STE can be reached in an RCT. As a
202 sensitivity analysis, we performed a leave-one-out cross-validation (LOOCV) analysis to assess the
203 robustness of the models regarding Kendall's τ , R^2_{trial} , the STE, and the prediction performance. The model
204 selection was based on both the prediction performance of the models and the goodness of fit assessed

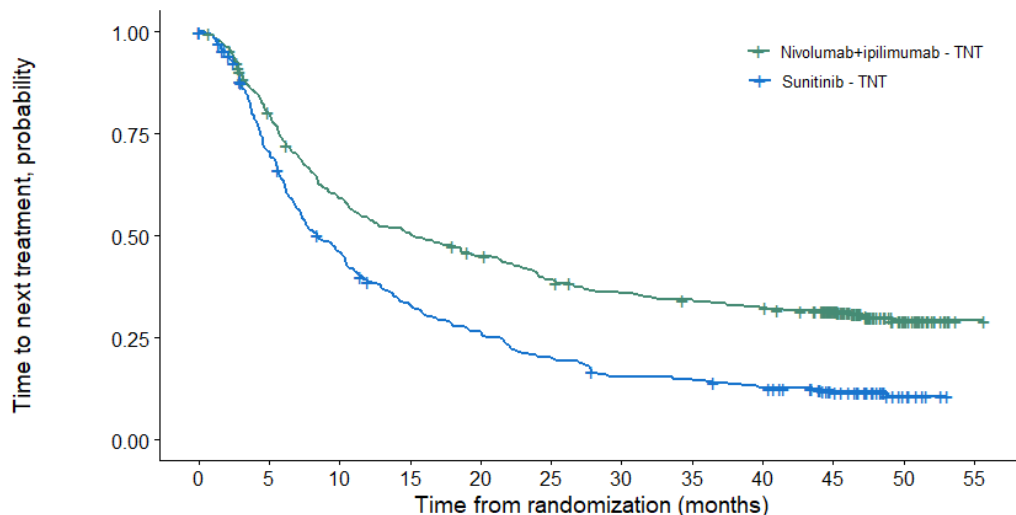
205 through the approximate likelihood cross-validation criterion (LCV α), which is equivalent to the Akaike
 206 information criterion in the penalized likelihood framework.^{12,20}
 207 We reported results following the ReSEEM guidelines for the reporting of surrogate endpoint evaluation.²¹
 208 All analyses were performed using R software v3.6.1 and the version 3.2.0 of the freely available R package
 209 Frailtypack.²²

210
 211 **Results**
 212 Full patient characteristics and efficacy results were reported previously.^{13,14} Twenty-two clusters were
 213 generated for this study. Table 1 summarizes results for TNT and OS outcomes. Out of 847 patients, 76.5%
 214 had initiated a subsequent systemic anticancer therapy or died by the end of the follow-up. Median TNT
 215 was reached approximately 3 times sooner than median OS in the nivolumab combined with ipilimumab
 216 arm (median TNT= 15.2 months, 95% CI [11.5; 19.4], median OS= 47.0 months 95% CI [35.6; Not
 217 estimable]), as well as in the sunitinib arm (median TNT = 8.5 months, 95% CI [7.3; 10.2], median OS=26.6
 218 months 95% CI [22.1; 35.5]). Figure 2 presents the Kaplan-Meier curves for TNT. The Kaplan-Meier
 219 estimates of the 3-year TNT in the nivolumab combined with ipilimumab, and sunitinib groups were 33.8%,
 220 and 14.4%, respectively.

221 **Table 1. Summary of the efficacy results of the primary efficacy population from CheckMate-214.** Overall
 222 survival outcomes from Motzer et al.¹⁴

	Nivolumab + ipilimumab (N=425)	Sunitinib (N=422)
Overall survival		
Median (months) [95% CI]	47.0 [35.6; NE]	26.6 [22.1; 33.5]
Events (n, (%))	207 (48.7)	255 (60.4)
Hazard ratio [95% CI]*	0.66 [0.55; 0.80]	
Time to next treatment		
Median (months) [95% CI]	15.2 [11.5;19.4]	8.5 [7.3; 10.2]
Events (n, (%))	288 (67.8)	360 (85.3)
Hazard ratio [95% CI]*	0.58 [0.49; 0.67]	

223 *Cox proportional-hazards model stratified according to IMDC risk score and geographic region; CI: confidence interval; NE:
 224 Not estimable.



	Number at risk											
	0	5	10	15	20	25	30	35	40	45	50	55
Nivolumab+ipilimumab - TNT	425	333	245	209	184	159	144	136	129	103	28	1
Sunitinib - TNT	422	289	186	132	106	80	61	58	49	34	7	0

225
 226 **Figure 2. Kaplan-Meier curves for TNT for nivolumab combined with ipilimumab and for sunitinib in**
 227 **previously untreated advanced intermediate- to poor-risk aRCC patients from CheckMate-214 trial. TNT: time**
 228 **to next treatment; ‘+’ signs on the Kaplan-Meier curves represent censorings.**

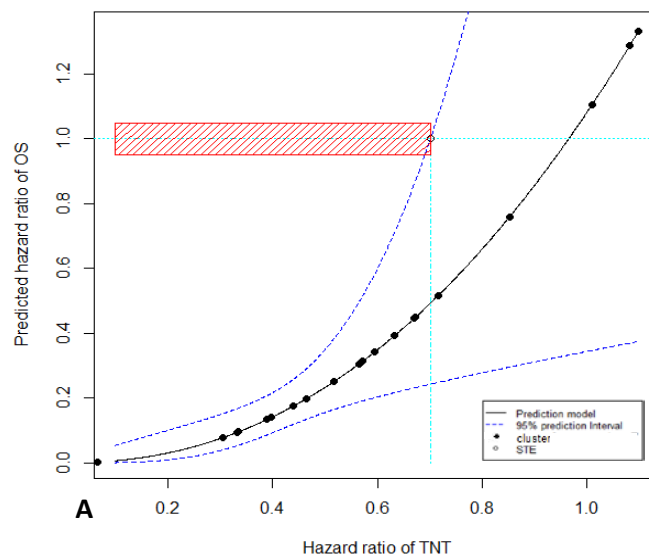
229
 230 Estimated parameters following surrogacy assessment according to the different statistical models are
 231 presented in Table 2.

232
 233 **Table 2. Estimated parameters for the assessment of the surrogate properties of TNT for OS in previously**
 234 **untreated advanced intermediate- to poor-risk aRCC patients treated with nivolumab combined with**
 235 **ipilimumab vs sunitinib based on the data from the CheckMate 214 trial.**

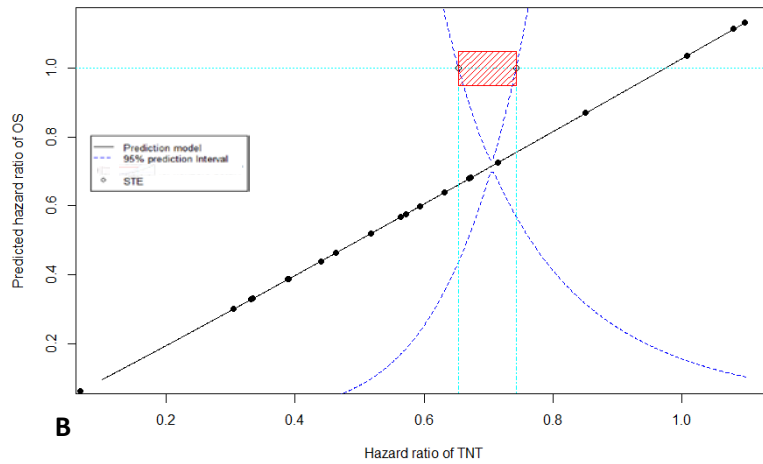
Criteria	Level of surrogacy	Point estimates [95%CI]
Joint surrogate model		
Kendall's τ	Individual	0.59 [0.56; 0.64]
R^2_{trial}	Trial	1.00 [0.97; 1.03]
STE (expressed in HR)	Trial	0.70
LCVα		9.75
Joint frailty-copula model (Clayton copula)		
Kendall's τ	Individual	0.70 [0.67; 0.74]
R^2_{trial}	Trial	1.00 [0.98; 1.02]
STE (expressed in HR)	Trial	>0.65 and <0.74
LCVα		8.05
Joint frailty-copula model (Gumbel copula)		
Kendall's τ	Individual	0.62 [0.59; 0.65]
R^2_{trial}	Trial	1.00 [0.80; 1.20]
STE (expressed in HR)	Trial	- [†]
LCVα		8.02

236 All values are rounded to 2 digits; CI: confidence interval; HR: Hazard ratio; LCV α : approximate likelihood cross-validation criterion; STE:
 237 Surrogate threshold effect; [†]No estimated value for the STE, the STE equation could not be solved. Upper bounds of the confidence intervals of
 238 R^2_{trial} cross 1 due to the delta method.
 239
 240
 241

242 Results from the three models were congruent. Individual association between TNT and OS was deemed
 243 acceptable (Kendall's τ ranged from 0.59, 95% CI [0.56; 0.64] to 0.70, 95% CI [0.67; 0.74]) (Table 2). At
 244 the trial-level, the coefficients of determination were close to 1 with relatively narrow confidence intervals,
 245 indicating a strong association between the treatment effects on TNT and on OS. Based on the $LCV\alpha$, the
 246 joint frailty copula models were those presenting the best goodness of fit, with similar $LCV\alpha$ between the
 247 Clayton and Gumbel copula approaches. Considering these acceptable correlations at both the individual
 248 and the trial levels, we assessed the STE (Table 2). Results on the STE slightly differed between the models.
 249 The joint frailty Gumbel copula model did not provide an estimate of STE. The joint surrogate and the joint
 250 frailty Clayton copula models provided similar threshold HR for TNT that would translate into OS benefit
 251 (0.70 and 0.74, respectively). The joint frailty Clayton copula model predicted a significant treatment effect
 252 on OS only when the HR for TNT was between 0.65 and 0.74. The wide 95% prediction limits of the
 253 predicted treatment effect from the joint frailty Clayton copula model is presented in Figure 3.



