université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR

DE L'UNIVERSITÉ DE BORDEAUX

École doctorale Sociétés, Politique, Santé Publique

Spécialité Santé Publique, Option Biostatistique

Par Quentin Le Coënt

Développement de méthodes pour la validation de critères de substitution : comparaison des approches causales et non causales pour des données de méta-analyses

---

Causal and non causal approaches for surrogate validation: joint modeling, mediation analysis and meta-analyses

---

Sous la direction de : Virginie Rondeau
Co-directrice : Catherine Legrand

Soutenue le 14 décembre 2022

Membres du jury :

M^me^ Proust-Lima Cécile	Dr, INSERM U1219, Bordeaux	Présidente
M. Liu Lei	Pr, Université de Washington, USA	Rapporteur
M. Michiels Stefan	Dr, Université Paris-Saclay, Paris	Rapporteur
M. Bouaziz Olivier	MCF, Université Paris-Cité, Paris	Examinateur
M^me^ Colantuoni Elizabeth	Ass. Pr, Université John-Hopkins, USA	Invitée
M^me^ Rondeau Virginie	Dr, INSERM U1219, Bordeaux	Directrice de thèse
M^me^ Legrand Catherine	Pr, UCLouvain, Belgique	Codirectrice de thèse

Contents

Remerciements **5**

Scientific production **7**

Complementary teaching allocation **8**

Notations **9**

Abbrevations **10**

Résumé substantiel **11**

I Introduction **19**

I.1 Cancer epidemiology . . . 19

I.1.1 Cancer treatments . . . 19

I.2 Clinical trials evaluation . . . 20

I.3 Surrogate endpoints . . . 22

I.3.1 Definitions . . . 22

I.3.2 The need for surrogate endpoints . . . 23

I.4 Statistical evaluation of surrogate endpoints . . . 24

I.5 Objectives of this thesis . . . 27

II Methods **29**

II.1 Validation of surrogate endpoint . . . 30

II.2 The meta-analytic approach . . . 33

II.2.1 Gaussian endpoints . . . 33

II.2.2 Time-to-event endpoints 38
II.2.3 Longitudinal surrogate and time-to-event final endpoint 39
II.2.4 The surrogate threshold effect 41
II.3 MEDIATION ANALYSIS 42
II.3.1 Counterfactual framework 43
II.3.2 Identifiability assumptions 43
II.3.3 Natural direct and indirect effects 48
II.3.4 Application to surrogate endpoint validation 50
II.4 JOINT MODELS IN SURVIVAL ANALYSIS 52
II.4.1 Joint model for a longitudinal surrogate and a time-to-event endpoint 53
II.4.2 Joint model for a time-to-event surrogate 64

III TIME-TO-EVENT SURROGATE VALIDATION USING MEDIATION ANALYSIS AND META-ANALYTIC DATA **67**

III.1 INTRODUCTION 67
III.2 MAIN ARTICLE 68
III.3 ADDTIONAL REMARKS 104

IV VALIDATION OF A LONGITUDINAL MARKER AS A SURROGATE USING MEDIATION ANALYSIS AND JOINT MODELING **105**

IV.1 INTRODUCTION 105
IV.2 MAIN ARTICLE 106
IV.3 ADDITIONAL REMARKS 134

V TUTORIAL FOR SURROGATE ENDPOINT VALIDATION USING JOINT MODELING AND MEDIATION ANALYSIS **135**

V.1 INTRODUCTION 135
V.2 MAIN ARTICLE 136

VI DISCUSSION **177**

VI.1 Joint modeling and mediation analysis 177

VI.2 Perspectives 179

VI.3 Conclusion 180

Bibliography **181**

Appendices **189**

A Proofs **191**

A.1 Mediation analysis formula 191

A.2 Kendall's τ 191

Remerciements

Mes premiers remerciements vont à mes deux directrices de thèse, Virginie Rondeau et Catherine Legrand. Merci d'avoir accepté d'encadrer cette thèse. Vos connaissances ainsi que votre enthousiasme m'ont été précieux durant ces trois années. Merci pour vos conseils, votre aide, vos analyses et vos relectures. J'ai énormément appris grâce à vous. J'espère pouvoir continuer à travailler avec vous à l'avenir.

Je remercie les membres du jury d'avoir accepté d'évaluer cette thèse. I would like to thank Lei Liu for accepting to be part of this jury, whose work on joint modeling and mediation analysis has been of great influence for me. Je remercie également Stefan Michiels d'avoir accepté de faire partie de mon jury, j'ai été très heureux d'avoir pu discuter avec vous de mes travaux de thèse à lors d'une session de poster à Louvain-la-Neuve. Je remercie également Cécile Proust-Lima et Olivier Bouaziz d'avoir accepté d'être examinateurs de ma thèse. I would also like to thank Elizabeth Colantuoni for accepting to be part of this jury as an invited member, I look forward to working with you.

À l'ensemble de l'équipe Biostatistique du BPH je vous remercie pour ces trois années. J'ai beaucoup appris lors de toutes ces réunions SOS. Je tiens à remercier Hélène Jacqmin-Gadda de m'avoir accueilie dans son équipe, Cécile Proust-Lima pour avoir accepté d'être dans mon jury ainsi que pour ces conseils lors des réunions SOS, les doctorants de l'équipe, présents et passés, pour l'ensemble des discussions que j'ai pu avoir avec vous. Enfin plus généralement je tiens à remercier l'ensemble des membres de cette très belle équipe pour ces années passées avec vous.

Je remercie également l'ensemble des administrateurs du MCIA pour m'avoir permis d'utiliser leur cluster de calcul sans lequel mes simulations auraient été sans doutes impossibles à conduire.

Merci à Rodolphe Thiébaut et Alioum Ahmadou pour m'avoir accepté comme doctorant contractuel pour enseigner dans leur module de mathématiques de l'École Sciences et Santé

de l'Université de Bordeaux. Ce fut un réel plaisir pour moi d'avoir l'opportunité de donner ces cours. Je remercie également tous les étudiants que j'ai pu avoir pour leur enthousiasme et surtout leur patience face à des notions de mathématiques pas toujours évidente pour des étudiants en deuxième années de parcours Santé.

Je remercie James Dignam pour m'avoir fournit un très joli jeu de données sur le PSA afin que je puisse y appliquer mon modèle.

Enfin je tiens à remercier ma famille et mes amis pour le soutien continu qu'ils m'ont apportés pendant ces trois années.

Scientific production

Articles

- Quentin Le Coënt, Catherine Legrand and Virginie Rondeau. Time-to-event surrogate endpoint validation using mediation analysis and meta-analytic data, *Biostatistics*, 2022, kxac044.
- Quentin Le Coënt, James Dignam, Catherine Legrand and Virginie Rondeau. Validation of longitudinal marker as a surrogate using mediation analysis and joint modeling: evolution of PSA as a surrogate of the DFS. *In preparation.*
- Quentin Le Coënt, Catherine Legrand and Virginie Rondeau. Tutorial for surrogate endpoint validation using joint modeling and mediation analysis. *In preparation.*

Oral Communications

- Le Coent Q., Rondeau V. and Legrand C. Causal assessment of surrogacy for time-to-event endpoints using meta-analytic data . *8th Channel Network Conference (CNC21)*, April 2021.
- Le Coent Q., Rondeau V. and Legrand C. Causal assessment of surrogacy for time-to-event endpoints using meta-analytic data . *42nd Conference of the International Society for Clinical Biostatistics (ISCB 2021)*, July 18-22, 2021.
- Le Coent Q., Rondeau V. and Legrand C. Validation of longitudinal outcome as a surrogate of a time-to-event endpoint using mediation analysis and meta-analytic data. *31st International Biometric Conference (IBC 2022)*, July 10-15 2022.

Poster communication

- Le Coent Q., Rondeau V. and Legrand C. Time-to-event surrogate endpoint validation using mediation and meta-analytic data. *Workshop "ISBA 30 years" - "Statistics and its role in societal challenges"*, May 12-13 2022.

Complementary teaching allocation

- **Institution**: Insitut de Santé Publique d'Épidémiologie et de Développement (ISPED), Universté de Bordeaux
- **Academic years**: 2019/2020, 2020/2021 and 2021/2022
- **Duration**: 96 hours (32 per academic year)
- **Subjects taught**
  - Introductory topics on undergraduate maths for second-year undergraduate medical students
    - Real analysis: real multivariate functions, partial derivatives, gradient vector
    - Optimzation methods
    - Ordinary differential equations
    - Basics of linear algebra: Matrices, vectors, scalar and operations.
    - Probability : Random variables, distributions, Law of large numbers, Central-Limit theoram
    - Statistics: Introduction to the concepts of estimation, inference and hypothesis testing
  - Introduction to programming through Python

Notations

- $\mathbb{P}(A)$: Probability of an event A
- $\mathbb{P}(A, B)$: Probability of the event $A \cap B$
- $\mathbb{P}(A \mid B)$: Probability of A given B
- $X \perp\!\!\!\perp Y$: X is indendpendent of Y
- $X \perp\!\!\!\perp Y \mid Z$: X is indendpendent of Y given Z
- $\mathbb{E}(X)$: Expectation of a random variable X
- $\mathbb{V}(X)$: Variance of a random variable X
- $\mathrm{Cov}(X, Y)$: Covariance between X and Y
- $f_X(\cdot)$: density function of a random variable X
- $\lambda(\cdot)$: Hazard/Instantaneous risk function
- $S(\cdot)$: Survival function
- $X \sim \mathcal{L}$: The random variable X follows the probability distribution $\mathcal{L}$
- $\mathcal{N}(\mu, \Sigma)$: Normal distribution with mean μ and covariance matrix Σ
- $\hat{\theta}$: Estimator of a parameter θ
- $L(\cdot)$: Likelihood function
- $l(\cdot)$: Log-likelihood function
- $A \propto B$: A is proportional to B
- $A \wedge B$: A and B

Abbrevations

- OS: Overall survival
- PSF: Progression-free survival
- DFS: Disease-free survival
- TTR: Time-to-recurrence
- ATE: Average treatment effect
- TTE: Total treatment effect
- NIE: Natural indirect effect
- NDE: Natural direct effect
- PTE: Proportion of treatment effect
- PSA: Prostate-specific antigen
- JM: Joint modeling
- SUTVA: Stable Unit Treatment Value Assumption
- DAG: Directed Acyclic Graph

Résumé substantiel

1. Introduction

En recherche clinique, le bénéfice clinique des patients se voit de plus en plus amélioré par le continuel développement de nouvelles thérapies. Bien que cette amélioriation soit une bonne chose, celle-ci pose cependant une difficulté du point de vue de la mise en place de nouveaux essais cliniques. Ces essais, et particulièrement en oncologie, utilisent très souvent comme critères de jugement principaux des temps d'évènements. Ces critères sont souvent défini comme la durée entre l'initiation d'un traitement et la survenue d'un évènement clinique majeur, tel le décès ou la progression de la maladie. Un traitement sera d'autant plus efficace qu'il rallonge cette durée. La conduite d'un essai clinique nécessite alors de suivre les patients de l'étude suffisamment longtemps afin d'observer un nombre d'évènement cliniques suffisant pour permettre de détecter statistiquement un effet du traitement. L'amélioration continue des nouveaux traitements conduit donc à devoir mettre en place des essais cliniques sur des durées de plus en plus longue. Cet allongement de la durée nécessaire des essais cliniques ralentit ainsi la mise à disposition des nouveaux traitements pour les patients, et peut même conduire à une infaisibilité de l'essai en lui-même du aux coûts trop important qu'induit un suivi sur le long terme des patients dans l'étude. Afin de pallier à cette difficulté, de plus en plus d'essais utilisent comme critères de jugements principaux des critères intermédiaires, plus rapidement ou plus fréquemment mesurables que les critères finaux. Un exemple se trouve en oncologie, ou dans certains essais sur dans cancers précoces de bon pronostic, le délai entre l'initiation (ou la randomisation) du traitement jusqu'au décès est remplacé par le délais jusqu'à la progression de la maladie. L'utilisation de tel critères intermédières (également appelés critères de *substitution*) présente ainsi un réel attrait afin de pouvoir réduire la durée des essais cliniques. Cependant se pose la question de savoir si un effet significatif du traitement sur le critère intermédiaire peut se généraliser au critère final (qui est le réel critère d'intérêt). Des essais ont été conduits sur des critères intermédiaires montrant une efficacité du traitement sur ceux-ci, avant de se révéler inefficace voir délétère vis-à-vis du critère final.

La validation de tels critères intermédiaires a donc émergé dans la littérature statisique depuis plusieurs années. L'idée de cette validation est de déterminer si un potentiel critère de substitution est intéressant dans le sens ou la mise en évidence d'une efficacité d'un traite-

ment sur celui-ci permettrait de pouvoir en déduire une efficacité du traitement sur le critère final. Ainsi, plusieurs méthodes statistiques ont été proposées, parmis lesquelles deux grandes classes peuvent se dégager.

La première classe est constituée de méthodes basées sur l'étude de l'association statistique entre les deux critères. Parmis celles, l'approche par méta-analyses a été la plus féconde. Cette approche utilise des données issus de plusieurs essais cliniques ayant déjà été réalisés (et pour lesquels on dispose de l'information sur les deux critères) afin de d'étudier cette association à deux niveaux. Au niveau individuel (patient) l'associatio entre les deux critères est déterminée, et au niveau de l'essai clinique l'association entre les effets du traitement sur les deux critères est étudiée. Un potentiel critère de substitution sera validé si l'association est suffisamment forte aux deux niveaux.

La seconde classe se base plutôt sur l'étude de l'effet du traitement sur le critère final et notamment sa décomposition en un effet direct et un effet indirect au travers son effet sur le critère intermédiaire. Les développements principaux de cette approche ont été realisés essentiellement au travers de l'analyse causale de médiation dont l'objectif est de d'identifier et de donné une interpretation causale à cette décomposition. La proportion de l'effet total du traitement sur le critère final opérant via son effet sur le critère de substitution est pris comme mesure pour valider un potentiel critère de substitution qui sera validé si cette proportion est suffisamment importante. Bien qu'intéressante, cette approche a connu moins de développement que celle par méta-analyse. Une de ces limitations est qu'elle est restreinte à l'analyse de données issues d'un seul essai clinique ce qui limite la généralisation de ses conclusions contrairement à une approche basée sur des méta-analyses.

2. Validation de temps d'évènements comme critères de substitution: utilisation de l'analyse de médiation et de méta-analyses

L'un des premiers objectifs de cette thèse était de faire le lien entre les deux principales approches présentées précedemment. Nous nous sommes d'abord intéressés au cas où le critère final et le critère de substitution à valider sont tous les deux des temps d'évènement. Par exemple, le critère final peut être le délais jusqu'au décès et le critère substitution le délai jusqu'à un critère clinique intermédiaire, comme par exemple la récurrence de la maladie ou encore

la survenue de métastases. Le point de départ fût un modèle conjoint ayant été précedemment développé. Ce modèle permettait, via l'introduction d'effets aléaoires, de mener une approche par meta-analyse, et de déterminer les deux niveaux d'associations. Cependant, comme pour de nombreux modèles conjoint, celui-ci faisait l'hypothèse d'une indépendance des critères conditionnellement aux effets aléatoires. Cette indépendance (conditionnelle) a pour conséquence une absence de lien direct entre les critères, et notamment d'un effet du critère de substitution sur le critère final. Par conséquent, du point de vue de la médiation ce modèle faisait l'hypothèse qu'un effet du traitement sur le critère final ne pouvait être médié au travers du critère de subsitution. Nous avons donc étendu ce modèle afin de permettre une potentielle médiation de l'effet du traitement. Cette extension se traduit dans le modèle par une modification du risque de survenue du critère final à la survenue du critère de substitution. À partir de ce modèle, nous avons également défini l'effet direct et indirect du traitement sur le critère final. Ces définitions font intervenir la notion de réponse contrefactuelles, très utilisées dans le domaine de l'inférence causale, leur permettant une interprétation causale. Une étude de simulation a également été conduite afin d'évaluer les performances de l'approche proposée, notamment concernant l'estimation des différents paramètres du modèle conjoint ainsi des mesures de médiations (effet direct et indirect) qui en découlent. Nous avons également appliquer cette approche à deux analyses de données réelles. Une première utilisait les données issues d'une méta-analyse en cancer gastrique portant sur l'étude du temps jusqu'à récurrence du cancer comme critère de substitution du temps jusqu'au décès (critère final). Cette application a montré que ce critère de subsitution médiait plus de 80% de l'effet traitement sur le critère final, ce qui en fait ainsi un critère intéressant. La seconde application portait sur l'étude du temps jusqu'à progression du cancer comme critère de substitution du temps jusqu'au décès dans le contexte de cancers colorectal traités en second-ligne. Nous nous sommes basés sur les données issues d'un essai clinique de phase 3 multi-centrique, les centres ayant fait l'object d'essais vis-à-vis de la méta-analyse. Cette analyse a révélé l'absence d'effet du traitement sur le critère final, rendant ainsi inutile l'analyse et la validation de critère de substitution dans ce contexte. Cependant, il peut tout de même être intéressant d'édutier la décomposition de l'effet du traitement en un effet direct et indirect pour voir si l'absence d'effet traitement s'explique par une compensation de l'effet direct et indirect, ce qui permettrait de mieux comprendre les mécanismes d'action du traitement. Ce travail a été accepté

pour publication dans la revue *Biostatistics* en 2022.