254



255

256 **Figure 3. Predictions of the treatment effect on OS based on the treatment effect on TNT for nivolumab in**
 257 **combination with ipilimumab versus sunitinib according to the joint surrogate model (A), and the joint frailty**
 258 **Clayton copula model (B).**

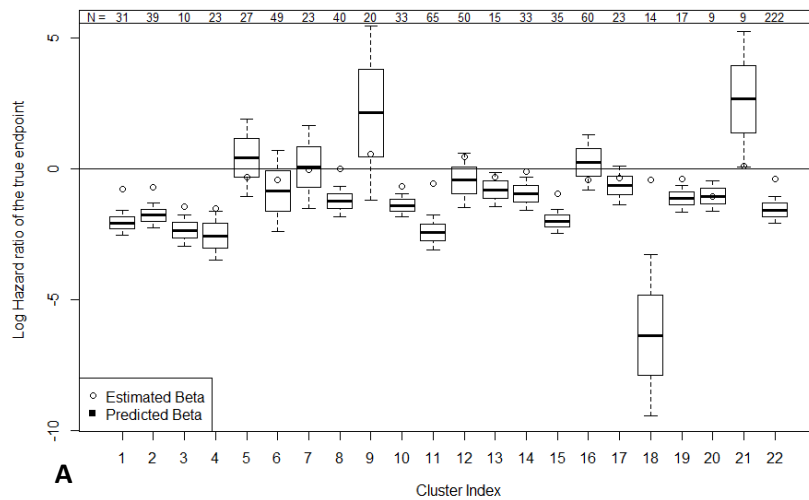
259 STE: surrogate threshold effect. The horizontal cyan line corresponds to a null hazard ratio (HR) for overall survival
 260 (OS), i.e., HR = 1. The vertical line crosses the upper boundary of the 95% prediction limit where the HR for OS is 1.
 261 The corresponding value on the x-axis corresponds to the STE. In (B), the model predicted a significant treatment on
 262 OS only when HR for TNT was between 0.65 and 0.74 due to wide 95% prediction limits.
 263

264 *Sensitivity analyses*

265 In supplementary Tables S1 to S3, we reported results of the LOOCV analyses. We faced convergence
 266 issues, more frequently with the copula approaches than with the joint surrogate model (convergence issues
 267 in the LOOCV analysis (N=22) for the joint frailty Clayton copula model: N=4; for the joint frailty Gumbel
 268 copula model: N=7; and for the joint surrogate model: N=0). Overall, for the three models, cross-validation
 269 results showed that, when convergence was reached, point estimates of the association measures were
 270 minimally influenced when the analyses were repeated removing one cluster iteratively. However,
 271 confidence intervals of the R^2_{trial} from the copulas models were slightly wider than those from the joint
 272 surrogate model. The different STE assessed through the LOOCV analysis ranged from 0.66 to 0.72 for the
 273 joint surrogate model. For the joint frailty copula models, the STE results were similar to those observed in
 274 the primary analysis.

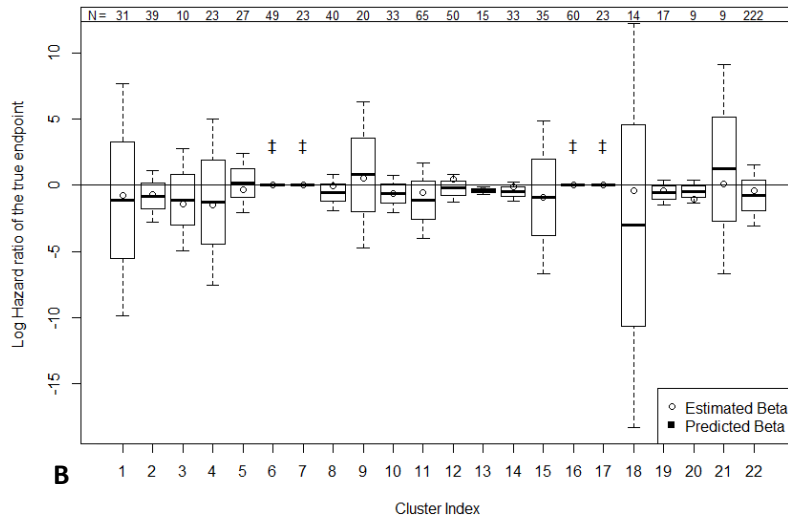
275 The ability of the models to predict the treatment effect on OS based on the treatment effect on TNT in

276 each cluster was different between the joint frailty and the copula-based models (Figure 4). The predictive
 277 accuracy of the joint frailty Clayton copula model was improved compared to the 2 others approaches
 278 (observed HR for OS included in the prediction intervals of the predicted HR for OS for 18 clusters),
 279 although the prediction intervals were not narrow. The predictive accuracy of the joint surrogate model was
 280 modest (observed HR for OS included in the prediction intervals of the predicted HR for OS for only 10
 281 clusters), with overall optimistic predictions of the treatment effect on OS. The joint frailty Gumbel copula
 282 model provided extremely wide confidence intervals for the prediction of the treatment effect on OS in 6
 283 clusters. The joint frailty Clayton copula approach was thus deemed the most appropriate surrogacy model
 284 in this study.



285

286



287

288 **Figure 4. Prediction of the log (HR) for OS based the log (HR) for TNT vs the observed log (HR) in each cluster**
289 **by leaving one cluster out at a time (LOOCV) for nivolumab combined with ipilimumab versus sunitinib. In**
290 **(A), results from the joint surrogate model; in (B), results from the joint frailty Clayton copula model; in (C),**
291 **results from the joint frailty Gumbel copula model.**

292 For each cluster, according to the LOOCV method, the surrogate models were estimated by excluding the cluster of
293 interest. Dotted bars are the 95% prediction intervals; “Estimated Beta”: log (HR) estimated by a simple Cox model;
294 “Predicted Beta”: log (HR) estimated by a joint frailty model; ‡: model did not converge. #: due to extremely wide
295 confidence intervals, the predictions were not displayed.

296

297 Discussion

298 This is the first study assessing the surrogacy properties of TNT for OS in ICI-treated aRCC patients. We
299 conducted surrogacy assessment based on IPD, relied on two distinct validated statistical approaches for
300 the assessment of the patient- and trial-level associations, and reported our findings following international
301 guidelines.²¹

302 Based on the results of this study, TNT may be an appropriate SE for OS in open-label RCT of previously
303 untreated aRCC patients treated with ICI. Results from the different statistical approaches and the
304 sensitivity analyses confirmed our findings. The joint frailty Clayton copula model was the one presenting
305 the best predictive accuracy and goodness of fit. The relatively large CIs of the prediction of the treatment
306 effect on OS are due to the limited sample size of patients in some clusters. The CIs of the estimated HR
307 for OS in each cluster were also not narrow and in many of the clusters the upper bound of the CIs were
308 above 1. These CIs yielded an uncommon interval for the prediction of the significant treatment effect on
309 OS, one expecting a unique STE as provided in the joint surrogate model. For an additional assessment of
310 the accuracy of the predictions, we wanted to pursue this work by predicting the HR for OS from published
311 HR for TNT in ICI-treated aRCC patients. As no publication of RCT of ICI combination in aRCC reported
312 TNT, we used IMDC favorable-risk patients who were enrolled in CheckMate 214, but who were not
313 included in the primary efficacy population (N=249). HR for TNT in this population was estimated to 0.85
314 (95% CI: [0.64; 1.14]) (Figure S1). Based on this estimate, the models would predict a non-significant OS
315 benefit for nivolumab combined with ipilimumab versus sunitinib. In CheckMate 214, HR for OS for the
316 favorable risk group was not significant, estimated to 1.19 (95% CI: [0.77; 1.85]).¹⁴ From our perspective,
317 this constitutes an additional argument in favor of the STEs assessed in this study.

318 TNT is a clinically relevant endpoint, and highly correlated with OS in ICI-treated aRCC patients. It is a
319 pragmatic and often measurable endpoint in all randomized subjects, reaches maturity (i.e. median) earlier
320 than the OS outcome, and may provide greater statistical power at the time of analysis.⁵ TNT reflects the
321 result of a therapeutic medical decision, a change in treatment usually occurring in response to a real change
322 in the patient status by integrating the efficacy and toxicity components.^{23,24} However, TNT may present
323 weaknesses such as the potential influences of the prescribing patterns of individual physicians.²⁵ The

324 availability of some therapies across geographic regions may also drive treatment decisions.²⁵ However,
325 TNT is a comprehensive assessment of the unique outcomes observed with ICI, as it may take into account
326 the time between interruption of the index treatment and the delay to initiation of the next systemic
327 treatment, as the patients still benefit from the ICI therapy through the prolonged treatment effect.⁵
328 In RCTs of aRCC patients, progression-free survival (PFS) is a common efficacy endpoint. We did not
329 assess the surrogacy properties of PFS since there is already evidence that PFS is weakly correlated with
330 OS in ICI-treated patients.²⁶ This may be due to the pseudo-progression and delayed treatment effect
331 phenomena, frequently reported in ICI-treated patients.^{27,28,29,30} We focused on TNT, because, by having
332 the subsequent systemic treatment initiation as an endpoint, TNT has a greater overlap in the events shared
333 with OS since it captures all patients dying prior to progression/subsequent-line of anti-cancer treatment
334 just as PFS would, but also those who progress but are never treated by a subsequent treatment. This
335 increased number of shared events is likely to yield a higher correlation.

336 From a methodological point of view, both innovative statistical approaches provided consistent results and
337 similar conclusions. However, the one-step joint frailty copula model might be the approach to be privileged
338 since: 1) predictions of HR for OS based on HR for TNT were much closer than those in the one-step joint
339 surrogate model; 2) the computational time was shorter with this approach, since less parameters have to
340 be estimated. Both surrogacy approaches definitely limit the issues faced with the copula model-based
341 approach introduced by Burzykowski et al., and applying both approaches might be relevant for confirming
342 results.

343 One characteristic of ICI agents is the potential delayed treatment effect due to the mechanism of action.³⁰
344 Therefore, HR might not properly represent the treatment benefit due to non-proportional hazards of
345 survival and long-term survival with ICI.³¹ All the surrogacy models currently available are based on Cox

346 model, in which the prerequisite is proportional hazards assumption. The use of the difference in *restricted*
347 *mean survival time* (RMST), a new metric for quantifying the treatment efficacy not dependent upon the
348 proportional hazards assumption, has been recently discussed for consideration in RCT of ICI-treated
349 patients.³⁰ The question of the surrogacy value of RMST is being explored in ovarian cancer,³² and may be
350 worth evaluating in other tumor types. In this dataset, the HR proportionality assumption was violated in
351 one cluster for TNT only (Figure S2).

352 The limited amount of data is the main limitation of this study. However, RCT of ICI combination in aRCC
353 patients are limited. The estimated R^2_{trial} were close to 1 and the confidence intervals were narrow. These
354 values, although high, are frequently reported.^{10,33} The point estimates and their confidence intervals need
355 however to be interpreted with caution since data from only one RCT was available for this study. The
356 cluster-level definition was therefore the country of enrollment instead of RCT as usually advised in the
357 meta-analytic framework. The heterogeneity in the treatment effects between clusters in this RCT might be
358 limited compared to that of traditional meta-analytic approaches. This may yield an overestimation of the
359 correlation at the cluster-level, and may have artificially narrowed the reported confidence intervals.

360 Since the validity of an SE depends both on the disease setting and on the mechanism of action of the
361 treatment,⁸ we recommend to estimate TNT and to assess the surrogacy properties of TNT in aRCC patients
362 treated with ICI in combination with vascular endothelial growth factor (VEGF) inhibitors, another
363 therapeutic option recently available in aRCC. Such combinations were assessed through three distinct
364 RCTs,^{34,35,36} the sample size for the surrogacy analysis would be larger than in this study. Finally,
365 conducting similar surrogacy analyses in ICI-treated patients with other tumor types would definitely be
366 relevant for evaluating whether TNT is also a candidate SE for OS in solid tumors other than advanced
367 melanoma and aRCC.

368 **Funding**

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

371

372

373

374 **Table S1. Sensitivity analyses in the Kendall's τ and R^2_{trial} values assessed with the joint surrogate model by removing one cluster iteratively.** All values are rounded to
 375 2 digits; CI: confidence interval; STE: surrogate threshold effect; HR: hazard ratio.

Cluster excluded	N of patients in the cluster	Kendall's τ Point estimates [95%CI]	R^2_{trial} Point estimates [95%CI]	STE (expressed in HR)
#1	31	0.57 [0.54; 0.60]	1.00 [0.97; 1.03]	0.69
#2	39	0.59 [0.55; 0.63]	1.00 [0.96; 1.03]	0.71
#3	10	0.59 [0.56; 0.64]	1.00 [0.97; 1.03]	0.71
#4	23	0.59 [0.53; 0.63]	1.00 [0.99; 1.01]	0.67
#5	27	0.59 [0.56; 0.63]	1.00 [0.96; 1.04]	0.66
#6	49	0.58 [0.54; 0.60]	1.00 [0.97; 1.03]	0.70
#7	23	0.59 [0.56; 0.63]	1.00 [0.96; 1.04]	0.66
#8	40	0.59 [0.56; 0.63]	1.00 [0.98; 1.02]	0.69
#9	20	0.57 [0.53; 0.60]	1.00 [0.97; 1.03]	0.63
#10	33	0.59 [0.56; 0.64]	1.00 [0.97; 1.03]	0.72
#11	65	0.60 [0.56; 0.64]	1.00 [0.97; 1.03]	0.72
#12	50	0.59 [0.56; 0.64]	1.00 [0.98; 1.02]	0.71
#13	15	0.59 [0.55; 0.64]	1.00 [0.97; 1.02]	0.70
#14	33	0.57 [0.54; 0.61]	1.00 [0.99; 1.01]	0.69
#15	35	0.59 [0.56; 0.64]	1.00 [0.97; 1.03]	0.71
#16	60	0.60 [0.56; 0.63]	1.00 [0.98; 1.02]	0.71
#17	23	0.59 [0.56; 0.63]	1.00 [0.98; 1.02]	0.69
#18	14	0.60 [0.57; 0.64]	1.00 [0.98; 1.02]	0.71
#19	17	0.59 [0.56; 0.63]	1.00 [0.97; 1.03]	0.70
#20	9	0.59 [0.56; 0.64]	1.00 [0.97; 1.02]	0.68
#21	9	0.59 [0.56; 0.64]	1.00 [0.97; 1.03]	0.68
#22	222	0.59 [0.56; 0.62]	1.00 [0.96; 1.03]	0.72

376
377

378 **Table S2. Sensitivity analyses in the Kendall's τ and R^2_{trial} values assessed with the joint frailty Clayton copula model by removing one cluster iteratively.** All values
 379 are rounded to 2 digits; CI: confidence interval; STE: surrogate threshold effect; HR: hazard ratio.