3. Validation d'un biomarqueur longitudinal comme critère de substitution d'un temps d'évènement par analyse de médiation et modélisation conjointe

Dans un second temps de la thèse, nous nous sommes intéressés au cas où le critère de substitution n'est plus un temps d'évènement mais un biomarqueur longitudinal, c'est à dire constitué de mesures répétées au cours du temps tel par exemple les mesure répétées du taux de PSA (prostate-specific antigen) chez des patients atteints d'un cancer de la prostate, le critère final d'intérêt étant toujours un temps d'évènement. La prise en compte de l'association entre les deux critères ainsi que la nature bruitée inhérente à beaucoup de biomarqueurs longitudinaux nécessite une modélisation. Comme pour la première partie, nous avons développé un modèle conjoint spécifique à ce contexte. Ce modèle inclut des effets aléatoires à la fois au niveau individuel afin de prendre en compte la dynamique d'évolution du biomarqueur au cours du temps ainsi qu'au niveau essai pour tenir compte de l'hétérogénéité de l'effet du traitement sur les deux critères entre les essais. L'association entre les deux critères est prise en compte dans le modèle grâce à une fonction de lien dans la partie du modèle portant sur la modélisation du risque de survenue du temps d'évènement. Cette fonction dépend de la trajectoire du biomarqueur. Cette trajectoire est la trajectoire moyenne, non bruitée, prédite par le modèle pour le critère longitudinal. Plusieurs types de fonctions de lien ont été proposées dans littérature des modèles conjoints. La valeur courante du biomarqueur à chaque instant est l'une des fonctions de lien les plus utilisées, et fait l'hypothèse que la composante du biomarqueur influançant le risque de survenue du critère final à un instant donné est la valeur du biomarqueur. Une autre possibilité est d'utiliser la pente du biomarqueur, auquel cas ce n'est non plus la valeur de celui-ci qui influence le risque d'évènement mais son taux d'accroissement. Enfin une troisième possibilité fréquemment rencontrée est une fonction de lien dite "à effets aléatoires partagés" où se sont les effets aléatoires au niveau individuels qui sont inclus dans la fonction de risque du critère final. Comme pour le cas deux temps d'évènements, ici une médiation de l'effet du traitement est possible seulement si la quantité dont dépend la fonction de lien dépend du traitement. Dans le cas des effets aléatoires partagés, ceux-ci sont indépen-

dants du traitement et donc ne permettent pas d'étudier un effet médier du traitement. De même pour le cas de la pente, afin "d'autoriser" une potentielle médiation, celle-ci doit pouvoir dépendre du traitement. Nous nous sommes ainsi intéressés essentiellement aux deux premières fonctions de lien (valeur courante et pente). À partir du modèle défini, l'utilisation du cadre contrefactuel permet, modulo certaines hypothèses d'identifiabilité, de définir et calculer la décomposition de l'effet total du traitement sur le critère final en un effet direct et un indirect au travers du biomarqueur. La proportion d'effet total égale à l'effet indirect est ensuite prise comme mesure d'évaluation pour le critère longitudinal. Une étude de simulation a été mise en place pour évaluer les performances de l'approche proposée. Cette approche a été appliqué à des données issues d'un essai randomisé multicentrique. Les centres de soins ont été pris comme unité de clustering afin d'étudier une hétérogénéité de l'effet du traitement. L'objectif initial de cet essai était d'évaluer le bénéfice d'une poursuite pendant 24 mois d'une hormonothérapie (visant à réduire le taux d'androgène) après radiothérapie, le bras contrôle ne recevait cette hormonothérapie que pendant la durée de la radiothérapie. Nous nous sommes intéressés à l'étude de l'évolution de taux de PSA chez ces patients comme critère de substitution de la survie sans maladie. Nous avons étudié le cas où la valeur courante du PSA était utilisé comme fontion de lien ou bien la pente du PSA. Le premier cas a montré que la valeur courante du PSA médiait un part significative de l'effet du traitement sur le risque de survenue de l'évènement d'intérêt avec un accroissement de cette proportion au cours du temps (d'environ 40% après 1 an jusqu'à environ 75% à 5 ans) tandis que la pente n'en médiait qu'une infime partie (moins de 3%). Ansi, l'évolution de la valeur du taux de PSA semble être un critère de subsitution intéressant pour la survie sans maladie dans cette contexte. De plus, l'étude de la variance des effets aléatoires représentants l'hétérogénéité de l'effet du traitement à la fois sur le temps d'évènement et sur le biomarqueur a montré que cette hétérogénéité était très faible (variances proche de 0). Ceci n'est pas surprenant étant donné que les centres de soins utilisés sur lesquels cette hétérogénéité a été étudiée faisait tous partis du même essai multicentrique, et suivaient donc tous le même protocole de soin, limitant la présence d'hétérogénéité. Ce travail est en cours de finalisation pour soumission dans une revue de statistique appliquée à la médecine.

4. Implémentation des méthodes développées dans un package statistique

Une majeure partie de ce travail de thèse a concerné le développement et l'implémentation de l'approche proposée dans un package statistique facilitant. Cette implémentation a été intégrée au package `frailtypack` du langage statistique R. Ces approches ont été dans deux fonctions du package, une pour le cas ou le critère de substitution et le critère final sont deux temps d'évènements et une autre pour un critère de subsitution continu et un critère final qui est un temps d'évènement. L'estimation des modèles conjoints développés ainsi que les mesures de substitution qui en découlent (et notamment la proportion d'effet du traitement) peuvent nécessiter des temps de calcul assez important. Afin de réduire ceux-ci, les méthodes ont été implémentées en `Fortran` et peuvent être appelées directement depuis le package R par interfaçage. De plus, ces méthodes peuvent supporter le calcul parallel, et plus particulièrement le multithreading grâce à `OpenMP`. Afin de faciliter l'utisitation des méthodes proposées, nous avons rédiger un article reprenant celles-ci, et illustrant leur utilisation au travers de trois applications sur des jeux de données issues de méta-analyses ou d'essais multicentrique. Une première illustration porte sur des données issues de quatres essais randomisés pour le cancer ovarien avancé et s'intéresse à l'étude du temps jusqu'à progression du cancer comme critère de substitution du temps jusqu'au décès via des critères "classiques" de méta-analyses, à savoir association au niveau individuel et au niveau essai. Une seconde porte sur l'application au cancer gastrique de la première partie de cette thèse et sur l'étude du temps jusqu'à récurrence du cancer comme critère de substitution du temps jusqu'au décès au travers de mesures liées à l'analyse de médiation à savoir la proportion d'effet du traitement. Une troisième illustration s'intéresse à un critère longitudinal dans le contexte du cancer colorectal. Le critère de substitution est l'évolution au cours du temps de la taille tumorale et le critère final est le temps jusqu'au décès. Comme pour la second illustration, l'approche présentée se base sur l'estimation de la proportion d'effet du traitement sur le critère final au travers du critère de substitution. Cet article est en cours de finalisation pour soumission dans une revue dédiée à l'implémentation logicielle de méthodes statistiques.

5. Discussion

Dans cette thèse nous avons proposé des développements pour la validation statistique de critères de substitutions, en s'appuyant sur l'analyse causale de médiation afin de déterminer la proportion d'effet du traitement sur le critère final opérant au travers du critère de substitution et également sur l'utilisation de données issues de méta-analyse. Deux modèles conjoints ont été developpés selon la nature (temps d'évènement ou biomarqueur longitudinal) du critère de substitution desquels des mesures liées à l'analyse de médiation ont été définies. Dans ce contexte, contrairement aux précédents développements de l'analyse de médiation qui se limitaient à des analyses sur des données issues d'un seul essai clinique, l'apport de données de méta-analyses permet une plus grande généralisabilité des résultats obtenus.

Les travaux de cette thèse ont été appliqués à des données issues d'étude en oncologie, lesquelles se basent très souvent sur des critères de jugements finaux de type temps d'évènement. Cependant les méthodes proposées peuvent être appliquées à d'autres domaines de la recherche clinique dans lesquels les mesures d'efficacité d'un traitement correspondent à délais jusqu'à la survenue d'un évènement clinique d'intérêt.

Plusieurs perspectives se dégagent de cette thèse. Premièrement, l'extension des méthodes proposées à des cas où le critère final ne correspond plus forcément à un temps d'évènement est envisagé. Par exemple, les critères binaires (succès/echec) sont également très utilisés dans certains champs de la recherche, ainsi que des critères continus. L'extension des méthodes proposées à ces différents types de critères finaux nécessite la définition de modèles conjoints appropriés, mais l'esprit globale de la validation resterait essentiellement le même; la proportion de l'effet du traitement sur le critère final sera déterminer à partir de ces nouveaux modèles.

Dans cette thèse nous avons chercher à rapprocher les deux grandes approches proposées dans littérature pour la validation de critère de substitution: l'approche par méta-analyse et l'approche par analyse de médiation. Ces approches ne sont pas exclusives mais complémentaires. La mesure d'association au niveau individuel couramment utilisée dans l'approche méta-analytique est complémentaire de la proportion de l'effet du traitement de l'approche par médiation. Dans le cas où un effet indirect du traitement sur le critère final existe cette mesure d'association nécessite d'être définie en conséquence afin de prendre cette dépendence

en compte. Dans ces travaux nous avons proposés de nouvelles définitions pour ces mesures d'associations, notamment pour le τ de Kendall. Cependant, l'introduction de cette dépendence entre les deux critères complexifie le calcul de ce τ dans le cas d'une médiation. Un travail en cours vise à l'implementation effective de cette mesure "corrigée".

En conclusion, les travaux de cette thèse ont permis le développements de nouvelles approches pour la validation de critères de substitution dans le cas où le critère final est un temps d'évènement, basées sur la modélisation conjointe, l'analyse de médiation et les méta-analyses. Ces approches ont été implémentées dans un package statistique afin d'en faciliter leur utilisation.

I

Introduction

I.1 Cancer epidemiology 19
I.1.1 Cancer treatments 19
I.2 Clinical trials evaluation 20
I.3 Surrogate endpoints 22
I.3.1 Definitions 22
I.3.2 The need for surrogate endpoints 23
I.4 Statistical evaluation of surrogate endpoints 24
I.5 Objectives of this thesis 27

I.1. | Cancer epidemiology

Cancer has been reported is one of the leading cause of death worldwide with more than 19 millions of new cancers been diagnosed in 2020 and nearly 10 million cancer-related deaths occuring during the same year (Worl Health Organization, 2022; Global Cancer Observatory, 2022). Cancer research focuses primarily on identifying risks factors, prevention, developing new therapies and improving the quality of life of cancer patients.

I.1.1. Cancer treatments

Treatments allocated to a cancer patient depend on the cancer type, its stage and the therapeutic background of the patient. It can also depend on the charateristics of the patient, such as his/her genetic profile. Common therapies are chemotherapy, radiotherapy, hormone therapy. Sometimes these therapies can be combined, for example adjuvant therapy can be

assigned to a patient after the main treatment to prevent the resurgence of the cancer. Neoadjuvant therapy can also be used before the main treatment (mainly surgery or radiotherapy) to reduce the tumor size and ease the course of the main therapy. More recently a new type of therapy, immunotherapy, has emerged. Its working principle is to help the immune system of a patient better fight cancer cells. The first immunotherapy for cancer was approved in 2011. Since then, several types of immunotherapies have been developed (immune checkpoint inhibitors, T-cell transfer therapy, monoclonal antibodies) for different cancers. One advantage of immunotherapies is to have less side effects than other therapies such as chemotherapy. Immunotherapy is currently one of the hottest area in cancer therapy research. In this continuing search for cancer therapies, the evaluation of their efficacy through clinical trials remains a essential part of the cancer research.

I.2. | Clinical trials evaluation

Over the last five years more than 33000 clinical studies per year have been registered on https://www.clinicaltrials.gov/ with a trend indicating an increase of the number of registered study each year (U.S. National Library of Medicine, 2022).

Most of these studies are clinical trials, i.e. interventional studies evaluating the benefit of an intervention on patients included in a trial. To properly quantify this "benefit", the efficacy of the intervention must be assessed on a clinical endpoint which is collected for each patient during or at the end of the study. A statistical analysis of the endpoint distribution between the different interventions arms can then be used to assess the efficacy of the intervention in the patient population.

The choice of the endpoint(s) is thus crucial in conducting a clinical trial. This choice depends strongly on the disease being studied. In oncology, several endpoints have been defined as the duration between the intervention assignment and the occurrence of a clinical event of interest (Delgado and Guddati, 2021). The gold standard is the overall survival (OS), defined as the duration between treatment assignment and death. A treatment that would show to substantially increase this duration will then be deemed efficient. Other comon endpoints include the progression-free survival (PFS), the time between treatment assignment and disease progression or death whichever occurred first.

With the ongoing development of new and more efficient therapies, these durations may,

in some type of cancers, be long. This is of course a good news for the patient as it improve their duration of survival, but from a clinical trial standpoint it creates a difficulty. In order to gather sufficient information to statistically exhibit a difference between the treatment arms in efficacy regarding the final endpoint, we need to observe enough events. This may require having to conduct the trial for a long period of time as most of the patients may experience the event long after being included in the trial.

For instance, the number of subject required in a randomized 1:1 trial with a log-rank test and assuming proportional hazards is proportional to the inverse of the proportion of observed events in the trial (Schoenfeld, 1983):

$$N \propto \frac{1}{d}.$$

For a trial with a fixed follow-up, d is decreasing as the survival probability increases. This result in N being larger when the survival increases.

For example, in early breast cancer, the 5-years survival is over 98% (Cancer Research UK, 2020), meaning that only 2% of the women will die before 5 years. As a result, if one has to conduct a new clinical trial on early breast cancer using OS as primary endpoint, one would have to either include a very large number of patients or conduct the trial over a very long period of time in order to observe enough events. In another word, it involves increasing the duration of the trial and/or its cost. Increasing the duration of the trial slows down the treatment development process and accesssibility while increasing the cost simply may make the trial unfeasible.

Motivated by the increasing need to shorten the duration of a treatment evaluation and the increasing length of trials due to availability of more efficient therapies, there has been growing reasearch interest in clinical research to discover and validate surrogate endpoints. A surrogate endpoint is an intermediate endpoint measured earlier than the final endpoint on which the treatment efficacy can be evaluated and should be extendable to the final endpoint. Hence, a new clinical trial could be conducted solely on the validated surrogate and with a reduced follow-up time and cost.

I.3. | SURROGATE ENDPOINTS

I.3.1. Definitions

The Biomarkers Definitions Working Group gives the following definition for a clinical endpoint: (BIOMARKERS DEFINITIONS WORKING GROUP, 2001):

> **Clinical endpoint**
> A characteristic or variable that reflects how a patient feels, functions, or survives.

A surrogate endpoint is defined as:

> **Surrogate endpoint**
> A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.

This definition uses the term "biomarker", which is defined as

> **Biological marker (biomarker)**
> A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Following these definitions, surrogate endpoints are a subset of biomarkers that can be used in substitution of clinical endpoints. The Food and Drug Administration (FDA) gives the following definition of a surrogate endpoint (U.S. FOOD AND DRUG ADMINISTRATION, 2018):

> **Surrogate endpoint (FDA)**
> A surrogate endpoint is a clinical trial endpoint used as a substitute for a direct measure of how a patient feels, functions, or survives. A surrogate endpoint does not measure the clinical benefit of primary interest in and of itself, but rather is expected to predict that clinical benefit.

This definition does not label a surrogate as a biomarker and is more general than the one given by the Biomarkers Working Group. Hence a surrogate endpoint is not necessarily a biomaker measure but can also be a clinical outcome. For example, in oncology some clinical outcomes have been shown to be surrogates of the overall survival: disease-free survival (DFS) for stage 2 or 3 colon cancer (Buyse, Burzykowski, Michiels, et al., 2008) or for operable and locally advanced lung cancer (Mauguen et al., 2013); and progression-free survival (PFS) for advanced/metastatic colorectal cancer (Giessen et al., 2013; Buyse, Burzykowski, Carroll, et al., 2007) or locally advanced head and neck cancer (Michiels et al., 2009). It is important to note that the use of a surrogate takes place in a context where the treatment, the target population, and the disease characteristics have to be properly defined. A surrogate may be relevant for a particular class of therapies (but not for other), as well as for a particular stage of a disease. In oncology, some surrogate endpoints have been validated. In patients with non-small cell lung cancer, DFS has been shown to be a good surrogate endpoint for OS in the context of adjuvant chemotherapy while PFS was validated as a surrogate endpoint for OS for patients with locally advanced lung cancers treated with chemotherapy and radiotherapy (Mauguen et al., 2013). In colon cancer both DFS and time-to-recurrence have been shown to be good surrogate endpoint of OS (Sargent et al., 2005; Ghosh et al., 2012). In locally advanced head and neck cancer, DFS, PFS and duration of locoregional control have all been shown to be valid surrogate of OS (Michiels et al., 2009).

I.3.2. The need for surrogate endpoints

There is an increasing need for the use of surrogate endpoints. It has been reported that in oncology the proportion of trials using clinical outcomes different from the OS as primary endpoint has increased from 26% to 43% between 2004 and 2009 (Kemp and Prasad, 2017). As a result the share of drug approvals by the FDA based on clinical trials using endpoints other than OS increased from 23% between 2003-2007 to 67% betwen 2008 and 2012 (Kim and Prasad, 2015; Sridhara et al., 2010).

However, trials using different outcomes as primary endpoints may approve a treatment that shows significant benefit on the endpoint but is ineffective or even deleterious on OS; the gold-standard endpoint. An illustration of this is the validation of the bevacizumab for HER-2 negative metastatic breast cancer by the FDA in February 2008 on the basis of a clinical trial

showing a significant improvement of 5.9 months in median PFS for bevacizumab + paclitaxel versus paclitaxel alone (MILLER et al., 2007). However two additional trials showed that addition of bevacizumad had no significant benefit on OS (MILES et al., 2010; ROBERT et al., 2011) and little benefit on the PFS. The FDA then proposed to remove the indication of bevacizumab for metastatic breast cancer. Interestingly, the manufacturer argued that PFS is a direct rather than a surrogate measure of a clinical benefit in this setting while the Oncologic Drugs Advisory Committee advised for the removal of this drug for metastatic breast cancer (CARPENTER et al., 2011).

Moreover, the ongoing development of new therapies implies a constant evolution of the surrogate endpoints to be used. Surrogates that may be valid for treatment classes such as targeted hormone therapies or chemotherapies may not be valid anymore for new therapies whose mechanisms of action are different such as immunotherapy . This motivates the need for both i) the investigation of new surrogate endpoints and ii) the strengthening of their validation process.

I.4. | STATISTICAL EVALUATION OF SURROGATE ENDPOINTS

Prior to any use in a new clinical trial, a surrogate endpoint must have been carefully validated. Indeed, a surrogate is an endpoint from which the treatment effect on the final endpoint can be inferred based on the observation of its effect on surrogate. A statistical validation is therefore required to show that a potential surrogate can be used.

The process of obtaining a valid and interesting surrogates can be broadly divided in two stages:

i) Ascertainment of biological mechanism of putative surrogates;

ii) Their statistical validation.

The first stage requires a strong rationale of the surrogate biological or clinical link regarding the final endpoint and the treatment it could be used for, based on strong clinical and biological knowledge. The second stage, which is the aim of this thesis, focuses on the statistical validation of potential surrogates. This stage requires that both the surrogate and the final endpoint have previously been investigated together in one or several clinical trials. This can be the case for a trial using the final endpoint as a primary endpoint in which the potential

surrogate is measured as a secondary endpoint. The purpose of a statistical validation of a surrogate endpoint is to be able to infer of the treatment effect on the final endpoint based solely on the observation of its effect on the surrogate. The first formal proposal of a methodology for validating a surrogate endpoint is due to Prentice (Prentice, 1989). In his paper the author defined a surrogate as "a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint". If we denote by S the surrogate, T the final endpoint and Z the treatment assignment, this translates into four criteria:

i) S is associated with Z;

ii) T is associated with Z;

iii) T is associated with S

iv) T is independent of Z conditionnaly on S

The last criterion is the most important one, stating that the surrogate must captures all of the treatment effect on the final endpoint. However, some issues have been raised regarding this method. Freedman et al., 1992 noted that iv) can lead to "a weak form of validation" in that it can reject a poor surrogate but can hardly validate a good surrogate. Moreover, criterion iv) requires the adjustment on a posttreatment variable (the surrogate) and does not takes into account potential common causes (confounders) of the final endpoint and surrogate. The issue raised by raw adjustment on a posttreatment variable has been largely investigated, and not only in the field of surrogate validation, (Rosenbaum, 1984; J. Robins, 1986; Frangakis and Rubin, 2002). Broadly speaking, when confounders of S and T are not properly accounted for, conditioning on S induces an association between Z and theses confounders, and thereby between Z and T. Hence, in that situation criterion iv) may not hold even if S perfectly mediates the treatment effect on T. Both these issues lead to the development of other approaches, which can be divided into two classes (Joffe and Greene, 2009):

- Methods based on "associations"
- Methods based on "effects"

In their paper, Joffe and Green put Prentice's approach in the "effect" paradigm.

The main approach of the "association" paradigm is the meta-analytic approach (Buyse, Molenberghs, Burzykowski, et al., 2000) whose purpose is to use data from several trials to derive associations between the surrogate and the final endpoint at both the invidual (patient) level and trial level. The individual level association is related to the patient follow-up and prediction of his future outcomes while, the trial-level association relates to the prediction of the treatment effect on the final endpoint based on its effect on the surrogate in a new trial. The meta-analytic approach has been extensively discussed and is the most developed and recognized approach for surrogate validation. This approach has been developed for final and surrogate endpoints of different natures (binary, time-to-event and so on) (Burzykowski, Buyse, and Molenberghs, 2005; Michiels et al., 2009). A review of these different implementations has been published (Buyse, Molenberghs, Paoletti, et al., 2016). To derive a trial-level association measure, two-stages approaches were developed in wich a bivariate mixed model is estimated at the second stage based on the treatment effects estimated at the first stage. An adjusment on the error in the estimations at the first stage can be used at the second stage. Numerical issues were often reported for the two-stages approach (Shi et al., 2011; Renfro et al., 2012; Rotolo et al., 2019). To overcome these numerical issues methods based on a single stage using joint modeling were developed. In particular, Sofeu et al. (2019) proposed a joint modeling approach using frailty models when both the surrogate and the final endpoint are time-to-event.

In the validation through "effects" paradigm, the main objective is often to derive and distinguish the treatment effect on the final endpoint due to its effect on the surrogate and its effect independent of the surrogate. This naturally leads to the idea of a "mediated" effect in which a good surrogate is a variable for which most of the treatment on the final endpoint goes through the surrogate. Criterion iv) of Prentice is in line with this idea that all the treatment effect must be captured by the surrogate, but did not take carefully into account the very specific nature of the surrogate as a posttreatment variable that may share common causes with the final endpoint.

Proper developments of the mediation analysis have been made, relying on the causal inference framework and specifically on the use of potential outcomes to derive a causal definition of the indirect effect of the treatment on the surrogate and a direct effect independent of it

(Rubin, 1974; J. M. Robins and Greenland, 1992; Pearl, 2022) . The proportion of effect that corresponds to the indirect effect can then be used as a surrogacy measure.

One limitation of the mediation approach is that available methods are often limited to single-trial data and do not take into account the potential informative heterogeneity that come from meta-alyses or multi-centric data.

I.5. | Objectives of this thesis

The objective of this thesis is to bind the two main approaches for surrogate validation, the meta-analytic approach and the mediation analysis approach, into a single one using joint modeling. The use of a joint model overcomes both the computational issues of the two-stages approach of the meta-analytic approach and the limitation to single-trial data of the mediation analysis. The developments presented in this thesis focus mainly on applications in oncology which rely heavily upon time-to-event but other fields of clinical research where such endpoints are used are also in their scope. The proposed methods were implemented in the R package `frailtypack` in order to make these developments more easily accessible and usable by other statisticians or clinicians.

A first part of this thesis focuses on the specific case where both the surrogate endpoint and the final endpoint are time-to-event. Sofeu et al. (2019) proposed a joint frailty model for meta-analytic data and two event times. However, their proposed model made the assumption of independence between the surrogate and the final endpoint. Therefore, from a mediation analysis perspective, no indirect treatment effect on the final endpoint through the surrogate was possible. To objective was to extend this model in order to interpret it in term of causal mediation analysis. We proposed an extension of this model in which a direct effect of the surrogate on the final endpoint is introduced. That way the assumption of conditional independence does not hold anymore and an indirect treatment effect is enabled. This direct effect translates as a modification of the occurrence risk of the final endpoint from time occurrence of the surrogate. The magnitude of this effect was taken into account through a function depending on the surrogate occurrence time. From the definition of the model and using appropriate identifiability assumptions, we can derive and compute the indirect treatment effect on the final endpoint through the surrogate as well as a direct effect independent of it, which carry a causal interpretation. A strength of this approach is that the use of random effects in

the joint frailty model helps to capture potential unmeasured confounding between the surrogate and the final endpoint which allows to make less stringent identifiablity assumptions. A simulation study was conducted to investigate the performance of the proposed approach regarding the estimation of i) the joint model parameters and ii) the related direct and indirect effects. Then we applied this method to a meta-analysis on gastric cancer investigating the time-to-relapse as a surrogate of the overall survival.

In a second part of this thesis we extended the approach developed in the first part to the case where the final endpoint is a time-to-event but the surrogate is a longitudinal continuous biomarker. A joint model was proposed by RENARD et al. (2003) but was limited to the case of shared-random effects. As for the case with two event times, here the shared random effects joint models assume an conditional independence between the longitudinal biomarker process and the time-to-event. Our objective was to extend this model to other types of association between the two endpoints, including the "current level" association in which the association between the two endpoints is made through including the value of the longitudinal biomarker in the risk function of the final endpoint, or the "current slope" association. These associations enables an indirect effect of the treatment on the final endpoint through the surrogate.

Finally, a third objective was to implement to proposed approaches in a statistical package to facilitate both its diffusion and utilisation.

The following of this thesis is organised as follows: Chapter II introduces the notations and main methods for surrogate validation, mediation analysis and joint modeling. Chapters III, IV and V present the articles related to the three objectives presented above. Finally, Chapter VI gives a general discussion regarding the proposed work, some perspectives and a conclusion.

II

Methods

II.1 Validation of surrogate endpoint 30

II.2 The meta-analytic approach 33

II.2.1 Gaussian endpoints 33

II.2.2 Time-to-event endpoints 38

II.2.3 Longitudinal surrogate and time-to-event final endpoint 39

II.2.4 The surrogate threshold effect 41

II.3 Mediation analysis 42

II.3.1 Counterfactual framework 43

II.3.2 Identifiability assumptions 43

II.3.3 Natural direct and indirect effects 48

II.3.4 Application to surrogate endpoint validation 50

II.4 Joint Models in survival analysis 52

II.4.1 Joint model for a longitudinal surrogate and a time-to-event endpoint 53

II.4.2 Joint model for a time-to-event surrogate 64

In this chapter we first introduce the main approaches that were proposed for validating a potential surrogate endpoint. Joffe and Greene (2009) separated these approaches within two main categories: approaches based on "associations" or "effects". In this chapter both categories will be presented through their main developments: the meta-analytic approach for validation through associations and the mediation analysis approach for the validation through effects.

In the first section we provide a brief historical summary of the early developments of surrogate validation and then give in the second section an introduction to the meta-analytic

approach which is the most popular approach.

The third section is dedicated to an introduction to the mediation analysis and the counterfactual framework in which most of the current developments are made. We also present applications of mediation analysis for surrogate endpoint validation. Finally in the fourth section we present the conventional joint modeling in survival analysis and the two new joints models. The first model considers a longitidunal biomarker (the surrogate of interest) and a time-to-event endpoint designed while the second one is joint frailty model for two event times (in which one of the time-to-event will be the surrogate). These two new models allow to take into account meta-analytic data and will be proved to be a useful tool for surrogate validation. We indeed will link each model to both type of surrogacy evaluation either by association or by indirect effects and provide the related surrogacy measures.

In the following we denote by S the surrogate, T the true (final) endpoint and Z the treatment.

II.1. | Validation of surrogate endpoint

One of the first formal attempt to tackle the issue of validating a surrogate variable was undertaken in the pioneering work of Prentice (Prentice, 1989). In this work, Prentice defined a surrogate as a variable S for which three criteria are verified:

a) S is correlated with T;

b) S is affected by the treatment Z;

c) $T \perp\!\!\!\perp Z \mid S$.

The last criterion states than in a regression model on T adjusted on both Z and S, the test of the null hypothesis of the coefficient of Z being equal to zero should not be rejected, while it should be rejected on a model adjusted only on Z. The motivation behind this criterion is that even if the treatment is associated with T, S must capture all the effect of Z on T. This corresponds to the situation depicted in Figure II.1(a). In this figure, the surrogate will verify Prentice's criterion c) if the arrow from the treatment towards the final endpoint does not exist.

Two issues related to criterion c) have been raised. First Freedman et al. (1992) noted that this criterion implies that S is a surrogate if the null hypothesis of no treatment effect on a

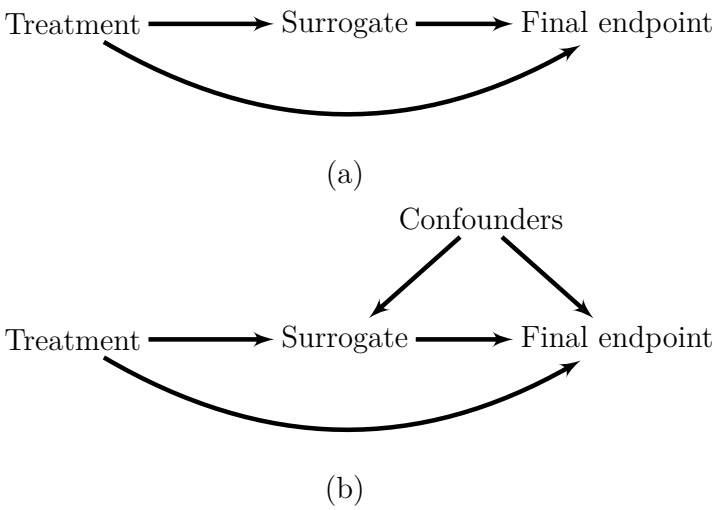


Figure II.1.: DAG for surrogate endpoints. (a) Following Prentice criteria and (b) in a more comprehensive representation.

model adjusted on S and Z is true. However proving a null hypothesis is difficult and the only thing that can be said if the null hypothesis is not rejected is that "the data are not inconsistent with this hypothesis", which is a "weak form of validation". Therefore, this criterion could be useful to reject a surrogate (by rejecting the null hypothesis) but not to validate an interesting one. To overcome this limitation, Freedman et al. (1992) defined the "proportion of treatment explained" defined as,

$$\text{PTE} = 1 - \frac{\beta_S}{\beta}, \tag{II.1}$$

where β_S is the regression coefficient of the treatment effect on T in a model adjusted on S and β the regression coefficient in a model not adjusted on S.

However, both Prentice's and Freedman's criterion assume the situation represented in Figure II.1(a), that is, S and T do not share common causes (confounders). A more plausible situation is given in Figure II.1(b) in which there is some confounding between S and T. In such a case, adjustment on a posttreatment variable (S) without appropriately taking into account confouding between S and T will lead to a biased estimation of the treatment effect (Rosenbaum, 1984). In that case it is not true in general that $T \perp\!\!\!\perp Z \mid S$ even if S perfectly mediates the treatment effect on T. Buyse and Molenberghs (1998) argued that for a surrogate to be useful it should be possible to use it to predict the treatment effect on the final endpoint. Following this idea, they introduced the "relative effect" (RE) defined as the ratio of

the treatment effect on the final endpoint over the treatment effect on the surrogate,

$$\mathrm{RE} = \frac{\beta}{\alpha},$$

where α is the estimated treatment effects on S. If $\mathrm{RE} = 1$ then both effects are equal and yield good surrogacy. If RE were to be known then it can be used to predict β based on an estimation of γ. In addition to the RE, BUYSE and MOLENBERGHS (1998) also proposed another measure which they called "adjusted association", noted ρ_Z, which corresponds to the individual level association between S and T. In the case where both S and T are jointly gaussian, i.e. for a subject i,

$$\begin{aligned} S_i &= \mu_S + \alpha Z_i + \varepsilon_{S_i}; \\ T_i &= \mu_T + \beta Z_i + \varepsilon_{T_i}, \end{aligned}$$

with ε_{S_i} and ε_{T_i} are correlated error terms such that,

$$\begin{pmatrix} \varepsilon_{S_i} \\ \varepsilon_{T_i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_S^2 & \sigma_{ST} \\ \sigma_{ST} & \sigma_T^2 \end{pmatrix} \right),$$

then

$$\rho_Z = \frac{\sigma_{ST}}{\sigma_S \sigma_T}.$$

Just like the surrogate will be perfectly validated at the population level (association between the two treatment effects) if $\mathrm{RE} = 1$, the surrogate will be perfectly validated at the individual level if $\rho_Z = 1$. Moreover, in the case of gaussian endpoints, the authors showed the following relation between their proposed measures and PTE,

$$\mathrm{PTE} = \left(\frac{\sigma_T}{\sigma_S}\right)^{\frac{1}{2}} \frac{\rho_Z}{\mathrm{RE}},$$

and noted that PTE is a mixture of fixed effect through RE and random components through ρ_Z which makes its interpretation difficult. They noted even that in order to achieve suffient accurate estimation of RE for it to be of pratical use, the sample size should be large and based on a large-scale clinical trial. Such trials are not always available and to overcome this, the

use of data from several trials seems natural. Therefore, the use of meta-analyses has since then been widely developed and is currently the most used approach for validating a surrogate endpoint.

II.2. | The meta-analytic approach

The core idea of the meta-analytic approach is to validate a surrogate endpoint at two levels: the individual and the trial-level (Buyse, Molenberghs, Burzykowski, et al., 2000; Daniels and Hughes, 1997). In the context of meta-analyses, let K be the number of trials, n_i the number of subjects in the ith trial and T_{ij}, S_{ij} and Z_{ij} the final endpoint, surrogate and assigned treatment respectively of the jth patient of the ith trial.

II.2.1. Gaussian endpoints

When S_{ij} and T_{ij} are gaussian variables, Buyse, Molenberghs, Burzykowski, et al. (2000) proposed two methods, one based on a two-stage fixed effects model and another based on mixed models. For the two-stage approach, the models for S and T are defined as,

$$\begin{aligned} S_{ij} \mid Z_{ij} &= \mu_{S_i} + \alpha_i Z_{ij} + \varepsilon_{S_{ij}}; \\ T_{ij} \mid Z_{ij} &= \mu_{T_i} + \beta_i Z_{ij} + \varepsilon_{T_{ij}}. \end{aligned} \tag{II.2}$$

At the first stage this model is estimated, where (μ_{S_i}, μ_{T_i}) are trial-specific fixed intercepts, (α_i, β_i) are trial-specific fixed treatment effects and $(\varepsilon_{S_{ij}}, \varepsilon_{T_{ij}})$ are correlated error terms,

$$\begin{pmatrix} \varepsilon_{S_{ij}} \\ \varepsilon_{T_{ij}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} , \begin{pmatrix} \sigma_{SS}^2 & \sigma_{ST} \\ \sigma_{ST} & \sigma_{TT}^2 \end{pmatrix} \right).$$

Then at a second stage, the model

$$\begin{cases} \mu_{S_i} = \mu + m_{S_i} \\ \mu_{T_i} = \mu + m_{T_i} \\ \alpha_i = \alpha + a_i \\ \beta_i = \beta + b_i \end{cases}$$

is estimated where $(m_{S_i}, m_{T_i}, a_i, b_i)'$ are random effects assumed to follow a zero-mean gaussian distribution with covariance matrix,

$$D = \begin{pmatrix} d_{SS} & d_{ST} & d_{Sa} & d_{Sb} \\ & d_{TT} & d_{Ta} & d_{Tb} \\ & & d_{aa} & d_{ab} \\ & & & d_{bb} \end{pmatrix}.$$

The second approach uses only one step and is based on the following mixed effects model,

$$\begin{aligned} S_{ij} \mid Z_{ij} &= \mu_S + m_{S_i} + (\alpha + a_i) Z_{ij} + \varepsilon_{S_{ij}} \\ T_{ij} \mid Z_{ij} &= \mu_T + m_{T_i} + (\beta + b_i) Z_{ij} + \varepsilon_{T_{ij}}, \end{aligned}$$

where (μ_S, μ_T) and (α, β) are fixed intercepts and treatment effects and (m_{S_i}, m_{T_i}) and (a_i, b_i) are random intercepts and treatment effects. The random effects $(m_{S_i}, m_{T_i}, a_i, b_i)$ are assumed to follow a zero-mean gaussian distribution with covariance matrix D. The error terms are similar to those used in the two-stage approach. The advantages of the second approach lies in its more natural interpretation as it combines fixed and random effects to take into account the heterogeneity of the treatment effects between trials. The authors argued that both approaches yield different results only in very pathological situations. However, as they noted, the second approach is more likely to suffer from computational and convergence issues. They found that these issues were principally due to the number of trials (more convergence issues with less trials) and the heterogeneity of the treatment effect (more convergence issue with less heterogeneity). Based on these two approaches, BUYSE, MOLENBERGHS, BURZYKOWSKI, et al. (2000) defined two measures of surrogacy, at the individual and trial level.