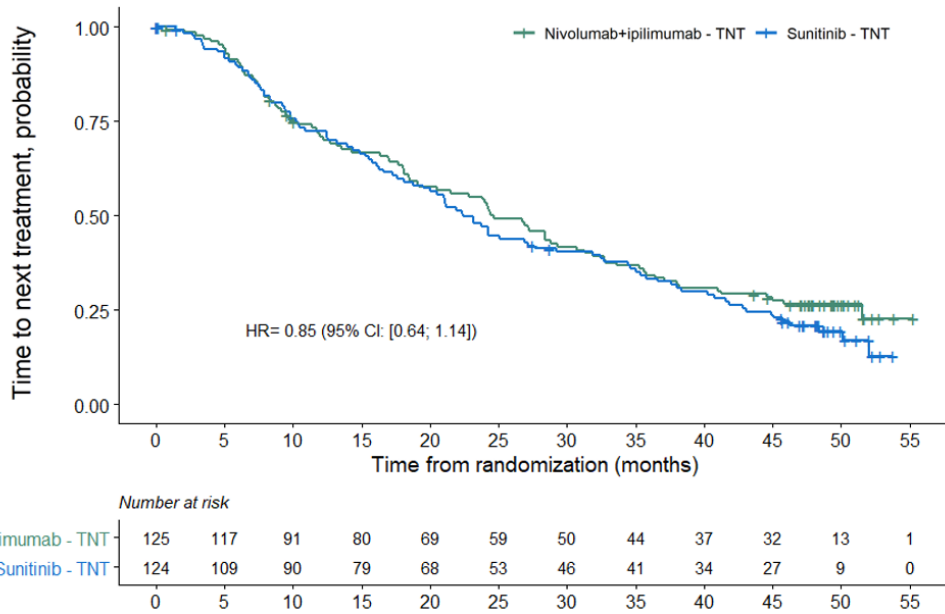
Cluster excluded	N of patients in the cluster	Kendall's τ Point estimates [95%CI]	R^2_{trial} Point estimates [95%CI]	STE (expressed in HR)
#1	31	0.71 [0.67; 0.74]	1.00 [0.81; 1.20]	-
#2	39	0.70 [0.67; 0.74]	1.00 [0.98; 1.02]	> 0.64 and < 0.76
#3	10	0.71 [0.67; 1.00]	1.00 [1.00; 1.00]	> 0.66 and < 0.75
#4	23	0.70 [0.66; 0.73]	1.00 [0.86; 1.14]	-
#5	27	0.70 [0.67; 0.74]	1.00 [0.95; 1.05]	> 0.64 and < 0.74
#6	49	<i>Model did not converge</i>		
#7	23	<i>Model did not converge</i>		
#8	40	0.71 [0.67; 0.75]	0.98 [0.45; 1.50]	> 0.65 and < 0.72
#9	20	0.70 [0.66; 0.74]	1.00 [0.97; 1.03]	> 0.63 and < 0.73
#10	33	0.71 [0.67; 0.75]	1.00 [1.00; 1.00]	> 0.65 and < 0.76
#11	65	0.71 [0.68; 0.75]	1.00 [0.98; 1.02]	> 0.65 and < 0.78
#12	50	0.72 [0.68; 0.75]	1.00 [0.99; 1.01]	> 0.61 and < 0.73
#13	15	0.70 [0.66; 0.74]	1.00 [0.98; 1.02]	> 0.65 and < 0.75
#14	33	0.71 [0.67; 0.75]	1.00 [0.89; 1.10]	> 0.66 and < 0.74
#15	35	0.71 [0.68; 0.75]	1.00 [0.80; 1.19]	-
#16	60	<i>Model did not converge</i>		
#17	23	<i>Model did not converge</i>		
#18	14	0.71 [0.67; 0.74]	1.00 [0.90; 1.09]	> 0.66 and < 0.76
#19	17	0.71 [0.67; 0.75]	1.00 [0.97; 1.03]	> 0.64 and < 0.74
#20	9	0.70 [0.67; 0.74]	1.00 [0.98; 1.02]	> 0.65 and < 0.75
#21	9	0.70 [0.66; 0.74]	1.00 [0.87; 1.13]	> 0.64 and < 0.73
#22	222	0.70 [0.67; 0.74]	1.00 [0.98; 1.02]	> 0.66 and < 0.74

380

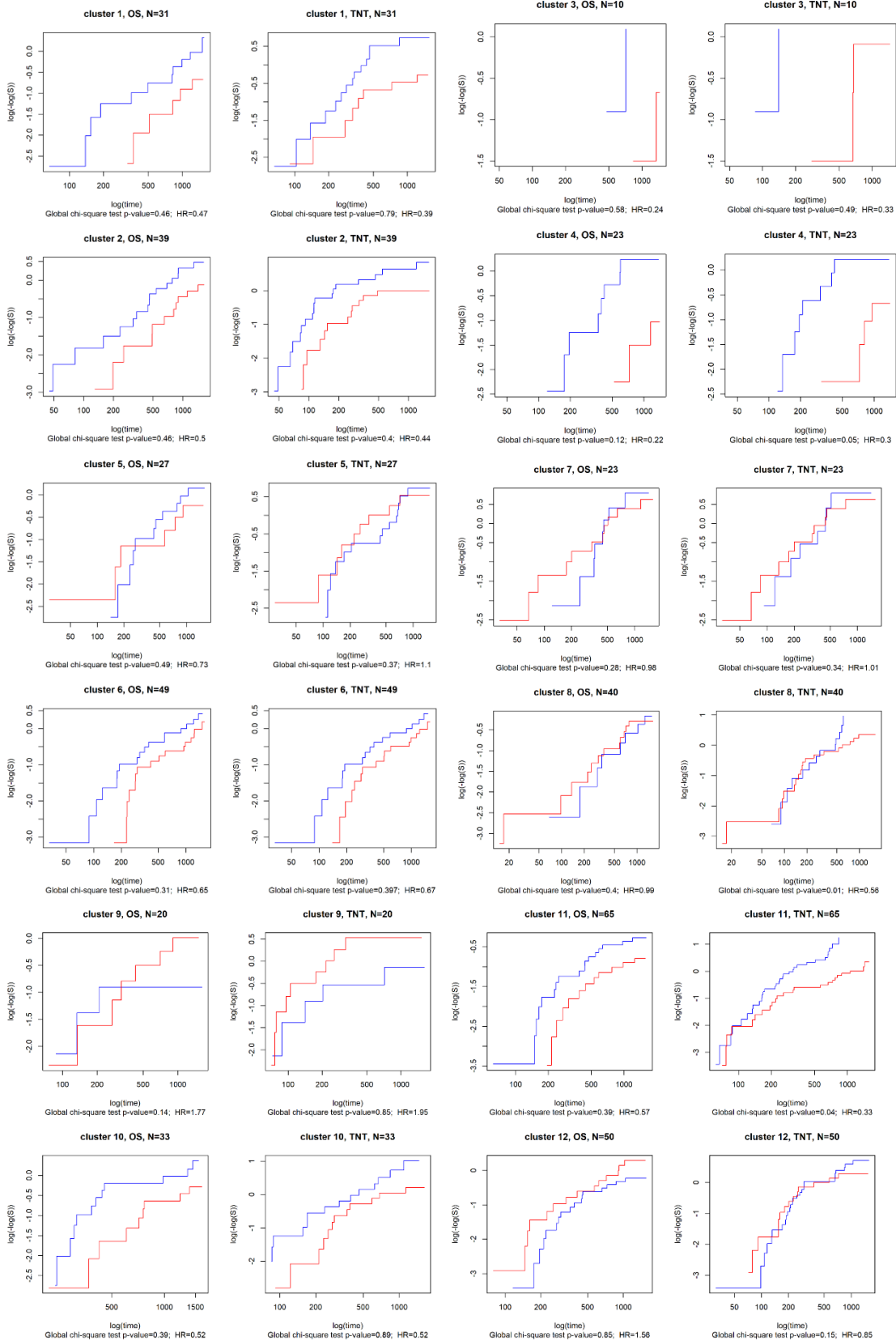
381 **Table S3. Sensitivity analyses in the Kendall's τ and R^2_{trial} values assessed with the joint frailty Gumbel copula model by removing one cluster iteratively.** All values
 382 are rounded to 2 digits; CI: confidence interval; STE: surrogate threshold effect; HR: hazard ratio.

Cluster excluded	N of patients in the cluster	Kendall's τ Point estimates [95%CI]	R^2_{trial} Point estimates [95%CI]	STE (expressed in HR)
#1	31	0.62 [0.59; 0.65]	1.00 [0.98; 1.02]	-
#2	39	0.62 [0.58; 0.65]	0.99 [0.74; 1.25]	-
#3	10	0.62 [0.59; 0.65]	1.00 [0.87; 1.13]	-
#4	23	<i>Model did not converge</i>		
#5	27	<i>Model did not converge</i>		
#6	49	0.60 [0.56; 0.64]	1.00 [0.78; 1.21]	-
#7	23	<i>Model did not converge</i>		
#8	40	<i>Model did not converge</i>		
#9	20	<i>Model did not converge</i>		
#10	33	0.62 [0.59; 0.66]	1.00 [0.96; 1.04]	-
#11	65	0.62 [0.59; 0.66]	1.00 [0.85; 1.15]	>0.59 and <0.68
#12	50	<i>Model did not converge</i>		
#13	15	0.62 [0.59; 0.65]	1.00 [0.94; 1.06]	-
#14	33	0.62 [0.59; 0.66]	1.00 [0.90; 1.10]	-
#15	35	0.63 [0.59; 0.66]	1.00 [0.94; 1.06]	-
#16	60	0.64 [0.60; 0.67]	1.00 [0.96; 1.04]	>0.46 and <0.66
#17	23	0.61 [0.57; 0.64]	0.99 [0.39; 1.60]	-
#18	14	0.62 [0.59; 0.65]	1.00 [0.88; 1.12]	-
#19	17	<i>Model did not converge</i>		
#20	9	0.62 [0.58; 0.65]	1.00 [0.86; 1.14]	-
#21	9	0.62 [0.58; 0.65]	1.00 [0.64; 1.36]	-
#22	222	0.62 [0.59; 0.66]	1.00 [0.81; 1.19]	-

383



384
 385 **Figure S1. Kaplan-Meier curves for TNT for nivolumab combined with ipilimumab and for sunitinib in**
 386 **previously untreated advanced favorable-risk aRCC patients from CheckMate-214 trial. HR=hazard ratio; TNT:**
 387 **time to next treatment; '+' signs on the Kaplan-Meier curves represent censorings.**
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405