Individual-level surrogacy

As suggested by BUYSE and MOLENBERGHS (1998), the association between S and T after adjusting on the treatment can be used as a measure of surrogacy at the individual level. If,

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ & \Sigma_{22} \end{pmatrix} \right),$$

then following properties of multivariate gaussian distributions, $X_1 \mid X_2$ follows a gaussian distribution such that,

$$\mathbb{E}\left[X_1 \mid X_2\right] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2),$$

and

$$\mathbb{V}\left[X_1 \mid X_2\right] = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}.$$

For the distribution of T_{ij} and S_{ij} given Z_{ij} this gives,

$$\begin{aligned}
\mathbb{E}\left[T_{ij} \mid (S_{ij}, Z_{ij})\right] &= \mathbb{E}\left[T_{ij} \mid Z_{ij}\right] + \frac{\mathrm{Cov}(T_{ij}, S_{ij} \mid Z_{ij})}{\mathbb{V}\left[S_{ij} \mid Z_{ij}\right]}(S_{ij} - \mathbb{E}\left[S_{ij} \mid Z_{ij}\right]); \\
\mathbb{V}\left[T_{ij} \mid (S_{ij}, Z_{ij})\right] &= \mathbb{V}\left[T_{ij} \mid Z_{ij}\right] - \mathrm{Cov}(T_{ij}, S_{ij} \mid Z_{ij})^2\mathbb{V}\left[S_{ij} \mid Z_{ij}\right],
\end{aligned}$$

where,

$$\begin{aligned}
&\mathbb{E}\left[T_{ij} \mid Z_{ij}\right] = \mu_T + m_{T_i} + (\beta + b_i)Z_{ij}; \\
&\mathbb{E}\left[S_{ij} \mid Z_{ij}\right] = \mu_S + m_{S_i} + (\alpha + a_i)Z_{ij}; \\
&\mathrm{Cov}(T_{ij}, S_{ij} \mid Z_{ij}) = \sigma_{ST}; \\
&\mathbb{V}\left[S_{ij} \mid Z_{ij}\right] = \sigma_{SS}; \\
&\mathbb{V}\left[T_{ij} \mid Z_{ij}\right] = \sigma_{TT}.
\end{aligned}$$

Therefore we have,

$$\begin{aligned}
&\mathbb{E}\left[T_{ij} \mid (S_{ij}, Z_{ij})\right] = \mu_T + m_{T_i} + (\beta + b_i)Z_{ij} + \sigma_{ST}^{-1}\sigma_{SS}(S_{ij} - \mu_S + m_{S_i} + (\alpha + a_i)Z_{ij}); \\
&\mathbb{V}\left[T_{ij} \mid (S_{ij}, Z_{ij})\right] = \sigma_{TT} - \frac{\sigma_{ST}^2}{\sigma_{SS}}.
\end{aligned}$$

Finally the individual-level measure of surrogacy is given by the squared correlation between S_{ij} and T_{ij} after adjusting on the treatment and the trial effects:

$$\begin{aligned}
R^2_{\text{indiv}} &= \left(\frac{\mathrm{Cov}(S_{ij}, T_{ij} \mid Z_{ij})}{\sqrt{\mathbb{V}\left(S_{ij} \mid Z_{ij}\right)\mathbb{V}\left(T_{ij} \mid Z_{ij}\right)}}\right)^2 \\
&= \frac{\mathrm{Cov}(S_{ij}, T_{ij} \mid Z_{ij})^2}{\sqrt{\mathbb{V}\left(s_{ij} \mid Z_{ij}\right)}\sqrt{\mathbb{V}\left(T_{ij} \mid Z_{ij}\right)}} \\
&= \frac{\sigma_{ST}^2}{\sigma_{SS}\sigma_{ST}}.
\end{aligned}$$

A surrogate will be validated at the individual level if R^2_{indiv} is sufficiently close to 1.

Trial-level surrogacy

According to BUYSE, MOLENBERGHS, BURZYKOWSKI, et al. (2000), the main motivation behind trial-level surrogacy is to be able to predict the treatment effect on the final endpoint based on the observed effect on the surrogate. Given a new trial $i = 0$ with n_0 subjects, a (fixed effects) model on the surrogate is first fitted:

$$S_{0j} = \mu_{S_0} + \alpha_0 Z_{0j} + \varepsilon_{S_{0j}}, \quad j = 1, \ldots, n_0.$$

Based on the estimations $\hat{\mu}_S$ and $\hat{\alpha}$ obtained from the meta-analysis, predictions of m_{S_0} and a_0 can be derived as:

$$\hat{m}_{S_0} = \hat{\mu}_{S_0} - \hat{\mu}_S;$$
$$\hat{a}_0 = \hat{\alpha}_0 - \hat{\alpha}.$$

where $\hat{\mu}_{S_0}$ and $\hat{\alpha}_0$ are estimated in the new trial. The main objective lies in predicting $\beta + b_0$ given m_{S_0} and a_0. Once again using the properties of normal distribution, we have that $(\beta + b_0 \mid m_{S_0}, a_0)$ follows a normal distribution with,

$$\mathbb{E}\left[(\beta + b_0 \mid m_{S_0}, a_0)\right] = \beta + \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}' \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} \mu_{S_0} - \mu_S \\ \alpha_0 - \alpha \end{pmatrix},$$

and

$$\mathbb{V}\left[(\beta + b_0 \mid m_{S_0}, a_0)\right] = d_{bb} - \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}' \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}.$$

Perfect trial-level surrogacy is achieved if $\mathbb{V}\left[(\beta + b_0 \mid m_{S_0}, a_0)\right] = 0$ or equivalently if

$$\frac{\begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}' \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}}{d_{bb}} = 1.$$

A trial-level measure of surrogacy can then be defined as,

$$R^2_{\text{trial(f)}} = \frac{\begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}' \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}}{d_{bb}}.$$

The authors argue that intuition can be gained if the prediction of b_0 is made independently of m_{S_0}, i.e. the prediction of $\beta_0 + b_0$ is only made given the trial heterogeneity (or deviation) from the average treatment effect on S, a_0. In that case we have,

$$\mathbb{E}\left[(\beta + b_0 \mid a_0)\right] = \beta + \frac{d_{ab}}{d_{aa}}(\alpha_0 - \alpha),$$

and

$$\mathbb{V}\left[(\beta + b_0 \mid a_0)\right] = d_{bb} - \frac{d^2_{ab}}{d_{aa}}.$$

which gives another trial-level measure of surrogacy

$$R^2_{\text{trial(r)}} = \frac{d^2_{ab}}{d_{aa} d_{bb}}.$$

Therefore the difference between $R^2_{\text{trial(f)}}$ and $R^2_{\text{trial(r)}}$ lies in the fact that the former is based on the prediction of $\beta + b_0$ given m_{S_0}, α_0 while the latter only uses the prediiction of $\beta + b_0$ given α_0. The notations $R_{\text{trial(f)}}$ and $R_{\text{trial(r)}}$ emphasize that the former is based on the full random effects model (with both m_{S_i} and a_i) while the latter only uses the reduced mixed effect model (using only a_i). Finally, a surrogate endpoint will be validated at the trial if either $R^2_{\text{trial(f)}}$ or $R^2_{\text{trial(r)}}$ are sufficiently close to 1. We presented here the meta-analysis approach in a setting where both S and T were assumed to be gaussian random variables. The meta-analytic approach has been extended to other types of endpoints but in all cases the spirit remains: the surrogacy is evaluated at both the individual and at the trial level. A review is provided in Burzykowski, Buyse, and Molenberghs (2005). In the next section we briefly present the proposed approach in the case where both S and T are time-to-event and the case where S is a longitudinal biomarker and T a time-to-event.

II.2.2. Time-to-event endpoints

The case where both S and T are time-to-event is frequently encountered, in particular in oncology when the time-to-progression, time-to-disease or time-to-relapse can be used as a surrogate of the time-to-death. In this setting, BURZYKOWSKI, MOLENBERGHS, et al. (2001) proposed to replace the first stage of BUYSE, MOLENBERGHS, BURZYKOWSKI, et al. (2000) by a copula model. They assumed that the joint survival function of (S_{ij}, T_{ij}) can be written as

$$F(s,t) = \mathbb{P}(S_{ij} > s, T_{ij} > t) = C_\theta(F_{S_{ij}}(s), F_{T_{ij}}(t)),$$

where $F_{S_{ij}}(s)$ and $F_{T_{ij}}(t)$ are the marginal survival functions of S_{ij} and T_{ij} and $C_\theta(\cdot)$ is a copula function. Even if any copula function could be used, they limited their choice to single parameter copula for simplicity. To model the effect of the treatment on the survival functions they used proportional hazards models, i.e.

$$\begin{aligned} F_{S_{ij}}(s) &= \exp\left(-\int_0^s \lambda_{S_i}(u)\exp(\alpha_i Z_{ij})\mathrm{d}u\right); \\ F_{T_{ij}}(t) &= \exp\left(-\int_0^t \lambda_{T_i}(u)\exp(\beta_i Z_{ij})\mathrm{d}u\right), \end{aligned}$$

where $\lambda_{S_i}(\cdot)$ and $\lambda_{T_i}(\cdot)$ are trial specific hazard functions and (α_i, β_i) are trial-specific treatment effects. When the hazard functions are specified, the joint model involving the two marginal survival functions and the copula can be estimated using maximum likelihood. At a second stage, similar to BUYSE, MOLENBERGHS, BURZYKOWSKI, et al. (2000), they assumed that,

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} a_i \\ b_i \end{pmatrix},$$

where (a_i, b_i) follows a zero-mean normal distribution with covariance matrix

$$D = \begin{pmatrix} d_{aa} & d_{ab} \\ d_{ab} & d_{bb} \end{pmatrix}.$$

The trial-level measure they used is

$$R^2_{\text{trial(r)}} = \frac{d_{ab}^2}{d_{aa}d_{bb}}.$$

For the individual-level measure, the time-to-event nature of S and T prevents the use of the R^2_{indiv} developed by Buyse, Molenberghs, Burzykowski, et al. (2000). Instead they used the Kendall's τ as a measure of association between S_{ij} and T_{ij} as it depends only on the copula function $C_\theta(\cdot)$ and is independent of the marginal survival functions. Indeed, in this setting, Kendall's τ can be written as:

$$\tau = 4 \int_0^1 \int_0^1 C_\theta(u, v) \mathrm{d}C_\theta(u, v) - 1.$$

This measure describes the strength of association between the two endpoints after adjustment.

In the next section we present the case when the surrogate is a longitudinal biomarker and the final endpoint is a time-to-event.

II.2.3. Longitudinal surrogate and time-to-event final endpoint

The case where the surrogate is a longitudinal biomarker and the final endpoint is a time-to-event endpoint was investigated using the meta-analytic approach by Renard et al. (2003). To be more consistent with the notations used in the litterature on longitidunal endpoint, we note by M the longitudinal surrogate and T the time-to-event endpoint.

Let $\{M_{ijk}, k = 1, \ldots, n_{ij}\}$ be the sequence of repeated measurements of subject j from trial i. The author defined a joint model as follows:

$$\begin{cases} M_{ijk} = \mu_{ij}(t_{ijk}) + W_{1ij}(t_{ijk}) + \varepsilon_{ijk} \\ \lambda_{ij}(t) = H_{ij}(t)\lambda_0(t) \exp\left(\beta_i Z_{ij} + W_{2ij}(t)\right). \end{cases} \tag{II.3}$$

with $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$. In this model $\mu_{ij}(\cdot)$ describes the mean trajectory and β_i is the trial-specific treatment effect on the time-to-event process. The trial-specific treatment effect on the longitidunal biomarker are included in $\mu_{ij}(\cdot)$. The function $H_{ij}(t)$ is the indicator that subject ij is still at risk at time t. The central feature of this approach is the assumption that the zero-mean bivariate gaussian process $W_{ij}(t) = (W_{1ij}(t), W_{2ij}(t))$ describe the association between the two outcomes. The specification of the latent gaussian process can take many form, for example

$$W_{1ij}(t) = \omega_{ij0} + \omega_{ij1} t,$$

corresponds to the classical random intercept and random slope model where $(\omega_{ij0}, \omega_{ij1})$ is a bivariate zero-mean gaussian vector. Then W_{2ij} can be for example,

$$W_{2ij}(t) = \gamma_1 \omega_{ij0} + \gamma_2 \omega_{ij1} t.$$

The estimation of the parameters of the model is carried out through maximum likelihood using the Expectation-Maximization algorithm (DEMPSTER et al., 1977).

The individual-level surrogacy measure can be taken as the association between W_{1ij} and W_{2ij}. Contrary to the case with two normal endpoints (in which the association between the two endpoints is considered), here the association between (W_{1ij} and W_{2ij}, the two components of the latent process W_{ij} that governs the association between both endpoints, is used to derive an individual-level measure of surrogacy. Another difference is that these processes are time-dependent so will be their association. For example if

$$W_{1ij}(t) = \omega_{ij0} + \omega_{ij1} t,$$

and

$$W_{2ij}(t) = \gamma_1 \omega_{ij0} + \gamma_2 \omega_{ij1} t,$$

with

$$\begin{pmatrix} \omega_{ij0} \\ \omega_{ij1} \end{pmatrix} \sim \mathcal{N}(0, G),$$

we have

$$\begin{aligned}
&\mathbb{V}\left[W_{1ij}(t)\right] = G_{11} + t^2 G_{22} + t G_{12}; \\
&\mathbb{V}\left[W_{2ij}(t)\right] = \gamma^2 G_{11} \gamma_2^2 t^2 G_{22} + \gamma_1 \gamma_2 t G_{12}; \\
&\mathrm{Cov}\left(W_{1ij}(t), W_{2ij}(t)\right) = \gamma_1 G_{11} + \gamma_2 t^2 G_{22} + (\gamma_1 + \gamma_2) t G_{12}.
\end{aligned}$$

The author proposed to take the squared correlation between $W_{1ij}(t)$ and $W_{2ij}(t)$ as an individual-level surrogacy measure, i.e.

$$R^2_{\text{indiv}}(t) = \frac{(\gamma_1 G_{11} + \gamma_2 t^2 G_{22} + (\gamma_1 + \gamma_2) t G_{12})^2}{(G_{11} + t^2 G_{22} + t G_{12})(\gamma^2 G_{11} \gamma_2^2 t^2 G_{22} + \gamma_1 \gamma_2 t G_{12})}.$$

The trial-level surrogacy measure is more straightforward but depends on the components in μ_{ij} that depends on Z_{ij}. For example if

$$\mu_{ij}(t_{ijk}) = \mu_{0i} + \mu_{1i}t_{ijk} + \alpha_{0i}Z_{ij} + \alpha_{1i}Z_{ij}t_{ijk},$$

then the R^2_{trial} can be taken as the coefficient of determination of the linear regression model (the hats emphasize that this regression is performed on the estimated values):

$$\hat{\beta}_i = \lambda_0 + \lambda_1\hat{\alpha}_{0i} + \lambda_2\hat{\alpha}_{1i} + \varepsilon_i.$$

The quantifies R^2_{trial} how much the variance in the treatment effect on the final endpoint is explained by the treatment effect on the surrogate. In the case where there is only a single treatment component in the longitudinal part, this coefficient of determination will be equivalent to a squared correlation.

II.2.4. The surrogate threshold effect

Burzykowski and Buyse (2006) introduced a new measure of surrogacy called "surrogate threshold effect"(STE). They defined it as "the minimum treatment effect on the surrogate required to predict a significant effect on the final endpoint". Using the notations introduced in Section II.2.1, for a new trial $i = 0$, the $(1-\gamma)100\%$ prediction interval for $\beta + b_0$ is given by

$$\mathbb{E}\left(\beta + b_0 \mid \alpha + a_0\right) \pm z_{1-\gamma/2}\sqrt{\mathbb{V}\left(\beta + b_0 \mid \alpha + a_0\right)},$$

where $z_{1-\gamma/2}$ is the $(1-\gamma/2)$ percentile of the standard normal distribution. In order to predict a significant effect of the treatment on the final endpoint, the lower bound of this interval should be greater than 0 (or the upper bound lower than 0 depending on the type of effect one is interested in). This interval only depends on $\alpha + a_0$, the treatment effect on the surrogate. Let $l_o(\alpha + a_0)$ be the lower bound of this interval, then the STE is the value of $\alpha + a_0$ such that

$$l_o(\alpha + a_0) = 0.$$

One advantage of the STE is its straightforward interpretation in terms of the treatment effect that we have to observe on the surrogate in order to predict a significant effect on the final

endpoint. As noted by BURZYKOWSKI and BUYSE (2006), the STE can be computed for any type of surrogate and one only needs to use an appropriate joint model to derive the measures of treatment effect. Another key feature of the STE is that it can be used to reassess the sample size of an ongoing trial based on the observed effect on the surrogate. For example, if the lower bound of the prediction interval for the treatment effect on the final endpoint computed using the observed effect on the surrogate and is larger than the STE, then a larger treatment effect on the final endpoint might be predicted.

We have presented the meta-analytic approach for surrogate endpoint validation which is the main approach that has been developed in the past two decades. However, another approach has also been developed based on the identification of the indirect effect. The early developments of this approach can be traced back to Prentice (PRENTICE, 1989) and Freedman (FREEDMAN et al., 1992) who relied on the idea that a surrogate is validated if it captures enough of the treatment effect on the final endpoint. However as described in the beginning of this chapter, these approaches were likely to be biased as they did not take into account potential confounding between the surrogate and the final endpoint. Methods to overcome this issues have been developed within the causal framework of mediation analysis which we present in the next section. We give a general introduction to the causal mediation analysis and give examples of its application to surrogate validation.