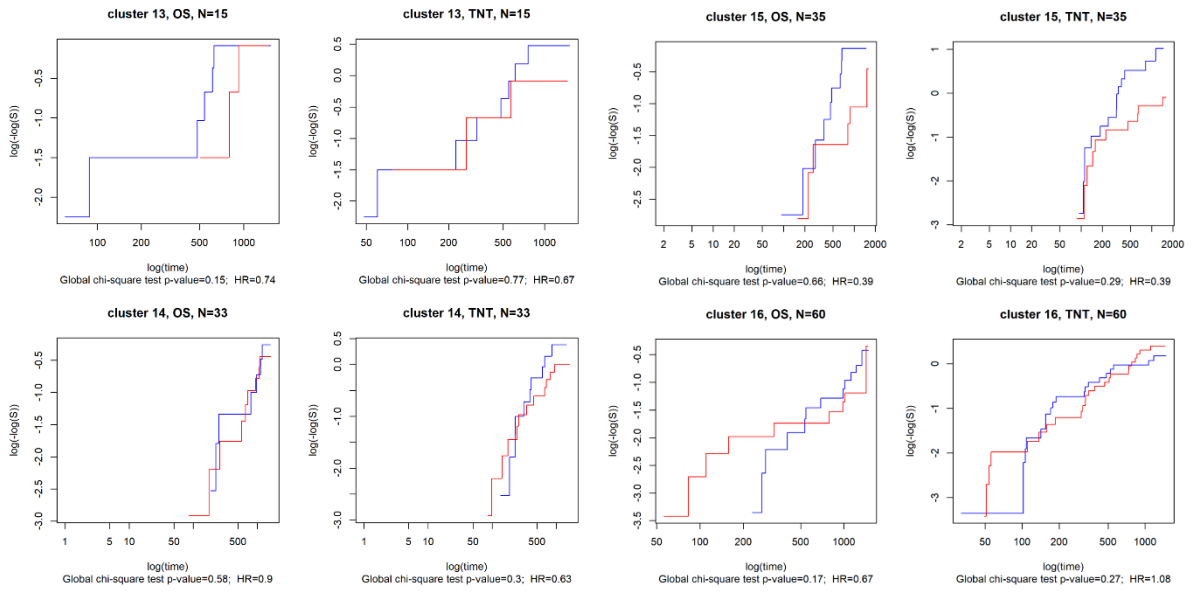


406

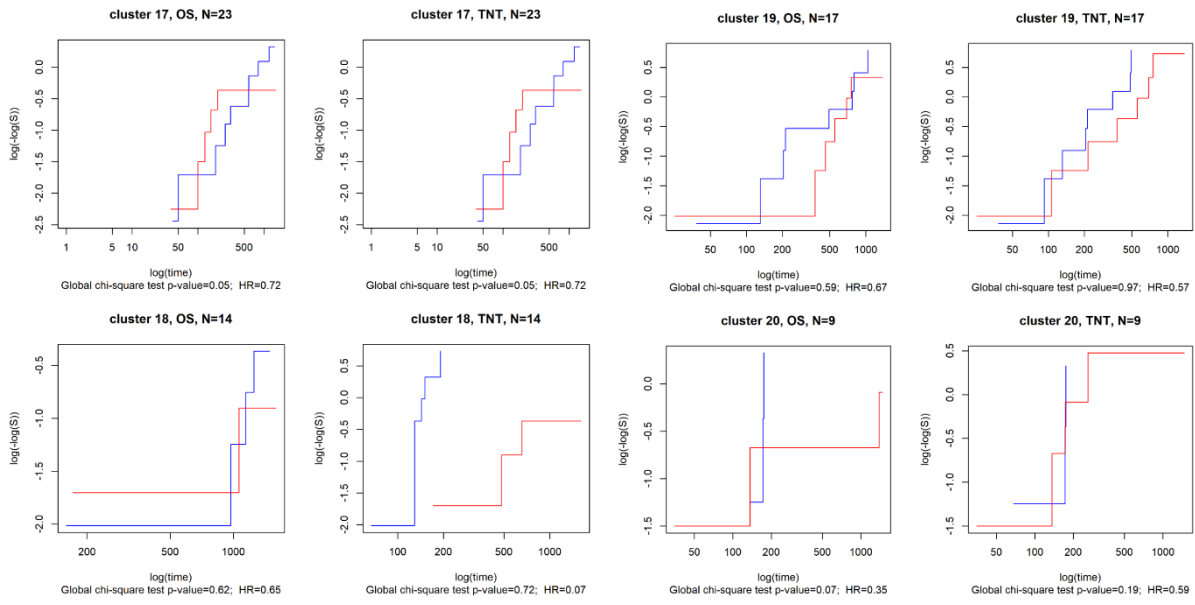
407

408
409
410
411

412



413



414

415

416

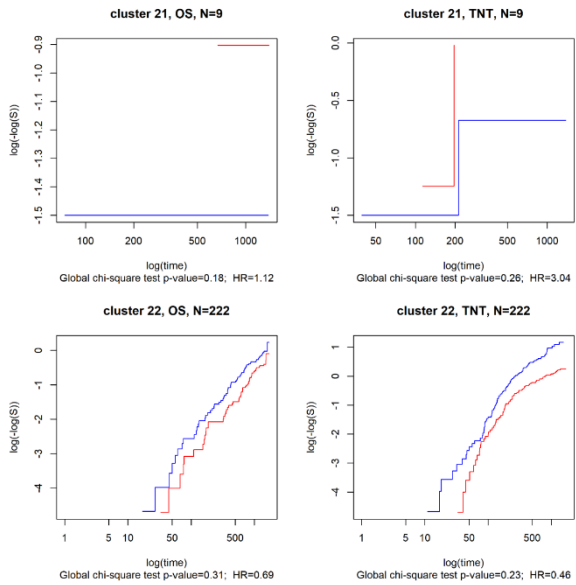


Figure S2. Evaluation of the proportional-hazards assumption by cluster for TNT for nivolumab combined with ipilimumab and for sunitinib in previously untreated advanced intermediate- to poor-risk aRCC patients from CheckMate-214 trial.

References

1. Takeo Fujii T, Naing A, Rolfo C, Hajjar J. Biomarkers of response to immune checkpoint blockade in cancer treatment. *Crit Rev Oncol Hematol* 2018; **130**:108-120.
2. Harper K. Checkpoint Inhibitors Spur Changes in Trial Design. *Cancer Discov* 2017; **7** (11): 1209-10.
3. U.S. Department of Health and Human Services, Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry; 2018. Available from: <https://www.fda.gov/media/71195/download>
4. European Medicine Agency. Guideline on the evaluation of anticancer medicinal products in man, 2017. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man-revision-5_en.pdf
5. Branchoux S, Sofeu C, Gandin AF et al. Time to next treatment 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. *Submitted to J of Clin Oncol*
6. Fleming TR, Powers JH. Biomarkers and Surrogate Endpoints in Clinical Trials. *Stat Med* 2012; **31**(25): 2973–84.
7. Burzykowski T, Molenberghs G, Buyse, M. The Evaluation of Surrogate Endpoints. New-York: Springer-Verlag, 2005.
8. 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 of Oncol*. 2019; **30**(5):v393.
9. Burzykowski T, Molenberghs G, Buyse M et al. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *Appl Statist* 2001; **50**: 405-22.
10. Sofeu 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):2928-42.
11. 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.
12. Sofeu 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; Online ahead of print doi: 10.1002/bimj.201900306.
13. Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med* 2018; **378**:1277-90.
14. Motzer RJ, Escudier B, McDermott DF, et al. Survival outcomes and independent response assessment with nivolumab plus ipilimumab versus sunitinib in patients with advanced renal cell carcinoma: 42-month follow-up of a randomized phase 3 clinical trial. *J Immunother Cancer* 2020; **8**:e000891.
15. Alonso A, Bigirimurame T, Burzykowski T et al. Applied surrogate endpoint evaluation methods with SAS and R. Chapman & Hall CRC biostatistics series, 2016.
16. 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):6139-48.
17. Coart E, Suci 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**:171-74.
18. Sofeu CL, Rondeau V. How to use frailtypack for validating failure-time surrogate endpoints using individual patient data from meta-analyses of randomized controlled trials. *Plos One* 2020; **15**(1):e0228098.