II.3. | MEDIATION ANALYSIS

In the context of surrogate endpoint validation, the purpose of mediation analysis is to decompose the total treatment effect on the final endpoint T as an indirect effect through the surrogate (the mediator) S and a direct effect independent of the latter. This decomposition has been the core of the mediation analysis, whose developments since the past decades have been made within the framework of causal inference (J. M. ROBINS and GREENLAND, 1992) which relies on the use of the potential (or counterfactual) outcomes framework (HOLLAND, 1986; LANGE et al., 2017).

II.3.1. Counterfactual framework

Let $S(z)$ be the value of the surrogate that would have been observed had the patient been assigned to treatment $Z = z$. Therefore, for a binary treatment this subject has two counterfactual surrogate outcomes, $S(1)$ and $S(0)$. At the individual level the quantity,

$$S(1) - S(0),$$

carries a causal interpretation of the treatment effect as the only element that differs between $S(1)$ and $S(0)$ is the treatment status. Averaging over the population,

$$\mathbb{E}\left[S(1) - S(0)\right],$$

yields the average treatment effect (ATE). In most clinical studies only one potential outcome will be observed for each subject, either $S(1)$ or $S(0)$ depending on which treatment was assigned to this subject. In the same way we can define $T(z, s)$ as the value of the final endpoint that would be observed for the subject if the treatment is set to $Z = z$ and the surrogate set to $S = s$. Being a posttreatment outcome, the surrogate cannot, in general, be "controlled" or fixed to any particular value. Therefore it is often more reasonable to consider the "nested" counterfactual outcome $T(z, S(z'))$ corresponding to the value of the final endpoint that would be observed if the treatment is set to $Z = z$ for T but to z' for S. The main difference between $T(z, s)$ and $T(z, S(z'))$ is that in the latter, S is not fixed to any value s but is random and "behaves" as if it was under the treatment arm $Z = z'$.

II.3.2. Identifiability assumptions

In order to be able to infer on the distribution of the counterfactual outcomes $S(z)$ and $T(z, S(z'))$ based on the observed outcomes S and T, we need to rely on several identifiability assumptions.

The Stable Unit Treatment Value Assumption (SUTVA)

First, SUTVA (Rubin, 1980; Holland, 1986) ensures that for a subject each potential outcome is uniquely defined. This assumption is composed of two parts:

1. The potential outcome of a subject does not depend on the treatment assignment of

other subjects

2. Each subject receives the same version of the treatment

The first part of the SUTVA states that for a subject receiving treatment $Z = z$, his associated potential outcomes are $S(z)$ and $T(z, S(z))$. Since $S(z)$ and $T(z, S(z))$ only depend on z, the subecject's treatment assignement, it does not depends on the treatment other subject have received. Violations of the first part of the SUTVA can arise for example in infectious diseases where the probability of a subject to be contaminated not only depends on his vaccination status but also on the vaccination status of other subjects in the study he may be in contact with. Another example can be found in social sciences when, for example, within a school some students were assigned a certain tutoring program and others were not; it is likely that the interactions between students will also influence those who were not assigned to the tutoring program. In a randomized trial, decisions can be made at the design stage to ensure this first part of SUTVA.

The second part of the SUTVA states that each version of the treatment is uniquely defined. In clinical trials investigating drugs or medical devices this assumption holds as each patient receive the same version of a drug or the same device. However if, say, in a trial on quality of life, the treatment (or intervention) is "physical activity 4 hours per week", this is not uniquely defined as there are multiple ways of doing 4 hours of physical activity per week, either by doing it on a daily basis or by doing 4 hours of activity in one day. It that case the two versions of the same intervention may yield a different outcome. An extension of the "no multiple version of a treatment" assumption was introduced by VanderWeele and M. A. Hernan (2013) as the "treatment variation irrelevance" in that different versions of a treatment should result in the same potential outcome.

Consistency

The consistency assumption states that for a subject, the observed outcome equals the potential outcome associated with the treatment assigned to this subect. This assumption is what makes

the link between the potential outcomes and the observed outcomes. It can be written as:

$$\{Z = z \Rightarrow S = S(z)\} \iff \{S = S(Z)\},$$
$$\{Z = z \ \wedge S = s \Rightarrow T = T(z, s)\} \iff \{T = T(Z, S(Z))\}.$$

Conditional exchangeability

The conditional exchangeability (also called "conditional ignorability assumption") broadly states that there is no unmeasured confounders on the pathway between S and T, and that the measured confounders have appropriately been taken into account.

A (much) more stringent assumption is the (unconditional) *exchangeability* assumption which for the surrogate S it can be written as,

$$S(z) \perp\!\!\!\perp Z, \ \forall z = 0, 1.$$

The exchangeability means that the potential outcomes $S(0)$ and $S(1)$ are independent of the treatment assignment Z. Therefore, patients in the control group ($Z = 0$) have, on average, the same potential outcomes $(S(0), S(1))$ as the patients in the experimental group ($Z = 1$). A patient in the experimental group has (on average) the same $S(1)$ (and $S(0)$) as a patient in the control group. It just happens that for the patient in the experimental group we observe $S(1)$ while for the patient in the control group we observe $S(0)$. Other exchangeability assumptions can be stated (M. Hernan and J. Robins, 2020), such as the *strong exchangeability*:

$$(S(1), S(0)) \perp\!\!\!\perp Z,$$

or the *average exchangeability*:

$$\mathbb{E}\left[S(z) \mid Z = 1\right] = \mathbb{E}\left[S(z) \mid Z = 0\right].$$

The randomization of the treatment ensures the strong exchangeability, and we have the following chain of implications,

$$\text{Strong exchangeability} \Rightarrow \text{Ignorability} \Rightarrow \text{Average exchangeability}.$$

It should be also noted that for a binary variable the exchangeability is equivalent to the av-

erage exchangeability (since the distribution is fully identified by its mean), but it does not hold for other types of variables whose distribution may depends on other parameters than the mean. The exchangeability does not imply strong exchangeability since marginal independence does not (in general) imply joint independence. The use of the consistency and exchangeability allows us to be able to derive the ATE based on the observed outcomes:

$$\begin{aligned}\text{ATE} &= \mathbb{E}\left[S(1)\right] - \mathbb{E}\left[S(0)\right] \\ &= \mathbb{E}\left[S(1) \mid Z = 1\right] - \mathbb{E}\left[S(0) \mid Z = 0\right] \quad \text{(exchangeability)} \\ &= \mathbb{E}\left[S \mid Z = 1\right] - \mathbb{E}\left[S \mid Z = 0\right] \quad \text{(consistency)}.\end{aligned}$$

Hence when these two assumptions are met, the ATE which carries a causal interpretation (being related to potential outcomes) can be estimated from observed outcomes (last line). The exchangeability assumption is also labelled "no confounders". In the presence of confounding influencing both Z and S, then $S(1)$ and $S(0)$ may be distributed differently across the treatment arms. For example if we have,

$$\begin{aligned}S(z) &\sim \beta_0 + \beta_1 z + \beta_2 U; \\ \text{logit}\left[\mathbb{P}(Z = 1)\right] &\sim \gamma_1 + \gamma_2 U,\end{aligned}$$

then $S(z)$ is distributed differently given $Z = 1$ or $Z = 0$. If either β_2 or γ_2 or both are equal to 0 then $S(z) \perp\!\!\!\perp Z$. The *conditional exchangeability* assumption is simply the exchangeability assumption conditioned on measured covariates. Let X be such a covariate, the conditional exchangeability is:

$$S(z) \perp\!\!\!\perp Z \mid X, \, \forall z,$$

that is X is the only confounder between S and Z, or equivalently, there is no unmeasured confounders (X being measured).

Sequential ignorability

In mediation analysis, the introduction of nested potential outcomes requires an extented version of the ignorability assumption called the "sequential ignorability" (SI) assumption (IMAI, KEELE, and YAMAMOTO, 2010). Several sequential ignorability assumptions have been pro-

posed in the litterature (Pearl, 2022; Petersen et al., 2006; J. M. Robins, 2003), and we use the following (Imai, Keele, and Tingley, 2010) :

$$\{T(z', s), S(z)\} \perp\!\!\!\perp Z \mid X;$$
$$T(z', s) \perp\!\!\!\perp S(z) \mid Z = z, X = x.$$

The first part of this sequential ignorability assumption is similar to the "strong ignorability" defined earlier and states that X should capture all the confounding between the treatment and S and T. As stated before, if Z is randomized this assumption holds. The second part can be understood as the adaptation of the first part in which the treatment Z is replaced by $S(z)$ and the covariate X is replaced by $(X, Z = z)$. Like the ignorability assumption which states that both group $Z = 0$ and $Z = 1$ should have the same potential outcomes distributions, the second part of the SI states that for any two groups $S(z) = s_1$ and $S(z) = s_2$ the distribution of the potential outcomes $T(z', s)$ is the same within these groups. To better understand this, consider the simple model:

$$S(z) = \beta_0 + \beta_1 z + \beta X + \varepsilon_S;$$
$$T(z', s) = \gamma_0 + \gamma_1 z' + \gamma_2 s + \gamma X + \varepsilon_T,$$

with $\varepsilon_S \perp\!\!\!\perp \varepsilon_T$. Then, given $Z = z$ and $X = x$, $S(z) = \beta_0 + \beta_1 z + \beta x + \varepsilon_S \perp\!\!\!\perp T(z', s) \sim \gamma_0 + \gamma_1 z' + \gamma_2 s + \gamma x + \varepsilon_T$. However, had we not taken into account the common cause X between S and T, this independence would not hold. Hence, while the first part of the SI states that there is no confounders between Z and (S, T) besides X and holds even inconditionally if Z is randomized, the second part is much more stringent and assumes that S and Tdo not share unmeasured common cause besides X. Moreover, the covariates X must be pre-treatment covariates as the SI does not hold if X depends on Z even if it is measured (Avin et al., 2005).

II.3.3. Natural direct and indirect effects

From the counterfactual outcomes $T(z, S(z'))$ we can decompose the total treatment effect (TTE) on T, defined as,

$$\text{TTE} = \mathbb{E}\left[T(1, S(1))\right] - \mathbb{E}\left[T(0, S(0))\right],$$

as a direct effect and an indirect effect through the surrogate:

$$\text{TTE} = \underbrace{\mathbb{E}\left[T(1, S(1))\right] - \mathbb{E}\left[T(1, S(0))\right]}_{\text{Indirect Effect}} + \underbrace{\mathbb{E}\left[T(1, S(0))\right] - \mathbb{E}\left[T(0, S(0))\right]}_{\text{Direct Effect}}.$$

The indirect effect is the average difference between the potential outcome when the treatment is set to 1 for both S and T, $T(1, S(1))$ and the potential outcome when the treatment is set to 1 for T but to 0 for S. The difference between the two potential outcomes therefore only lies in the treatment status for S hence their difference quantifies the indirect effect of the treatment on T through S. Similarly, the direct effect is the average difference between the two potential outcomes $T(1, S(0))$ and $T(0, S(0))$ whose only difference is the treatment status on T. It should be noted that $T(1, S(1)) - T(0, S(1))$ and $T(0, S(1)) - T(0, S(0))$ also quantifies a direct effect and an indirect effect. J. M. ROBINS (2003) called $T(0, S(1)) - T(0, S(0))$ the "pure indirect effect" and $T(1, S(1)) - T(1, S(0))$ the "total indirect effect" while PEARL (2022) called them "natural indirect effects". In addition to the natural direct effects, controlled direct effects (CDE) are effects in which the surrogate value is fixed (VANDERWEELE, 2011b):

$$\text{CDE}(s) = T(1, s) - T(0, s).$$

The difference between natural direct effects and controlled direct effects is that in the former the surrogate is not fixed to any particular value but "behaves" as if it was under treatment $Z = z$. Controlled direct effects are useful in settings where the variable S can explicitly be set and controlled, such as policy decision making. However, from a surrogate endpoint perspective in clinical trials, the natural effects are more relevant as such a posttreatment variable cannot be controlled (and if a surrogate could be controlled the whole process of validating it would be pointless). In the following we will use NDE for natural direct effect

and NIE for natural indirect effect.

Given the assumptions defined in the previous section (SUTVA, consistency, sequential ignorability) we have the following important result:

$$\mathbb{E}\left[T(z, S(z')) \mid X\right] = \int_{\mathcal{S}} \mathbb{E}\left[T \mid S = s, Z = z, X\right] f_{S|Z=z',X}(s)\mathrm{d}s. \tag{II.4}$$

Here $\mathcal{S}$ denotes the support of S and $f_U(\cdot)$ denotes the a density function of a random variable U. This equation links the potential outcomes to the *observed* outcomes S and T conditional on the treatment Z and covariable X. Therefore, the causal quantity $\mathbb{E}\left[T(z, S(z')) \mid X\right]$ (causal because it is based on potential outcomes) can be derived from the observed data. Proof of Equation II.4 is given in Annex A.1.

A populational value of $T(z, S(z'))$ can be obtained by marginalizing over the distribution of X,

$$\mathbb{E}\left[T(z, S(z'))\right] = \int_{\mathcal{X}} \int_{\mathcal{S}} \mathbb{E}\left[T \mid S = s, Z = z, X = x\right] f_{S|Z=z',X=x}(s) f_X(x)\mathrm{d}s\mathrm{d}x.$$

Finally, the natural direct and indirect effect can be obtained as,

$$\begin{aligned}\mathrm{NDE} = \int_{\mathcal{X}} \int_{\mathcal{S}} \Big\{ &\mathbb{E}\left[T \mid S = s, Z = 1, X = x\right] - \mathbb{E}\left[T \mid S = s, Z = 0, X = x\right] \Big\} \\ &\times f_{S|Z=0,X=x}(s) f_X(x)\mathrm{d}s\mathrm{d}x,\end{aligned}$$

and

$$\mathrm{NIE} = \int_{\mathcal{X}} \int_{\mathcal{S}} \mathbb{E}\left[T \mid S = s, Z = 1, X = x\right] \times \Big\{ f_{S|Z=1,X=x}(s) - f_{S|Z=0,X=x}(s) \Big\} f_X(x)\mathrm{d}s\mathrm{d}x.$$

The ATE can be obtained in the same way as

$$\begin{aligned}\mathrm{ATE} = &\int_{\mathcal{X}} \int_{\mathcal{S}} \mathbb{E}\left[T \mid S = s, Z = 1, X = x\right] f_{S|Z=1,X=x}(s) f_X(x)\mathrm{d}s\mathrm{d}x \\ &- \int_{\mathcal{X}} \int_{\mathcal{S}} \mathbb{E}\left[T \mid S = s, Z = 0, X = x\right] f_{S|Z=0,X=x}(s) f_X(x)\mathrm{d}s\mathrm{d}x.\end{aligned}$$

However, in the case of a randomized trial, $T(1, S(1))$ and $T(0, S(0))$ are the endpoints observed in the experimental and control arm respectively and ATE can simply be estimated as the difference of the sample average of T in each arm. The above formula are useful as one can not observe endpoints such as $T(1, S(0))$ whose average intervenes in both the NDE and NIE. Although we have focused on the expectation of the random variables $T(z, S(z'))$, similar results applies to any function h of these variables:

$$\mathbb{E}\left[h\left(T(z, S(z'))\right)\right] = \int_{\mathcal{X}} \int_{\mathcal{S}} \mathbb{E}\left[h\left(T\right) \mid S = s, Z = z, X = x\right] f_{S|Z=z',X=x}(s) f_X(x) \mathrm{d}s \mathrm{d}x.$$

For example if T is a time-to-event and we are interested in estimating the NDE and NIE on a survival scale, one could use $h(x; t) = \mathbb{1}_{x>t}$ such that,

$$\begin{aligned}\mathbb{E}\left[h\left(T(z, S(z')); t\right)\right] &= \mathbb{E}\left[\mathbb{1}_{T(z,S(z'))>t}\right] \\ &= \mathbb{P}\left(T(z, S(z')) > t\right),\end{aligned}$$

which then gives time-dependent NDE and NIE,

$$\begin{aligned}\mathrm{NDE}(t) &= \mathbb{P}\left(T(1, S(0)) > t\right) - \mathbb{P}\left(T(0, S(0)) > t\right); \\ \mathrm{NIE}(t) &= \mathbb{P}\left(T(1, S(1)) > t\right) - \mathbb{P}\left(T(1, S(0)) > t\right).\end{aligned}$$

II.3.4. Application to surrogate endpoint validation

Applications of causal mediation analysis have been made in the contex of surrogate endpoints validation. Most of these applications use the proportion of treatment effect (PTE) as a measure of surrogacy, defined as the ratio of indirect effect over total treatment effect (TTE):

$$\mathrm{PTE} = \frac{\mathrm{NIE}}{\mathrm{TTE}}. \tag{II.5}$$

This PTE is closely related to the one defined in Equation II.1. However, there is two key differences between the two. First the PTE defined in Equation II.5 carries a causal intepretation as in relies on potential outcomes and secondly, it is not attached to any particular model in contrary to the PTE given in II.1. In the case of a gaussian outcomes the PTE in Equation II.1 has a natural intepretation in terms of average effects (since β and β_S will be then

be average treatment effects), but for more complex or non linear models this is not the case anymore. For example in a simple proportional hazards models,

$$\lambda_T(t \mid Z) = \lambda \exp(\beta Z);$$
$$\lambda_T(t \mid S, Z) = \lambda' \exp\left(\beta_S Z + \gamma S\right),$$

even if $\beta_S = \frac{1}{2}\beta$ and thus $\mathrm{PTE} = 1 - \frac{\beta_S}{\beta} = \frac{1}{2}$ it does not necessarily expands to a 50% mediated effects in terms of survival probability. However, when the outcome is rare, the natural direct and indirect effect on the log-hazard scale can be derived directly from the regression parameters (VanderWeele, 2011a). On the other hand, the PTE defined in II.5 does not make use of naive ratio of coefficients but can be used to assess the proportion of treatment effect in more natural scales such as survival, cumulative hazards, odds-ratio etc. We now present two published applications of the mediation analysis for surrogate validation, but others applications can be found in the literature (Parast et al., 2017; Wang, Parast, Tian, et al., 2022; Parast et al., 2022).

Vandenberghe et al. (2018) used causal mediation analysis to investigate the proportion of treatment effect on a time-to-event endpoint T through a binary surrogate S. They defined their measure of surrogacy as the PTE expressed in the survival scale:

$$\mathrm{PTE}(t) = \frac{\mathbb{P}(T(1, S(0)) > t) - \mathbb{P}(T(1, S(0)) > t)}{\mathbb{P}(T(1, S(1)) > t) - \mathbb{P}(T(0, S(0)) > t)}.$$

They applied their proposed method to a clinical trial dataset on breast cancer patients on neo-adjuvant chemotherapy and investigated if the pathological complete response (pCR), S, at the time of surgery was a surrogate of the disease-free survival. They concluded that a very small proportion of the treatment effect on the DFS (less than 5%) was due to its effect on pCR.

Weir et al. (2022) used a multi-state model approach to investigate mediation of a treatment on a time-to-event endpoint T through another time-to-event endpoint S. They defined two causal estimands expressed as a proportion of treatment effect in two different scales. First in the cumulative risk,

$$\mathrm{PTE}_1(t) = \frac{H_{1S_1}(t) - H_{1S_0}(t)}{H_{1S_1}(t) - H_{0S_0}(t)},$$

where $H_{zS_{z'}}(t)$ is the cumulative hazard of T if the treatment is set to z for T but to z' for S.

Their second estimand was the proportion of treatment effect in the restricted mean survival time(ROYSTON and PARMAR, 2013),

$$\mathrm{PTE}_2(t) = \frac{\mu_{1S_1}(t) - \mu_{1S_0}(t)}{\mu_{1S_1}(t) - \mu_{0S_0}(t)},$$

where $\mu_{zS_{z'}}(t) = \int_0^t H_{zS_{z'}}(u)\mathrm{d}u$. They applied their proposed method to a dataset of a randomized clinical trial on prostate cancer patients to investigate the time-to-metastasis as a surrogate of the time to prostate cancer-specific death for radical prostatectomy ($Z = 1$) and waitfull watching ($Z = 0$) with death from other causes as a competing evetn. For both their measures they found a proportion of treatment mediated of about 35% at 18 years and suggested that time-to-metastasis might be an interesting surrogate in this setting.

II.4. | JOINT MODELS IN SURVIVAL ANALYSIS

In survival analysis, joint models were developed to circumvent issues related to endegenous time-dependent covariates effects on a time-to-event outcome (WULFSOHN and TSIATIS, 1997; KALBFLEISCH and PRENTICE, 2011; TSIATIS and DAVIDIAN, 2004; RIZOPOULOS, 2012b).

An endogeneous variable is a variable whose distribution is affected by the time-to-event outcome. These variables are often subject-specific, such as blood pressure, HIV viral load or prostate-specific antigen levels. Classical regression models such as Cox model for time-dependent covariates are likely to be biased in this setting by not appropriately taking into account the dependency between endogeneous variate and the time-to-event. Joint models jointly modelize the endogenous and time-to-event processes over time by assuming a latent structure between the two processes. This latent structure often takes the form of random effects or latent classes.

In the following sections we introduce our proposed joint models for surrogate validation: the first one for a longitudinal continuous surrogate and the second one for a time-to-event surrogate. In both cases the final endpoint is a time-to-event. These models are defined based on counterfactual endpoints to allow their intepretation in the realm of causal mediation analysis. Furthermore we consider the setting of meta-analyses, with K the number of trials (or centers). The link with both the meta-analytic and mediation analysis valiation approaches is presented for each model. We first present the case when the surrogate is a longitudinal

biomarker.

II.4.1. Joint model for a longitudinal surrogate and a time-to-event endpoint

In this section we assume that the potential surrogate endpoints is a longitudinally repeated continuous endpoint. Henceforth the terms "longitudinal endpoint" and "surrogate endpoint" are used interchangeably. An issue with continuous longitudinal outcomes is that they are often measured with error. When investigating their influence on a time-to-event, these measurements errors will bias the estimated effects towards 0(Hughes, 1993). Two-steps approaches were first proposed, in which the longitudinal endpoint is first modeled at each event time on the patients still at risk at that time using a linear mixed model and then the prediction of the (error-free) variable for each individuals at risk (Self and Pawitan, 1992) is used in a proportional hazards model. However these two-step approaches were shown to produce biased results (Sweeting and Thompson, 2011). Joint models were therefore developed to appropriately take into account both the association between the two outcomes and the measurement errors of the longitudinal variable.

Notations and potential outcomes

To be more consistent with the notations used in the literature, we will used M to denote the longitudinal continuous biomarker instead of S. Let i denotes the ith trial and ij the jth subject of trial i, $j = 1, \cdots, n_i$. We assume that the follow-up is bounded by a maximal follow-up time τ. In the following we denote by $Z_{ij} \in \{0, 1\}$ and $X_{ij} \in \mathbb{R}^p$ the treatment and baseline covariates of subject ij. In the counterfactual outcome framework, let M_{ij}^z be the potential longitudinal outcome of subject ij had his/her treatment been set to z,

$$M_{ij}^z = \{M_{ij}^z(t), 0 \leq t \leq \tau\},$$

$T_{ij}(z, m)$ and $C_{ij}(z, m)$ the time-to-event and censoring time of subject ij if the treatment and the longitudinal biomarker are fixed to z and m respectively. The *potentially observable* outcomes are $M_{ij} = M_{ij}^{Z_{ij}}$, $T_{ij} = T_{ij}(Z_{ij}, M_{ij}^{Z_{ij}})$ and $C_{ij} = C_{ij}(Z_{ij}, M_{ij}^{Z_{ij}})$. Since M_{ij} is a longitudinal continuous variable, it could be measured with an error thus we denote by $\tilde{M}_{ij}$ the observed longitudinal outcome. The observed follow-up time is $T_{ij}^* = \min(T_{ij}, C_{ij})$ with its associated indicator $\delta_{ij} = \mathbb{1}_{T_{ij} \leq C_{ij}}$.

Proposed joint model

Our proposed joint model for the potential outcomes $M_{ij}^{z_1}$ and $T_{ij}(z_2, m)$ is:

$$\begin{cases} \tilde{M}_{ij}^{z_1}(t) = M_{ij}^{z_1}(t) + \varepsilon_{ij}(t) \\ \qquad\quad = \theta'_{ij} f(t) + \beta'_M X_{ij}^M + (\beta_{Z,M} + \nu_{M,i}) z_1 + \varepsilon_{ij}(t) \\ \lambda_{ij}^{z_2 m}(t) = \lambda_0(t) \exp\left((\beta_{Z,T} + \nu_{T,i})\, z_2 + \beta'_T X_{ij}^T + \eta' h\,(m) \right). \end{cases} \tag{II.6}$$

This joint model is composed of two submodels: one for the longitudinal surrogate and the other for time-to-event endpoint. In this model the covariate vectors X_{ij}^T and X_{ij}^M are subvectors of X_{ij}. These baseline covariates are associated with fixed parameters vector β_M and β_T respectively. In the longitudinal submodel the error terms ε_{ij} are assumed to be independent and normally distributed, $\varepsilon_{ij}(t) \sim \mathcal{N}(0, \sigma^2)$. The term $\theta'_{ij} f(t)$ represent the individual trajectory of the biomarker over time where $f(t)$ is a time-dependent vector and θ_{ij} is its individual association which is the sum of a population-level effect θ and individual-level random effects ω_{ij}. For example, a linear evolution can be described with $f(t) = (1, t)'$ for which $\theta = (\theta_0, \theta_1)'$ and $\omega_{ij} = (\omega_{ij0}, \omega_{ij1})'$ leading to $\theta_0 + \omega_{ij0} + (\theta_1 + \omega_{ij1})t$, i.e. a random intercept and slope model.

The individual random effects are assumed gaussian with an unconstrained covariance matrix D:

$$\omega_{ij} \sim \mathcal{N}\,(0, D)\,.$$

Note that it is not required that each component of $f(t)$ is associated with an individual-level random effect. Once again taking the example of $f(t) = (1, t)$ then it is possible to only include a random intercept $\omega_{ij} = \omega_{ij0}$. The term $\beta_{Z,M} + \nu_{M,i}$ is the treatment effect on the biomarker equal to the sum of the fixed effect $\beta_{Z,M}$ and the trial-level random effect ν_{M_i}.

In the survival part, the function $\lambda_{ij}^{z_2 m}(t)$ is the hazard function of T_{ij} given the treatment value is z_2 and the biomarker path m. The function λ_0 is the baseline hazard function that is modeled parametrically either by assuming a common distribution (such as Weibull or Gamma) or by using more flexible approaches such as piecewise-constant functions or splines functions. As for the longitudinal submodel, the term $\beta_{Z,T} + \nu_{T,i}$ represents the trial-level effect of the treatment on the risk of occurrence of the time-to-event endpoint. The trial-level

random effects are correlated zero-mean normal variables,

$$\begin{pmatrix} \nu_{M,i} \\ \nu_{T,i} \end{pmatrix} \sim \mathcal{N}(0, \Sigma), \tag{II.7}$$

where Σ is a 2×2 unstructured covariance matrix. The longitudinal and survival submodels are linked together through the "link function" $h(\cdot)$ which describes the dependence structure between the longitudinal marker and the time-to-event. The function itself is not estimated but has to be specified *a priori* based on both biological and clinical grounds. On the other hand, the association parameter η is estimated. This parameter quantifies the magnitude of the dependence between both outcomes. For a biomarker trajectory $m = \{m(t), 0 \leq t \leq \tau\}$, the most common link functions found in the literature are the following:

- The "current level" link: $h(m) = m(t)$. This is probably the most natural link between the biomarker and the time-to-event that assumes that the risk occurrence of T at any time t depends only on the current value of the biomarker $m(t)$.
- The "current slope" link: $h(m) = \frac{\mathrm{d}m(t)}{\mathrm{d}t}$. Rather than considering an association between the hazard function of T with the current value of m at time t, we rather consider an association with the increase (or decrease) of m at time t.
- $h(m) = \omega f(t)$. This link function assumes that the risk of occurrence of T depends on the trajectory $m(t)$ only through the deviation of individual ij from the average trajectory.
- $h(m) = \omega$. This is the "shared random-effects" link function. As it is time-independent, it assumes that dependence structure between the two outcomes is constant over time regardless of the biomarker evolution.

These links are either time-fixed (shared random effects) or time-dependent but are all a single summary measure taken at time t. Note that these links could also be combined together, for example the combination of the current level and the current slope is possible $h(m) = \left(m(t), \frac{m(t)}{\mathrm{d}t}\right)$. In some cases it may be more relevant to consider a delay ζ between the longitudinal biomarker and its effect on the risk of occurrence of T, for example instead of taking the current value at time t one could include the current value at time $t - \zeta$. Of

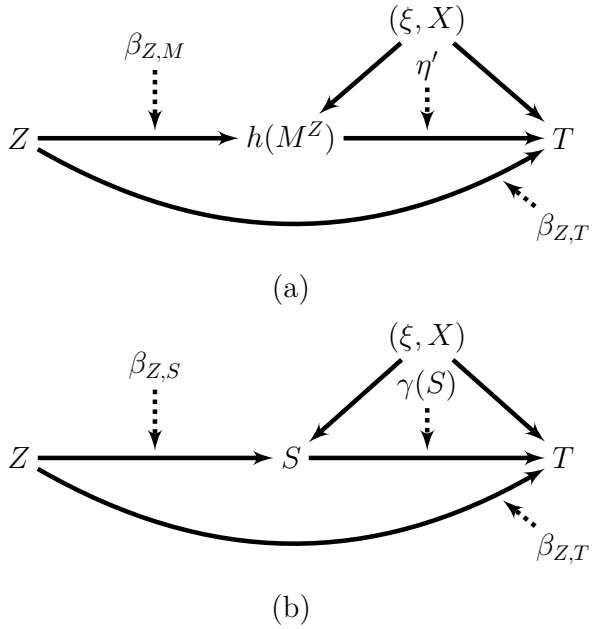


Figure II.2.: DAG associated with the joints models: (a) for a longitudinal surrogate (M) and (b) for a time-to-event surrogate (S). Here ξ represents the random effects in these models.

course for $t < \zeta$ this value is undefined and thus one can consider the link to be null. Hence we would have $h(m) = m(t-\zeta)\mathbb{1}_{t\geq\zeta}$.

Another possibility is to take an aggregate measure of the overall trajectory up to time t such as the weighted sum of past biomarker value $h(m) = \int_0^t w(u)m(u)\mathrm{d}u$ where $w(\cdot)$ is a weight function allowing to put more emphasis on earlier or latter biomarker values. The different parameters of model II.6 are represented in the DAG given in Figure II.2(a) (where ξ denotes the vector of all random effects).

Likelihood computation and estimation

Let ϕ be the vector of all parameters in Model II.6. For computational purposes, the covariance matrices D and Σ are parametrized through their Cholesky decomposition as to not impose positivity constraints on the underlying parameters,

$$D = LL', \Sigma = BB',$$

Therefore we have,

$$\phi = (\lambda_{0,T}, \eta, \sigma, \theta, L, B, \beta_M, \beta_T, \beta_{Z,M}, \beta_{Z,T}).$$

In a meta-analysis, the complete likelihood of the data is the product of the contribution of each trial:

$$L(\phi) = \prod_{i=1}^{K} L_i(\phi),$$

where $L_i(\phi)$ is the likelihood for the ith trial. Given the random effetcs $\nu_i \equiv (\nu_{M,i}, \nu_{T,i})$, the observations from trial i are independent, hence we have,

$$\begin{aligned} L_i(\phi) &= \int_{\mathbb{R}^2} L_i(\phi, \nu_i) \mathrm{d}\nu_i \\ &= \int_{\mathbb{R}^2} L_i(\phi \mid \nu_i) f_\nu(\nu_i) \mathrm{d}\nu_i \\ &= \int_{\mathbb{R}^2} \left(\prod_{j=1}^{n_i} L_{ij}(\phi \mid \nu_i) \right) f_\nu(\nu_i) \mathrm{d}\nu_i, \end{aligned}$$

where $L_{ij}(\phi \mid \nu_i)$ is the contribution of subject ij given ν_i. The function $f_\nu(\cdot)$ is the density function of a zero-mean bivariate gaussian vector with covariance matrix Σ:

$$f_\nu(\nu_i) = \frac{1}{(2\pi)\sqrt{\det(\Sigma)}} \exp\left\{ -\frac{\nu_i^\intercal \Sigma^{-1} \nu_i}{2} \right\}.$$

The individual likehood $L_{ij}(\phi \mid \nu_i)$ can be averaged over the individual-level random effects ω_{ij}. Omitting the dependence on the covariates X_{ij} and Z_{ij} to ease the notations, we have,

$$\begin{aligned} L_{ij}(\phi \mid \nu_i) &= p_\phi(T_{ij}^*, \delta_{ij}, \tilde{M}_{ij} \mid \nu_i) \\ &= \int_{\mathbb{R}^p} p_\phi(T_{ij}^*, \delta_{ij}, \tilde{M}_{ij}, \omega_{ij} \mid \nu_i) \mathrm{d}\omega_{ij} \\ &= \int_{\mathbb{R}^p} p_\phi(T_{ij}^*, \delta_{ij}, \tilde{M}_{ij} \mid \nu_i, \omega_{ij}) f_\omega(\omega_{ij} \mid \nu_i) \mathrm{d}\omega_{ij} \\ &= \int_{\mathbb{R}^p} p_\phi(T_{ij}^*, \delta_{ij}, \tilde{M}_{ij} \mid \nu_i, \omega_{ij}) f_\omega(\omega_{ij}) \mathrm{d}\omega_{ij} \\ &= \int_{\mathbb{R}^p} p_\phi(T_{ij}^*, \delta_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}) p_\phi(\tilde{M}_{ij} \mid \nu_i, \omega_{ij}) f_\omega(\omega_{ij}) \mathrm{d}\omega_{ij}, \end{aligned} \tag{II.8}$$

where p is the number of individual-level random effects and $f_\omega(\cdot)$ is the density function of a multivariate normal distribution with mean 0 and (unstructured) covariance D:

$$f_\omega(\omega_{ij}) = \frac{1}{(2\pi)^{p/2}\sqrt{\det(D)}} \exp\left\{ -\frac{\omega_{ij}^\intercal D^{-1} \omega_{ij}}{2} \right\}.$$

The fourth equality in Equations II.8 comes from of independence between ν_i and ω_{ij}.

Assuming that the censoring time C_{ij} is independent of T_{ij} and non-informative we have,

$$\begin{aligned} p_\phi(T^*_{ij}, \delta_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}) &= \left[\lambda(T^*_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}) S(T^*_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij})\right]^{\delta_{ij}} S(T^*_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij})^{1-\delta_{ij}} \\ &= \lambda(T^*_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij})^{\delta_{ij}} S(T^*_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}), \end{aligned}$$

where,

$$\lambda(T^*_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}) = \lambda_0(T^*_{ij}) \exp\left\{(\beta_{Z,T} + \nu_{T,i})\, Z_{ij} + \beta'_T X^T_{ij} + \eta' h\left(M_{ij}\right)\right\} \tag{II.9}$$

and

$$\begin{aligned} S(T_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}) = \exp\Big\{ - \int_0^{T^*_{ij}} \lambda_{0,T}(u) \exp\Big\{&(\beta_{Z,T} + \nu_{T,i}) Z_{ij} \\ &+ \beta'_T X^T_{ij} + \eta' h\left(M_{ij}\right) \Big\} \mathrm{d}u \Big\}. \end{aligned} \tag{II.10}$$

By assumption of the conditional independence of the repeated measurements of the biomarker of a subject given the random effetcs, the longitudinal part of the conditional likelihood of individual ij given the random effects is:

$$p_\phi\left(\tilde{M}_{ij} \mid \nu_i, \omega_{ij}\right) = \frac{1}{(2\pi\sigma^2)^{n_{ij}/2}} \exp\left\{ - \frac{\|\tilde{M}_{ij} - \theta'_{ij} f(t) + \beta'_M X^M_{ij} + (\beta_{Z,M} + \nu_{M,i}) Z_{ij}\|^2}{2\sigma^2} \right\}.$$

Finally the full log-likelihood is

$$\begin{aligned} l(\phi) = \sum_{i=1}^{K} \Big\{ \int_{\mathbb{R}^2} \Big[\prod_{j=1}^{n_i} \Big(\int_{\mathbb{R}^p} \Big(&p_\phi\left(T^*_{ij}, \delta_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}\right) \\ &\times \prod_{k=1}^{n_{ij}} p_\phi\left(\tilde{M}_{ijk} \mid \nu_i, \omega_{ij}\right) f_\omega(\omega_{ij}) \Big) \mathrm{d}\omega_{ij} \Big) \Big] f_\nu(\nu_i) \mathrm{d}\nu_i \Big\}. \end{aligned} \tag{II.11}$$

An estimation of ϕ can be carried-out by maximizing this log-likelihood,

$$\hat{\phi} = \mathrm{argmax}_\phi\, l(\phi),$$

using an optimization algorithm such as the Marquardt algorithm (MARQUARDT, 1963).