19. Burzykowski T, Buyse M. Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceut Statist* 2006; **5**: 173–86.
20. Commenges, D, Joly P, Gégout-Petit A, and Liqueur B. Choice between semi-parametric estimators of Markov and non-Markov multi-state models from coarsened observations. *Scand J of Stat* **34**:33–52.
21. Xie W, Halabi S, Tierney JF, et al. A systematic review and recommendation for reporting of surrogate endpoint evaluation using meta-analyses. *JNCI Cancer Spectrum*. 2019; **3**(1)pkz002.
22. Krol A, Mauguen A, Mazroui Y et al. Tutorial in Joint Modeling and Prediction: a Statistical Software for Correlated Longitudinal Outcomes, Recurrent Events and a Terminal Event. *J Stat Soft* 2017; **81**(3):1-52.
23. Marshall J, Schwartzberg LS, 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**(15S): 11521.
24. 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 Medicine* 2017; **15**(1):78.
25. Campbell B A, Scarisbrick J J , Kim Y H. Time to next treatment as a meaningful endpoint for trials of primary cutaneous lymphoma. *Cancers* 2020; **12**(8): 2311.
26. Branchoux S, Bellera C, Italiano A et al. 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.
27. Mushti SL, Mulkey F and 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):2268–75.
28. 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 Network Open* 2018; **1**(2):e180416.
29. Kaufman H, Schwartz L, William Jr W et al. Evaluation of classical 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):2245-61.
30. Buyse M, Saad ED, Burzykowski T. Assessing Treatment Benefit in Immuno-oncology. *Stat Biosci* 2020. **12**:83–103.
31. Mazarella L, Morganti S, Marra A et al. Master protocols in immuno-oncology: do novel drugs deserve novel designs? *Journal ImmunoTher Cancer* 2020; **8**:e000475.
32. Paoletti X, Lewsley L-A, Daniele G et al. Assessment of progression-free survival as a surrogate endpoint of overall survival in first-line treatment of ovarian cancer: a systematic review and meta-analysis. *JAMA Netw Open* 2020; **3**(1):e1918939.
33. Rotolo F, Paoletti X, Burzykowski T et al. A Poisson approach to the validation of failure time surrogate endpoints in individual patient data meta-analyses. *Stat Methods Med Res* 2019; **28**(1):170-83.
34. Rini BI, Plimack ER, Stus V et al. Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *N Engl J Med* 2019; **380**:1116-27.
35. Motzer RJ, Penkov K, Haanen J et al. Avelumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *N Engl J Med* 2019; **380**:1103-15.
36. Choueri TK, Powles T, Burotto M, et al. Nivolumab plus cabozantinib versus sunitinib in first-line treatment for advanced renal cell carcinoma: first results from the randomized phase 3 CheckMate 9ER trial. *Ann of Oncol* 2020; **31**(S4): S1142-S1215.

4.5 Additional discussion

TNT seems a promising candidate SE in previously untreated aRCC patients. We had the opportunity to assess the surrogacy properties of TNT in a more advanced setting, i.e. in previously treated aRCC patients, based on CheckMate 025 trial. CheckMate 025 is a randomized, open-label, phase III trial of nivolumab monotherapy versus everolimus monotherapy, conducted in previously treated aRCC patients.⁶³

The Kendall's τ coefficient of correlation between TNT and OS was assessed to 0.49 (95% CI: [0.44; 0.55]). Based on this relatively weak association at the individual level, TNT might not be a valuable SE in previously treated aRCC patients. In previously treated setting, fewer patients are likely to experience a durable response and a time free of systemic anticancer therapy versus previously untreated patients. Also, in CheckMate 025, a cross-over from everolimus arm to nivolumab arm was permitted once the OS benefit was shown. These are the suggested explanations of this weaker correlation compared to in previously untreated aRCC patients.

5 General Discussion

Cancer patients are living longer due to improved therapies, powering a study for OS will become increasingly challenging.²⁷ Endpoints other than OS are therefore important for regulatory approval of anticancer agents and it is of importance to identify and validate novel SE for OS for ICI.²⁰ SE are particularly relevant while it take a significant period of time to demonstrate a significant OS benefit or to get mature survival data, like in advanced melanoma and renal cell carcinoma (aRCC).

In the first part of this work, we summarized the current evidence on clinical alternative endpoints associated with OS in ICI-treated patients. Overall, there was insufficient data to support validated SE for OS in this specific population. ICI development is quite recent and it may take an important period of time before surrogacy analyses could be performed. Moreover, access to IPD, required for applying an appropriate surrogacy approach, is challenging. These may be the two main reasons for the absence of published research works in this field (only 24 references were retrieved).

Recently, the immune-related response criteria, the iRECIST, were developed considering some immunotherapy-specific phenomena, such as delayed treatment effect and pseudo-progression.²⁹ We did not assess the surrogacy properties of iPFS in the frame of this doctoral thesis since they were some evidence that the association between iPFS and OS was weak. Indeed, Mulkey et al. recently conducted a retrospective assessment of response which was defined according to RECIST 1.1 or iRECIST in 4,751 ICI-treated patients by pooling RCTs data from multiple cancer types.⁶⁴ Despite refined consideration of these ICI-specific criteria, correlations between PFS and OS did not improve substantially.

Even though our systematic literature review was inconclusive over validated SE for OS in ICI-treated patients, it helped us on appraising the important characteristics that an SE for OS in this population should account for, such as the duration of response (or a proxy of it).

In the second part of this doctoral thesis, we leveraged learnings from the systematic literature review (i.e. weak correlation of tumor-based intermediate endpoints, relevance of the duration of response) and from the performance of ICI agents during the TFI. Therefore, we assessed the surrogacy properties of TNT in advanced melanoma patients considering an innovative statistical approach. Even though the available data were limited (1 RCT of 945 patients), this is the first research work, based the meta-analytic approach, suggesting important individual- and trial-level correlations between an intermediate endpoint and OS in ICI-treated patients with advanced cancer. The last objective of this thesis was to assess whether this alternative endpoint may be appropriate in another

tumor type. We applied two different statistical approaches in aRCC, yielding similar conclusive results.

It was of interest to pursue this work by estimating HR for OS from HR for TNT based on published RCT data in advanced melanoma and aRCC in order to assess the quality of prediction of the developed models, as a complement to the leave-one-out cross-validation analyses. Unfortunately, HR for TNT were not reported in any published RCT of drugs sharing the same mechanism of action (MoA) and conducted in advanced melanoma or aRCC patients.

Results from this doctoral thesis are promising, and could be considered as the first conclusive step in the research of SE in ICI-treated patients with advanced or metastatic cancer. From our perspective, TNT fulfills the main requested criteria for being an SE: (i) to be easily measurable, (ii) to predict the course of the disease in an individual patient, (iii) to predict the effect of a therapy on the clinical endpoint of interest. Moreover, TNT takes into account the specificity of ICI, such as durable response. To date, only one published RCT of ICI-treated patients considered TNT as a secondary endpoint,⁶⁵ although this endpoint is not listed as a relevant clinical endpoint by the Food and Drug Administration nor by the European Medicines Agency in their guidance for the approval of cancer drugs.^{35,36} In these guidance documents, time to treatment failure (TTF), defined as time from randomization to discontinuation of therapy for any reason, is listed as a relevant endpoint although not generally recommended as a regulatory one. We did not assess the surrogate properties of TTF since some characteristics of this metric would not allow TTF to be a candidate SE in ICI-treated patients. For instance, in this population, treatment discontinuation due to toxicity is not uncommon and patients may continue to experience long-term disease control even after cessation of ICI therapy.⁶⁶ There are also some recent discussions around on the importance of the *time in response*, another efficacy measure of ICIs, as a relevant endpoint in ICI-treated patients.⁶⁷ This endpoint includes the intention-to-treat population in the analysis, in which non-responders are included with an outcome of “zero” duration of response. In order to assess the surrogacy properties of this clinical endpoint, new surrogate approaches need to be developed, since the statistical methods currently available could not be applied when there is an important number of “zero” in the data. From our perspective, this metric is partially but indirectly captured by TNT as non-responders are likely to initiate a subsequent treatment or die, while responders are not.

Even though TNT presents multiple advantages over some traditional clinical endpoints, it is not a perfect endpoint. Perceived weaknesses of TNT are the potential influences of the prescribing patterns of individual physicians, especially when selecting the timing of switching over.⁶⁸ Differences in multidisciplinary care may also exist across geographic regions: the availability of

treatments may drive treatment decisions and lead to evolving prescription.⁶⁸ Also, we recommend to collect in the case report form (CRF) of future RCTs detailed information after discontinuation of index therapy in RCT such as the initiation date of subsequent therapy (since it has been reported to be heterogeneously captured⁶⁹), or the reason of not receiving subsequent treatment. Bristol Myers Squibb already modified the CRFs in order to routinely gather both information in the frame of RCTs.

Surrogacy analyses considering multiple RCTs (i.e. through the conventional meta-analytic approach) of ICI-treated patients are warranted for confirming our findings. It is of importance to have data from multiple RCTs (positive or negative RCTs on the final endpoint) for an appropriate assessment of the associations. Academic researchers who could solicit sponsors to get access to RCT data, or researchers for health regulatory authorities (HRA), who have a direct access to RCT data, may be interested in running similar analyses (and relevant subgroup analyses) considering multiple RCTs of advanced melanoma or aRCC patients treated with ICI. Then, if the results are conclusive, an updated HRA guidance documentation for the approval of anticancer drugs might be relevant for suggesting to estimate TNT. In such case, strong effect on TNT of the ICI agent(s) might expedite regulatory approval. Submission to HRA considering the preliminary data would be performed, and, later on, OS data (which remain crucially important) would be provided for confirming the treatment benefit. As an estimate of the time which could be saved, median TNT was reached approximately three times sooner than median OS in CheckMate 067 and in CheckMate 214.