In pratice the basline hazard function $\lambda_0(\cdot)$ can be specified parametrically using a standard hazard function (like Weibull) or through more flexible M-splines (JOLY et al., 1998). In the

latter case, the estimated function $\hat{\lambda}_0(\cdot)$ can be highly irregular. As it is preferable to have a smooth function, a penalized likelihood approach can be used, with a penalization on the second-order derivative of $\lambda_{0,T}$. Therefore, the penalized log-likelihood to be maximized is

$$l_{\text{pen}}(\phi) = l(\phi) - \kappa \int_0^\infty \frac{\mathrm{d}^2 \lambda_{0,T}(u)}{\mathrm{d}u^2} \mathrm{d}u, \tag{II.12}$$

where κ is the smoothing parameter which will result in smoother estimation of $\lambda_{0,T}$ as it increases. This parameter has to be fixed before maximizing the penalized likelihood. As it is difficult to chose a particular value *a priori*, an optimal value of κ can be obtained through cross-validation.

One technical aspect in the computation of the likelihood concerns the computation of the two integrals over the trial-level and individual-level random effects. These integrals do not have a closed-form solution and therefore need to be numerically approximated. Such approximations can be carried out using MCMC techniques (the sampling distributions being multivariate normal distributions) or through gaussian quadrature rules. The issue with MCMC methods is that they often require many sampling points in order to achieve satisfaying accuracy. While gaussian quadratures requires less points and thus less computation for integral of low dimensions, the number of points increases exponentially with the dimension and can become burdensome. Moreover, gaussian quadratures may yield poor results if the nodes used in their approximation do not appropriately cover the shape and scale of the integrand. To overcome this difficulty, pseudo-adaptive gaussian-hermite quadrature has been proposed (Rizopoulos, 2012a). In this approach, a longitudinal model for the biomarker is first estimated from which bayesian predictions of the random-effects are then derived. These predictions are then used to recenter and rescale the integrand and achieve better numerical approximation. In our implementation, we used pseudo-adaptive gauss-hermite quadrature for approximating the integral over the individual-level random effects (ω_{ij}) and MCMC for the integral over the trial-level random effects (ν_i).

Moreover, the survival $S(T_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij})$ also involves an integral (over time). This integral has a closed-form solution only when the predictor $(\beta_{Z,T} + \nu_{T,i}) Z_{ij} + \beta_T' X_{ij}^T + \eta' h\left(M_{ij}\right)$ depends linearly on time. For more complex time-dependency this integral also needs to be approximated. In most implementations of joint model this integral approximation is performed

using a Gauss-Kronrod rule (LAURIE, 1997) which is also the method we used.

Finally, the covariance matrix of $\hat{\phi}$ can be obtained as the inverse of the Hessian matrix of $l_{\text{pen}}(\phi)$ evaluated at $\hat{\phi}$ from which inference on ϕ can be derived.

Surrogacy measures

In this Section we give derivations of the surrogacy measures used in the meta-analytic and mediation analysis approach (see Sections II.2 and II.3). For the meta-analytic approach, the trial-level surrogacy measure is straightforward and defined as the squared correlation between $\nu_{M,i}$ and $\nu_{T,i}$. Recall that,

$$\begin{pmatrix} \nu_{M,i} \\ \nu_{T,i} \end{pmatrix} \sim \mathcal{N}(0, \Sigma),$$

where Σ is a 2×2 matrix $\begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_{2,2} \end{pmatrix}$. Then the trial-level surrogacy measure R^2_{trial} is given by,

$$R^2_{\text{trial}} = \frac{\Sigma^2_{1,2}}{\Sigma_{1,1}\Sigma_{2,2}}. \tag{II.13}$$

The standard-error of $\hat{R}^2_{\text{trial}}$ can be derived from that of the maximum likelihood estimator, $\hat{\phi}$, using the delta-method. The individual-level association measure depends on the nature of the link function $h(\cdot)$. One could use the approach used by RENARD et al. (2003) and detailed in Section II.2.3. Another approach could be to used the Kendall's τ. The Kendall's τ is a measure of association between two variables X and Y and defined as follow: for two random draws (X_1, Y_1) and (X_2, Y_2),

$$\begin{aligned}
\tau &= \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0) \\
&= \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - (1 - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0)) \\
&= 2\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - 1 \\
&= 2\left(\mathbb{P}(X_1 > X_2, Y_1 > Y_2) + \mathbb{P}(X_1 < X_2, Y_1 < Y_2)\right),
\end{aligned}$$

when the couples (X_1, Y_1) and (X_2, Y_2) are independent draws from the same joint distribution, $\mathbb{P}(X_1 > X_2, Y_1 > Y_2) = \mathbb{P}(X_1 < X_2, Y_1 < Y_2)$ and therefore,

$$\tau = 4\mathbb{P}(X_1 > X_2, Y_1 > Y_2) - 1.$$

However in the case of a time-dependent variable like M, there is no natural order relation $M_1 > M_2$ between the trajectories. One way to circumvent this is to define a time-dependent τ, $\tau(t)$, on the variables T and $M(t)$:

$$\tau(t) = 4\mathbb{P}\left(M_1(t) > M_2(t), T_1 > T_2\right) - 1.$$

For the sake of simplicity, we present a derivation of $\tau(t)$ as "covariate-free" in which we do not consider the dependence on X and Z. Therefore, at a given time t, $M_i(t)$ is gaussian variable,

$$M_i(t) = (\beta + \omega_i)' f(t),$$

whose mean is $\beta' f(t)$ and variance $f(t)' D f(t)$. We have,

$$\begin{aligned}\mathbb{P}\left(M_1(t) > M_2(t), T_1 > T_2\right) &= \mathbb{P}\left(T_1 > T_2 \mid M_1(t) > M_2(t)\right) \mathbb{P}\left(M_1(t) > M_2(t)\right) \\ &= \frac{\mathbb{P}\left(T_1 > T_2 \mid M_1(t) > M_2(t)\right)}{2},\end{aligned}$$

since $\mathbb{P}\left(M_1(t) > M_2(t)\right) = \mathbb{P}(M_1(t) < M_2(t))$. The probability $\mathbb{P}\left(T_1 > T_2 \mid M_1(t) > M_2(t)\right)$ can be computed by conditionning and averaging over ω_1 and ω_2. Since the event $M_1(t) > M_2(t)$ is equivalent to $\omega_1' f(t) > \omega_2' f(t)$, we have:

$$\mathbb{P}\left(T_1 > T_2 \mid M_1(t) > M_2(t)\right) = \int_{\mathbb{R}} \left(\int_{x_2}^{\infty} \mathbb{P}\left(T_1 > T_2 \mid U_1 = u_1, U_2 = u_2\right) f_{U_1}(u_1) \mathrm{d}x_1 \right) f_{U_2}(u_2) \mathrm{d}u_2,$$

where for simplicity we note $U_i = \omega_i' f(t) \sim \mathcal{N}(0, f(t)' D f(t))$. The probability,

$$\mathbb{P}\left(T_1 > T_2 \mid U_1 = u_1, U_2 = u_2\right),$$

can be computed as,

$$\mathbb{P}\left(T_1 > T_2 \mid U_1 = u_1, U_2 = u_2\right) = \int_0^\infty \mathbb{P}\left(T_1 > t \mid X_1 = x_1\right) f_{T_2|X_2=x_2}(t)\mathrm{d}t,$$

where the quantities,

$$\begin{aligned}
\mathbb{P}\left(T_1 > t \mid X_1 = x_1\right) &= \exp\left(-\int_0^t \lambda_{0,T}(u)\exp(\eta' h(x_1))\mathrm{d}u\right) \\
f_{T_2|X_2=x_2}(t) &= \lambda_{0,T}(t)e^{\eta' h(x_2)}\exp\left(-\int_0^t \lambda_{0,T}(u)\exp(\eta' h(x_2))\mathrm{d}u\right),
\end{aligned}$$

can be derived from the model.

Unlike R^2_{trial}, the standard-error of $\hat{\tau}(t)$ cannot be estimated using the delta-method as is computation involves complex integration. One possibility is to use nonparametric bootstrap based on resampling of the original dataset, obtaining a new value of $\hat{\phi}_2$ and computing $\hat{\tau}(t)$ for $\hat{\phi}_2$. By doing so a sufficient number of times, one can approximate the sampling distribution of $\hat{\tau}(t)$ and derive an estimation of its standard-error and a quantile-based confidence interval. However since joint models rely on heavy computations they can take prolonged time to run which makes the use of nonparamatric bootstrap prohibitive. Another possibility is to use "parametric bootstrap" in which the first step is replaced by sampling $\hat{\phi}$ directly from its estimated distribution. Since ϕ is estimated via maximum likelihood its estimator is asymptotically gaussian. One can thus sample $\hat{\phi}_k$ from a normal $\mathcal{N}(\hat{\phi}, \hat{G})$, where $\hat{G}$ is the estimated covariance matrix of $\hat{\phi}$.

For the mediation analysis, a measure of surrogacy can be taken as the proportion of treatment effect PTE, as defined by the formula II.5. Since T is a time-to-event endpoint, it is more relevant to measure this effect in the survival scale rather than in terms of hazard-ratio as the latter may lack causal interpretability (HERNÁN, 2010). Therefore, let denote by $\mathbb{S}^{zz'}(t \mid x)$ the survival at time t of the counterfactual variable $T(z, M^{z'})$ given covariates x:

$$\mathbb{S}^{zz'}(t \mid x) = \mathbb{P}(T(z, M^{z'}) > t \mid X = x). \tag{II.14}$$

Given the assumptions introduced in Section II.3.2, the survival in II.14 can obtained from the

joint model as (Zheng and Liu, 2022):

$$\mathbb{S}^{zz'}(t \mid x) = \int_{\xi} S_T\left(t \mid h(M^{z'}), \xi, Z = z, X = x\right) f_{\xi}(\xi)\mathrm{d}\xi, \tag{II.15}$$

where ξ denotes the random effects (trial and individual levels) in the model and

$$S_T\left(t \mid h(M^{z'}), \xi, Z = z, X = x\right) = \exp\left(-\int_0^t \lambda_{0,T}(u)\exp((\beta_{Z,T} + \nu)z + \beta_T' x^T + \eta' h(M^{z'}))\mathrm{d}u\right).$$

From the "counterfactual survival functions" $\mathbb{S}^{zz'}(t \mid x)$ we can obtain the natural direct and indirect effect of the treatment on T through S as,

$$\begin{aligned}\mathrm{NDE}(t \mid x) &= \mathbb{S}^{10}(t \mid x) - \mathbb{S}^{00}(t \mid x);\\ \mathrm{NIE}(t \mid x) &= \mathbb{S}^{11}(t \mid x) - \mathbb{S}^{10}(t \mid x).\end{aligned}$$

Averaging over the distribution of X yieds the marginal effects,

$$\begin{aligned}\mathrm{NDE}(t) &= \int_{\mathcal{X}} \mathrm{NDE}(t \mid x) f_X(x)\mathrm{d}x\\ &= \int_{\mathcal{X}} \left(\mathbb{S}^{10}(t \mid x) - \mathbb{S}^{00}(t \mid x)\right) f_X(x)\mathrm{d}x;\end{aligned}$$

$$\begin{aligned}\mathrm{NIE}(t) &= \int_{\mathcal{X}} \mathrm{NIE}(t \mid x) f_X(x)\mathrm{d}x\\ &= \int_{\mathcal{X}} \left(\mathbb{S}^{11}(t \mid x) - \mathbb{S}^{10}(t \mid x)\right) f_X(x)\mathrm{d}x.\end{aligned}$$

The proportion of treatment effect can then be taken as a measure of surrogacy,

$$\begin{aligned}\mathrm{PTE}(t) &= \frac{\mathrm{NIE}(t)}{\mathrm{NIE}(t) + \mathrm{NDE}(t)}\\ &= \frac{\mathrm{NIE}(t)}{\mathrm{TTE}(t)}.\end{aligned} \tag{II.16}$$

The PTE is defined as a ratio of survival functions, and as such is not defined for $t = 0$ since no effect of the treatment can be observed at baseline and $\mathrm{TTE}(0) = 0$. But even for $t > 0$ it is not defined if there is no treatment effect. However it is not relevant to investigate any surrogacy at all for a treatment that has no effect on the final endpoint. Additionally, even if

it is commonly labelled as a "proportion", there is no guarantee that the PTE lies in the unit interval $[0, 1]$. For example if $\text{NIE} > 0$ and $\text{NDE} < 0$ then $\text{NIE} > \text{TTE}$ and thus $\text{PTE} > 1$. Therefore, even if the PTE is the common surrogacy measure used in mediation analysis, it has to be analyzed jointly with its associated direct and indirect effect.

From a mediation perspective it is important to note the role played by the link function $h(\cdot)$. In order to have a possible mediated effect, the argument of this function should depend on the treatment. For example in the case of shared random-effects, commonly used in joint modeling, no mediation is possible, by definition, since the random effects do not depend on the treatment. This can be seen in the Figure II.2(a). If the quantity in $h(M^Z)$ is independent of Z then there is not arrow from Z to $h(M^Z)$ and thus no mediated effect (even if $\eta' \neq 0$). Finally, like $\tau(t)$, the standard-errors for $\text{NIE}(t)$, $\text{NDE}(t)$ and $\text{PTE}(t)$ can be obtained via nonparametric or parametric bootstrap.

We now present the second proposed joint model in which the surrogate is no longer a longitudinal biomarker but is also itself a time-to-event.

II.4.2. Joint model for a time-to-event surrogate

We now return to the notation S and T for the surrogate and the final endpoint respectively. In the case when both S and T are censored event times, we proposed to adapt the joint frailty model proposed by SOFEU et al. (2019). As before, let i denotes the ith trial, $i = 1, \cdots, K$ and ij the jth patient from trial i. Let S_{ij}, T_{ij} and C_{ij} be the value of the surrogate endpoint, the final endpoint and the censoring time respectively. The observed outcomes are $S^*_{ij} = \min(S_{ij}, T_{ij}, C_{ij})$ and $T^*_{ij} = \min(T_{ij}, C_{ij})$. It is important to note that S_{ij} can be censored by the occurrence of the final endpoint T_{ij} but not the opposite. For example S_{ij} could be the time-to-progression and T_{ij} the time-to-death. The model proposed by SOFEU et al. (2019) is as follows:

$$\begin{cases} \lambda_{S_{ij}}(t) = \lambda_{0,S}(t) \exp\left(\omega_{ij} + u_i + (\nu_{S,i} + \beta_{Z,S})Z_{ij} + \beta_S^\intercal X_{ij}\right) \\ \lambda_{T_{ij}}(t) = \lambda_{0,T}(t) \exp\left(\zeta\omega_{ij} + \alpha u_i + (\nu_{T,i} + \beta_{Z,T})Z_{ij} + \beta_T^\intercal X_{ij}\right). \end{cases} \tag{II.17}$$

In this model ω_{ij} and u_i are individual-level and trial-level random effects such that $\omega_{ij} \sim \mathcal{N}(0, \theta^2)$ and $u_i \sim \mathcal{N}(0, \rho^2)$. These effects represent the association between S_{ij} and T_{ij} (ω_{ij}) and the heterogeneity of the patients across trials (u_i). The parameters α and ζ allow

these random effects to influence S_{ij} and T_{ij} differently. For example $\zeta > 0$ corresponds to a positive association between both endpoints. The random effects $\nu_{S,i}$ and $\nu_{T,i}$ are correlated random effects representing the heterogeneity of the treatment effects across trials as defined by Equation II.7. The surrogacy measures Sofeu et al. proposed are similar to those introduced in the case of a longitudinal surrogate. The trial-level is the same as defined in Equation II.13. The individual-level measure is the Kendall' τ,

$$\tau = \int_{\mathbb{R}^4} \left[\frac{e^{\omega_2+u_2} + e^{\zeta\omega_2+\alpha u_2}}{\left(e^{\omega_1+u_1} + e^{\omega_2+u_2}\right)\left(e^{\zeta\omega_1+\alpha u_1} + e^{\zeta\omega_2+\alpha u_2}\right)} \right] \frac{1}{4\pi^2\theta\rho} \exp\left(-\frac{1}{2}\left(\frac{\omega_1+\omega_2}{\theta^2} + \frac{u_1+u_2}{\rho^2}\right)\right) \mathrm{d}\omega_1 \mathrm{d}\omega_1 \mathrm{d}u_1 \mathrm{d}u_2. \tag{II.18}$$

Proof of the formula II.18 is given in Annex A.2. However from a mediation analysis viewpoint this model assumes a conditional independence between S_{ij} and T_{ij} given the random effects and therefore does not allow a mediated treatment effect on T_{ij} through S_{ij}. To overcome this issue, one possibility is to introduce a dependence of the risk of occurrence of T_{ij} on S_{ij}. However, in this semi-competing risks setting, the hazard function of T_{ij}, $\lambda_{T_{ij}}(t)$ cannot depend on the value of S_{ij} for $t < S_{ij}$ as this would artificially create a dependence on a future event. Therefore we proposed the following extended model:

$$\begin{cases} \lambda_{S_{ij}}(t) = \lambda_{0,S}(t) \exp\left(\omega_{ij} + u_i + (\nu_{S,i} + \beta_{Z,S})Z_{ij} + \beta_S^\intercal X_{ij}\right) \\ \lambda_{T_{ij}}(t) = \lambda_{0,T}(t) \exp\Big(\zeta\omega_{ij} + \alpha u_i + (\nu_{T,i} + \beta_{Z,T})Z_{ij} + \beta_T^\intercal X_{ij} \\ \qquad\qquad +\gamma(S_{ij})\mathbb{1}_{(S_{ij}\leq t)}\Big). \end{cases} \tag{II.19}$$

The difference between model II.17 and II.19 is the introdution of the term $\gamma(S_{ij})\mathbb{1}_{(S_{ij}\leq t)}$ in the latter. This represents the modification of the risk of occurrence of T_{ij} at the occurrence of S_{ij} (through the indicator $\mathbb{1}_{(S_{ij}\leq t)}$). The magnitude of this effect is given by the function $\gamma(S_{ij})$ which is estimated using B-splines. A positive function will result in an increase of the hazard of T_{ij} at the occurrence of S_{ij} and a negative function in a reduced risk. The absence of direct link between S_{ij} and T_{ij} is equivalent to $\gamma(S_{ij}) = 0$. The different parameters of model II.19 are represented in the DAG in Figure II.2(b). As in the case of a longitudinal biomarker, with

this extended model the natural direct and indirect effect are obtained as,

$$\begin{aligned}\mathrm{NIE}(t) = \int_{\mathcal{X}} \int_0^\infty \int_{\mathbb{R}^4} & S_T(t \mid S = s, Z = 1, X = s, \xi) \Big[f_S(s \mid Z = 1, X = x, \xi) \\ & - f_S(s \mid Z = 0, X = x, \xi) \Big] \mathrm{d}\xi \mathrm{d}s \mathrm{d}x; \\ \mathrm{NDE}(t) = \int_{\mathcal{X}} \int_0^\infty \int_{\mathbb{R}^4} & \Big[S_T(t \mid S = s, Z = 1, X = s, \xi) - S_T(t \mid S = s, Z = 0, X = s, \xi) \Big] \\ & \times f_S(s \mid Z = 0, X = x, \xi) \mathrm{d}\xi \mathrm{d}s \mathrm{d}x,\end{aligned}$$

from which the PTE can be derived as in Equation II.16.