Beyond the regulatory approval of the ICI agents, we see multiple perspectives for TNT or field of research considering this endpoint.

For instance, within health economic evaluations of anticancer drug, survival modelling is required so that the survival impact of the new intervention can be taken into account alongside health related quality of life impacts.⁷⁰ Unless survival data from a clinical trial is very close to being complete, extrapolation is required such that survival data can be usefully incorporated in health economic models.⁷⁰ In case of immature OS data and strong association between TNT and OS, one could consider TNT within the health economics model for extrapolating OS. The uncertainty in the results of the economic modelling due to the consideration of the intermediate endpoint compared to the final endpoint is an interesting area of research.

Another research perspective worth highlighting is the analysis of the association between TNT and OS in real-life setting. For advanced melanoma, it might be interesting to consider the French cohort of melanoma patients (MELBASE⁷¹) and to assess whether the association between TNT and OS

is confirmed in real-life setting, although one should first consider how to handle treatment effect in the context of non-randomized patients.

Finally, in the current American Society of Clinical Oncology (ASCO) net health benefit assessment tool developed for assessing the value of cancer treatment regimens, only OS and PFS constitute the clinical benefit score of the anticancer drug.⁷² Health-related quality of life and TFI are considered as bonus points. One may want to update this tool by adding TNT into the clinical benefit score, at least as a complement to PFS for ICI agents, when TNT will be fully validated as an SE.

Results from this doctoral thesis also raise the following questions: “Is TNT a candidate SE in more advanced settings or in other tumor types? Is TNT a candidate SE for ICI in combination with chemotherapy or VEGF inhibitors?”

We tried to partially address this question using CheckMate 025 trial data. Based on the relatively weak association at the individual level, TNT might not be a valuable SE in previously treated aRCC patients. Results from this analysis suggest that TNT may be a more appropriate SE for OS in previously untreated rather than in previously treated patients. However, it would be interesting to perform similar work in previously treated advanced melanoma patients.

The other question around the appropriateness of TNT as an SE for OS for other ICI combinations (e.g. anti-PD-1 + chemotherapy, anti-PD-1 + anti-CTLA4 + chemotherapy, anti-PD-(L)1+TKI) is a promising field of research. However, before performing surrogacy analysis, one should first consider the relevance of identifying an SE in the disease setting of interest. For instance, in advanced non-small cell lung cancer (aNSCLC), it might not be particularly relevant to identify an SE as mature OS data might be available quite rapidly due to the poor prognosis of the disease. In the recent CheckMate 9LA RCT, a study comparing nivolumab combined with ipilimumab and 2 cycles of chemotherapy versus chemotherapy in 1L patients with stage IV or recurrent aNSCLC, the median OS for the ICI + chemotherapy combination was 14.1 months.⁷³

From a methodological point of view, both innovative statistical approaches used in this doctoral thesis provided consistent results and similar conclusions. However, the one-step joint frailty copula model might be the one to be privileged, since: 1) predictions of HR for OS based on HR for TNT were improved compared to those obtained from the one-step joint surrogate model (c.f. manuscript n°3); 2) the computational time was shorter with this approach (less parameters to be estimated). In any case, it is convenient to have two statistical approaches which are not strictly based on the same model, at least for using it for confirming the results. Both approaches definitely limit the issues faced with the copula model-based approach introduced by Burzykowski et al,⁹ such as model convergence issues, and large standard error of the coefficient of determination.¹²

As discussed in chapter I, one characteristic of ICI is the potential delayed treatment effect due to its MoA. Therefore, the estimated HR might not properly represent the treatment benefit due to non-proportional hazards of survival and long-term survival with ICI.^{28,74} All surrogacy models currently available are based on a Cox model, which assumes proportional hazards. We identified recent discussions around new approaches for taking into account this aspect in surrogacy analysis:

- Consideration of accelerated-failure time model;^{75,76}
- Use of a new metric for quantifying the treatment efficacy, which is not dependent upon the proportional hazards assumption: the difference in *restricted mean survival time* (RMST). RMST is an alternative measure to median survival, which quantifies the additional life expectancy due to the investigational treatment.⁷⁷ The question of the surrogacy value of RMST is being explored in ovarian cancer.⁷⁸

In our case, the proportionality assumption within each cluster for each outcome was in most the case not violated in both tumors (see manuscript n°3 for aRCC, data not shown for advanced melanoma). However, it would be of interest to perform surrogacy analyses after adaptation of the available models, or by considering difference in RMST as a measure of the treatment benefit, for confirming our findings.

6 Conclusion

Results from this doctoral thesis constitute the first conclusive step in the research of SE in ICI-treated patients with advanced or metastatic cancer. We identified a promising SE, which takes into account the specificity of ICI. We encourage to pursue this work by performing surrogacy analyses considering multiple RCTs of ICI-treated patients with advanced melanoma and aRCC, as well as conducting similar analyses in other tumor types for evaluating whether TNT is also a candidate SE for OS in other solid tumors.

7 References

-
- 1 World Health Organization. Cancer. Available from <https://www.who.int/health-topics/cancer>. Accessed on 09/22/2020.
 - 2 INCa. Les immunothérapies spécifiques dans le traitement des cancers / Rapport, France 2018, Collection états des lieux et des connaissances. Available from <https://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Les-immunotherapies-specifiques-dans-le-traitement-des-cancers-Rapport>. Accessed on 09/22/2020.
 - 3 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**:1535-46.
 - 4 Motzer RJ, Escudier B, McDermott DF, et al. Survival outcomes and independent response assessment with nivolumab plus ipilimumab versus sunitinib in patients with advanced renal cell carcinoma: 42-month follow-up of a randomized phase 3 clinical trial. *J Immunother Cancer* 2020; **8**:e000891.
 - 5 Brahmer JR, Rodriguez-Abreu D, Robinson AG. KEYNOTE-024 5-year OS update: First-line (1L) pembrolizumab (pembro) vs platinum-based chemotherapy (chemo) in patients (pts) with metastatic NSCLC and PD-L1 tumour proportion score (TPS) $\geq 50\%$. *Ann of Oncol* 2020; **31** (S4): S1142-15.
 - 6 Fleming TR, Powers JH. Biomarkers and Surrogate Endpoints in Clinical Trials. *Stat Med* 2012; **31**(25): 2973–84.
 - 7 International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. (1998). ICH harmonized tripartite guideline. Statistical principle for clinical trials. Available from: <http://www.ich.org/pdfICH/e9.pdf> Accessed on 09/22/2020.
 - 8 Burzykowski T, Molenberghs G, Buyse, M. The Evaluation of Surrogate Endpoints. New-York: Springer-Verlag, 2005.
 - 9 Burzykowski T, Molenberghs G, Buyse M et al. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *Appl Statist* 2001; **50**: 405-22.
 - 10 Rotolo F, Paoletti X, Burzykowski T et al. A Poisson approach to the validation of failure time surrogate endpoints in individual patient data meta-analyses. *Stat Methods Med Res* 2019; **28**(1):170-83.
 - 11 Renfro L, Shi Q, Sargent D, Carlin B. Bayesian adjusted R2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials. *Stat Med* 2012; **31**(8):743-61
 - 12 Sofeu 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):2928-42.
 - 13 Sofeu 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; Online ahead of print doi: 10.1002/bimj.201900306.
 - 14 Sofeu CL, Rondeau V. How to use frailtypack for validating failure-time surrogate endpoints using individual patient data from meta-analyses of randomized controlled trials. *Plos One* 2020; **15**(1):e0228098.
 - 15 Burzykowski T, Buyse M. Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceut Statist* 2006; **5**: 173–86
 - 16 Alonso A, Bigirimurame T, Burzykowski T et al. Applied surrogate endpoint evaluation methods with SAS and R. Chapman & Hall/ CRC biostatistics series, 2016.
 - 17 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 Medicine* 2017; **15**(1):78.
 - 18 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):3350-58.
 - 19 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**(5S):v393.

-
- 20 Saad ED, Buyse M. Statistical controversies in clinical research: end points other than overall survival are vital for regulatory approval of anticancer agents. *Ann Oncol* 2016; **27**(3):373-8.
- 21 Galluzzi L, Vacchelli E, Bravo-San Pedro J-M, et al. Classification of current anticancer immunotherapies. *Oncotarget* 2014; **5**(24):12472–508.
- 22 Martin-Liberal J, Hierro C, Ochoa de Olza M, Rodon J. Immuno-oncology: the third paradigm in early drug development. *Target Oncol* 2017; **12**(2):125–38.
- 23 The Nobel Assembly at Karolinska Institutet. Available at <https://www.nobelprize.org/prizes/medicine/2018/press-release>. Accessed on 09/22/2020.
- 24 European Medicine Agency. Summary of the product characteristics for ipilimumab. Available from https://www.ema.europa.eu/en/documents/product-information/yervoy-epar-product-information_en.pdf Accessed on 09/22/2020.
- 25 European Medicine Agency. Summary of the product characteristics for pembrolizumab. Available from https://www.ema.europa.eu/en/documents/product-information/keytruda-epar-product-information_en.pdf Accessed on 09/22/2020.
- 26 European Medicine Agency. Summary of the product characteristics for nivolumab. Available from https://www.ema.europa.eu/en/documents/product-information/keytruda-epar-product-information_en.pdf Accessed on 09/22/2020.
- 27 Harper K. Checkpoint Inhibitors Spur Changes in Trial Design. *Cancer Discov* 2017; **7** (11): 1209-10.
- 28 Mazarella L, Morganti S, Marra A et al. Master protocols in immuno-oncology: do novel drugs deserve novel designs? *Journal ImmunoTher Cancer* 2020; **8**:e000475.
- 29 Seymour L, Bogaerts J, Perrone A et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* 2017; **18**(3): e143–52.
- 30 Champiat S, Lambotte O, Barreau E et al. Management of immune checkpoint blockade dysimmune toxicities: a collaborative position paper. *Ann Oncol* 2016; **27**(4):559-74.
- 31 Nishino M, Jagannathan JP, Ramaiya NH et al. Revised RECIST Guideline Version 1.1: What Oncologists Want to Know and What Radiologists Need to Know. *Am J Roentgenol*. 2010; 195:281-9.
- 32 Eisenhauer EA, Therasse P, Bogaerts J et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; **45**(2):228–47.
- 33 Cortot AB. Evaluation de la réponse à la chimiothérapie, aux thérapies ciblées et à l’immunothérapie. *Rev Mal Respir Actu*. 2015; 7:455-61.
- 34 Eisenhauer EA, Twelves C, Buyse ME. Phase I cancer clinical trials: a practical guide. Oxford; New York: Oxford University Press, 2015.
- 35 U.S. Department of Health and Human Services, Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry; 2018. Available from: <https://www.fda.gov/media/71195/download>. Accessed on 09/22/2020.
- 36 European Medicine Agency. Guideline on the evaluation of anticancer medicinal products in man, 2017. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man-revision-5_en.pdf. Accessed on 09/22/2020.
- 37 Ocana A, Tannock I. F. When are “positive” clinical trials in oncology truly positive? *JNCI* 2001; **103**:16-20.

-
- 38 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):3721–26.
- 39 Alonso A, Bigirimurame T, Burzykowski T et al. Applied surrogate endpoint evaluation methods with SAS and R. Chapman & Hall/ CRC biostatistics series, 2016.
- 40 Savina M, Gourgou S, Italiano A et al. Meta-analyses evaluating surrogate endpoints for overall survival in cancer randomized trials: A critical review. *Crit Rev Oncol Hematol* 2018; **123**:21–41.
- 41 Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989; **8**(4):431–40.
- 42 Buyse M, Molenbergs G, Burzykowski T et al. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2001; **1**:49-67.
- 43 Emura T, Matsui S, and Rondeau V. Survival analysis with correlated endpoints. Joint-Fraily copula models. Springer Nature Singapore Pte Ltd. 2019.
- 44 Rondeau V. Statistical models for recurrent events and death: Application to cancer events. *Mathematical and Computer Modelling* 2010; **52**(7–8): 949-55.
- 45 Zhang J, Pilar MR, Wang X et al. Endpoint surrogacy in oncology Phase 3 randomised controlled trials. *British journal of cancer* 2020; **123**:333–34.
- 46 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):6139-48.
- 47 Coart E, Suci 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**:171-74.
- 48 Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group,. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLoS Med* 2009; **6**(7):e1000097.
- 49 Kaufman HL, Andtbacka RHI, Collichio FA et al. Durable response rate as an endpoint in cancer immunotherapy: insights from oncolytic virus clinical trials. *J Immunother Cancer* 2017; **5**(1):72.
- 50 Giobbie-Hurder A, Gelber RD, Regan MM. Challenges of Guarantee-Time Bias. *J Clin Oncol* 2013; **31**(23): 2963–69.
- 51 Hodi FS, Chesney J, Pavlick AC et al. Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol* 2016; **17**(11):1558-68.
- 52 Kaufman H, Schwartz L, William Jr W et al. Evaluation of classical 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):2245-61.
- 53 Commenges D, Jacqmin-Gadda H. Dynamical biostatistical models. Chapman & Hall. 2015.
- 54 Chudley L, McCann K, Mander A et al. DNA fusion-gene vaccination in patients with prostate cancer induces high-frequency CD8(+) T-cell responses and increases PSA doubling time. *Cancer Immunol Immunother* 2012; **61**:2161–70.
- 55 Liang C, Li L, Fraser CD et al. The treatment patterns, efficacy, and safety of nab(®)-paclitaxel for the treatment of metastatic breast cancer in the United States: results from health insurance claims analysis. *BMC Cancer* 2015; **15**:1019.
- 56 Teng CL, Wang CY, Chen YH, Lin CH, Hwang WL. Optimal sequence of irinotecan and oxaliplatin-based regimens in metastatic colorectal cancer: a population-based observational study. *PLoS One* 2015; **10**:e0135673.

-
- 57 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.
- 58 Branchoux S, Bellera C, Italiano A et al. 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.
- 59 Larkin J, Chiarion-Sileni V, Gonzalez R et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 2015; **373**:23-34.
- 60 Motzer RJ, Tannir NM, McDermott DF et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med* 2018; **378**:1277-90.
- 61 Emura T, Nakatochi M, Murotani K, Rondeau V. A joint frailty-copula model between tumour progression and death for meta-analysis. *Stat Methods Med Res* 2017; **26**(6):2649–66.
- 62 Emura T, Matsui S, Rondeau V. Takeshi Emura. Survival analysis with correlated endpoints joint frailty-copula models. SpringerBriefs in Statistics Springer Nature Singapore 2019.
- 63 Motzer RJ, Escudier B, McDermott DF et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* 2015; **373**(19):1803-13.
- 64 Mulkey F, Theoret MR, Keegan P et al. Comparison of iRECIST versus RECIST V.1.1 in patients treated with an anti-PD-1 or PD-L1 antibody: pooled FDA analysis. *J Immunother Cancer* 2020; **8**(1): e000146.
- 65 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): 288-93.
- 66 Dudani S, Graham J, Wells JC. 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):861-67.
- 67 DeMets DL, Psaty BM, Fleming TR. When Can Intermediate Outcomes Be Used as Surrogate Outcomes? *JAMA* 2020; **323**(12): 1184-85.
- 68 Campbell B A, Scarisbrick J J, Kim Y H. Time to next treatment as a meaningful endpoint for trials of primary cutaneous lymphoma. *Cancers* 2020; **12**(8): 2311.
- 69 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):830-38.
- 70 Davis S, Tappenden P, Cantrell A. A review of studies examining the relationship between progression-free survival and overall survival in advanced or metastatic cancer. Report by the decision support unit.2012. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/PFSOS-Report.FINAL_.06.08.12.pdf Accessed on 09/22/2020.
- 71 Allayous C, Da Meda L, Barthelemy M et al. MelBase, constitution et suivi d'une cohorte nationale de patients atteints de mélanome stade III inopérable ou stade IV avec collection d'une base de données clinico-biologiques. *Ann Dermatol Venereol* 2014 ; **141**(12S):S396.
- 72 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): 2925-34.
- 73 Reck M, Ciuleanu T-E, Cobo Dols M. Nivolumab (NIVO) + ipilimumab (IPI) + 2 cycles of platinum-doublet chemotherapy (chemo) vs 4 cycles chemo as first-line (1L) treatment (tx) for stage IV/recurrent non-small cell lung cancer (NSCLC): CheckMate 9LA. *J Clin Oncol* 2020; **38**(15S):9501.
- 74 Buyse M, Saad ED, Burzykowski T. Assessing Treatment Benefit in Immuno-oncology. *Stat Biosci* 2020. **12**:83–103.

75 Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Stat Med* 1992; **11**:1871-79.

76 Lee KH, Rondeau V, Haneuse S. Accelerated failure time models for semi-competing risks data in the presence of complex censoring. *Biometrics* 2017; **73**(4):1401–12.

77 Royston P, Parmar MK. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med*. 2011; **30**(19):2409-21.

78 Paoletti X, Lewsley L-A, Daniele G et al. Assessment of progression-free survival as a surrogate endpoint of overall survival in first-line treatment of ovarian cancer: a systematic review and meta-analysis. *JAMA Netw Open* 2020; **3**(1):e1918939.