III

Time-to-event surrogate validation using mediation analysis and meta-analytic data

III.1 Introduction 67
III.2 Main article 68
III.3 Addtional remarks 104

III.1. | Introduction

Motivated by the will to tie both the meta-analytic and mediation analysis approaches for surrogate validation, we developed a joint model for the case where the surrogate and the final endpoint are (possibly censored) time-to-event. Specifically, we assume a semi-competing risk setting where the occurrence of the final endpoint precludes the observation of the surrogate. This setting is of particular relevance as many fields of clinical research, especially in oncology, focus on time-to-event endpoints. In this work we propose an extension of the joint frailty model developed by Sofeu et al. (2019) to enable and quantify a direct link between the surrogate and the final endpoint, and therefore enable a potential mediated effect. The development of this model is articulated within the frameworks of causal mediation analysis and potential outcomes. This development proceeds in two main stages. The first stage is the specification and estimation of the joint model for the (counterfactual) surrogate and final endpoint, and the second stage is the derivation of the quantities related to the mediation analysis based on the estimated parameters of the model. Identifiabilty assumptions are made in order to derive the estimated quantities based on the observed outcomes in the counterfactual world. The proposed joint model uses random effects at the individual and trial levels to take into account the association of the two endpoints for an individual and the intra-correlation of the treatment

effects within a trial. The use of random effects is also of interest from a mediation analysis viewpoint. Indeed, one indentifiability assumption (the sequential ignorability) requires that there is no unmeasured confounder, or equivalently that the (observed) covariate of a subject must capture all the confounding between the two endpoint. This is a particularly important and stringent assumption, but the use of random effects allows us to relax this assumption a little: now instead of requiring that the confounding must be taken into account solely by the covariates, it must be taken into account by either the covariates, the random effects or both. Therefore, the random effects can be used to capture the confounding that is potentially not taken into account by the covariates. A simulation study was conducted to investigate the performances of the proposed approach regarding both the estimations of the parameters as well as the mediation and surrogacy measures. We applied this method to a meta-analysis of randomized clinical trials in gastric cancer to study the time-to-relapse as a surrogate of the overall survival. We also conducted a second application on data from a multi-centric clinical trial on second-line treated metastatic colorectal cancer to study the time-to-progression as surrogate of the overall survival. This article has been published in *Biostatistics* (LE COËNT et al., 2022)

III.2. | MAIN ARTICLE

Biostatistics (2022) **00**, 00, *pp.* 1–19
https://doi.org/10.1093/biostatistics/kxac044

C

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

Time-to-event surrogate endpoint validation using mediation analysis and meta-analytic data

QUENTIN LE COËNT*

Department of Biostatistics, Bordeaux Population Health Research Center, INSERM U1219, 146 rue Léo Saignat, 33076 Bordeaux, France

quentin.le-coent@u-bordeaux.fr

CATHERINE LEGRAND

ISBA/LIDAM, UCLouvain, 20 Voie du Roman Pays, B-1348 Louvain-la-Neuve, Belgium

VIRGINIE RONDEAU

Department of Biostatistics, Bordeaux Population Health Research Center, INSERM U1219, 146 rue Léo Saignat, 33076 Bordeaux, France

Summary

With the ongoing development of treatments and the resulting increase in survival in oncology, clinical trials based on endpoints such as overall survival may require long follow-up periods to observe sufficient events and ensure adequate statistical power. This increase in follow-up time may compromise the feasibility of the study. The use of surrogate endpoints instead of final endpoints may be attractive for these studies. However, before a surrogate can be used in a clinical trial, it must be statistically validated. In this article, we propose an approach to validate surrogates when both the surrogate and final endpoints are censored event times. This approach is developed for meta-analytic data and uses a mediation analysis to decompose the total effect of the treatment on the final endpoint as a direct effect and an indirect effect through the surrogate. The meta-analytic nature of the data is accounted for in a joint model with random effects at the trial level. The proportion of the indirect effect over the total effect of the treatment on the final endpoint can be computed from the parameters of the model and used as a measure of surrogacy. We applied this method to investigate time-to-relapse as a surrogate endpoint for overall survival in resectable gastric cancer.

Keywords: Joint modeling; Mediation analysis; Meta-analysis; Surrogacy; Time-to-event.

1. Introduction

Given the continued development of new and more effective therapies in oncology, the use of standard endpoints in clinical trials may affect the feasibility of these trials. For example, a significant increase in overall survival requires long-term follow-up of patients to ensure adequate statistical power and increases

*To whom correspondence should be addressed.

© The Author 2022. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

the duration and cost of the trial. The use of surrogate endpoints is therefore an important approach and a possible way to enable such otherwise unfeasible studies.

These surrogates or intermediate endpoints are faster to measure, potentially more frequent, and less costly to obtain. They would shorten the duration of the trial and allow new treatments to be available to patients more quickly. However, to maintain the validity of the clinical trial with respect to the final endpoint, statistical conclusions about the efficacy of the treatment on the surrogate endpoint should be extendable to the final endpoint. Consequently, putative surrogate endpoints should be statistically validated as good surrogates for the final endpoint before being used in an upcoming trial. For example, time-to-progression, defined as the duration between randomization and disease progression, is a commonly used endpoint in phase II/III oncology clinical trials. However, this endpoint has been shown to be a poor surrogate for overall survival in some cases (Burzykowski *and others*, 2008), while it has been shown to be promising in others (Lee *and others*, 2016). Therefore, it is of interest to use surrogate endpoints that have been properly validated as such.

Several approaches have been proposed to investigate the quality of an endpoint as a surrogate. First, Prentice (1989) defined a surrogate as a variable for which the effect of the treatment on the final endpoint is null when conditioned upon. Freedman *and others* (1992) proposed a measure of the proportion of treatment explained by the surrogate. Frangakis and Rubin (2002) used the causal inference framework to study a potential surrogate using principal stratification. An approach using meta-analytic data was proposed by Daniels and Hughes (1997) and Buyse *and others* (2000). Joffe and Greene (2009) provided an overview of existing methods and divided them into two categories: methods based on "effects" or on "associations."

Currently, methods based on data from meta-analyses are widely used and are the most common methods for analyzing potential surrogate endpoints (Buyse *and others*, 2016). The purpose of these methods is to study the association between a surrogate endpoint and a final endpoint at two different levels. Firstly, the association between the surrogate endpoint and the final endpoint is statistically quantified at the individual level. Secondly, the association between the treatment effect on the surrogate endpoint and the treatment effect on the final endpoint can be determined at the study level using meta-analytic data from multiple trials. A surrogate endpoint is considered to be of interest if the association is strong at both levels. The definition of these associations depends on the type of the endpoints, and these meta-analytic approaches have been developed for different types of surrogates and endpoints. An overview of the different proposed methods can be found in Buyse and Molenberghs (2005). However, because a surrogate is always a post-treatment variable, it may be confounded by post-treatment factors. In such cases, potential associations between surrogates and the final outcomes that are highlighted may not have a causal interpretation.

Causal inference methods, including mediation analysis, have therefore been proposed for validating surrogate endpoints (Zigler and Belin, 2012; Conlon *and others*, 2014; Vandenberghe *and others*, 2018). In this article, we focus on mediation analysis whose purpose is to decompose the total effect of the treatment on a (final) endpoint into a direct effect and an indirect effect through a mediator (here, the surrogate). The indirect effect of the treatment on the final endpoint through the surrogate can itself be decomposed into two parts: a direct effect of the treatment on the mediator and a direct effect of the mediator on the final endpoint. Recently, Zheng and Liu (2021) have proposed an approach to study the amount of treatment effect on a time-to-event outcome going through a longitudinal biomarker using a joint modeling approach. Our purpose is to consider the situation in which both the final endpoint and the surrogate are censored time-to-event and to extend the mediation analysis approach to account for meta-analytic data. A meta-analytic setting introduces heterogeneity into the data and strengthens the validation process of a surrogate compared with validation based on data from a single study.

This article is organized as follows. In Section 2, we introduce the causal and mediation analysis framework used to define our measure of surrogacy. We also define a joint model for time-to-event that

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

accounts for the meta-analytic nature of the data through trial-level random effects. In Section 3, we present a simulation study to investigate the finite sample-size performances of this method regarding the estimation of the parameters of the model as well as the measure of surrogacy. In Section 4, we apply this method to a meta-analysis of adjuvant chemotherapy for resectable gastric cancer to investigate time-to-relapse as a surrogate for overall survival. Finally, in Section 5, we conclude and provide some comments and further developments on the proposed methodology.

2. Mediation analysis and joint survival model

2.1. *Counterfactual framework for mediation analysis*

To give a causal interpretation of the "direct," "indirect," and "total" effects, we rely on the counterfactual results framework (Rubin, 1974; Holland, 1986). Based on this causal framework, Robins and Greenland (1992) gave conditions under which the identification of a direct and indirect effect of treatment by a post-treatment variable is possible, thus linking mediation analysis to the framework of causal inference. Much of the development of mediation analysis in the context of causal inference relies on the framework of potential outcomes, most commonly using the notion of nested counterfactual outcomes. Counterfactual mediation analysis for time-to-event outcomes was explored in Lin *and others* (2017) and in Tchetgen Tchetgen (2011).

Defining an indirect treatment effect in a setting where the intermediate outcome is a time-to-event that can be censored by the final event is challenging (Huang, 2020; Didelez, 2019). In this setting of semicompeting risks, Huang (2020) used counterfactual outcomes in a counting process to derive a nonparametric estimator of the indirect effect. In this article, we chose to use a flexible parametric model to estimate the direct and indirect effect of treatment on the final endpoint.

In the context of a meta-analysis, let ij denote the jth patient from the ith trial. Let Z_{ij} be a binary variable indicating whether the patient is in the experimental or control arm, X_{ij} a vector of baseline covariates, S_{ij} the uncensored surrogate (the mediator), and T_{ij} the uncensored final outcome. We assume an independent censoring C_{ij}. The observed censored outcomes are $S_{ij}^{*} = \min(S_{ij}, C_{ij}, T_{ij})$ and $T_{ij}^{*} = \min(T_{ij}, C_{ij})$ with their associated censoring indicators $d_{ij} = I_{S_{ij} \leq T_{ij}^{*}}$ and $\delta_{ij} = I_{T_{ij} \leq C_{ij}}$.

We define the counterfactual outcome $S_{ij}(z)$ as the counterfactual surrogate that would have been observed if a patient had been treated with treatment $Z_{ij} = z$. Furthermore, we define the counterfactual outcome $T_{ij}(z, s)$ as the final endpoint that would have been observed if treatment had been set to z and the surrogate had occurred at time s. We assume that for all $t < s$, the probability $\mathbb{P}\left(T_{ij}(z, s) \geq t\right)$ is independent of s, i.e., the occurrence of the surrogate at time s does not affect the distribution of $T_{ij}(z, s)$ before s. The nested counterfactual outcome $T_{ij}(z, S_{ij}(z'))$ is defined as the final endpoint that would have been observed if the treatment for the final endpoint had been set to z but the surrogate had "behaved" as if the treatment had been set at z'.

In the setting of semicompeting risks, such nested counterfactuals may be ill-defined, for example, if survival under z' is shorter than under z (Didelez, 2019). In this case, $T_{ij}(z, S_{ij}(z'))$ might not be properly defined if the support of $S_{ij}(z')$ is constrained by $T_{ij}(z', S_{ij}(z'))$.

We assume that:

1. $S_{ij}(z')$ can be censored by $T_{ij}(z, S_{ij}(z'))$
2. If $S_{ij}(z')$ is censored by $T_{ij}(z, S_{ij}(z'))$ it cannot have an effect on the occurrence of $T_{ij}(z, S_{ij}(z'))$
3. Thus $S_{ij}(z')$ can only have an effect on $T_{ij}(z, S_{ij}(z'))$ if it occurs first.

In the following, we will mainly focus on the hazard functions of $S_{ij}(z')$ and $T_{ij}(z, S_{ij}(z'))$.

Under the assumptions defined below, the observed outcome T_{ij} of a patient receiving treatment $Z_{ij} = z$ is equal to the counterfactual outcome $T_{ij}(z, S_{ij}(z))$. From these counterfactual outcomes, we can derive

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

a definition of the natural direct and indirect effects of treatment on T_{ij}. To identify the distributions of these counterfactual outcomes from the observed outcomes S_{ij} and T_{ij}, we must rely on a set of hypotheses (Imai *and others*, 2010).

In this counterfactual framework, we rely on the SUTVA (stable unit treatment value assumption). The SUTVA first assumes that each patient has only one outcome under each version of the treatment. This is plausible in settings where each version of the treatment is uniquely defined. The SUTVA also assumes that one's treatment assignment does not affect the outcome of another. This is also plausible in the setting of randomized trials except for some particular cases like in infectious diseases. In addition to the SUTVA, we must also rely on assumptions that are required in the context of mediation analysis.

First, we make the consistency assumption:

(A1) If a patient receives treatment $Z_{ij} = z$, then $S_{ij} = S_{ij}(z)$ and $T_{ij} = T_{ij}(z, S_{ij}(z))$.

This assumption states that a subject's observed outcomes S_{ij} and T_{ij} are the counterfactual outcomes associated with the treatment that the subject received.

We also make two positivity assumptions: let $f_{S_{ij}}(\cdot)$ be the density of S_{ij}, then we assume that,

(A2) $\forall z \in \{0, 1\}, \mathbb{P}(Z_{ij} = z \mid X_{ij}) > 0$
(A3) $\forall z, x, f_{S_{ij}}(s \mid Z_{ij} = z, X_{ij} = x) > 0, \forall s \geq 0$

Assumption (A2) states that there is no level of X that fully determines whether a patient is going to be treated or not. This assumption holds if Z is randomized. Assumption (A3) states that there is no level of (Z_{ij}, X_{ij}) for which a time of occurrence of S_{ij} would be impossible to observe: the support of the distribution of $S_{ij} \mid Z_{ij}, X_{ij}$ remains the same for all values of (Z_{ij}, X_{ij}).

Finally, the third assumption is the sequential ignorability assumption, which is also composed of two assumptions:

(A4) $\big(T_{ij}(z', s), S_{ij}(z)\big) \perp\!\!\!\perp Z_{ij} \mid X_{ij}, \forall z, z', s$
(A5) $T_{ij}(z', s) \perp\!\!\!\perp S_{ij}(z) \mid (Z_{ij} = z, X_{ij}), \forall z, z', s.$

Assumption A4 states that for patients in the same stratum of X_{ij}, the potential outcomes $\big(T_{ij}(z', s), S_{ij}(z)\big)$ are the same (in distribution) whether these patients are treated or not. This assumption usually holds when the treatment assignment is random. Assumption (A5) states that the potential outcomes $T_{ij}(z', s)$ are independent of $S_{ij}(z)$ for patients sharing the same baseline characteristics and receiving the same treatment $Z_{ij} = z$. This assumption holds under the assumption that there are no unmeasured confounders along the path from S_{ij} to T_{ij}, which is often a strong assumption even for randomized trials. However, as will be discussed in more details in the next section, the proposed approach relies on a parametric model that requires a less stringent assumption.

In the setting of mediation analysis, one must also rely on the so-called "cross-world" assumption (Andrews and Didelez, 2020) which in our case would be translated as follows:

(A6) $T_{ij}(z', s) \perp\!\!\!\perp S_{ij}(z), \forall s, z \neq z'.$

This is a strong assumption but we will see below that it can hold given the proposed statistical model.

We now define, for two treatment levels z and z' and for $t \geq 0$, the counterfactual survival function $\mathbb{S}_{z'z}(t)$ associated with the counterfactual outcome $T_{ij}(z', S_{ij}(z))$: $\mathbb{S}_{z'z}(t) = \mathbb{P}\left(T_{ij}(z', S_{ij}(z)) > t\right)$. From this function, the *natural indirect effect* (NIE) of Z_{ij} on T_{ij} through S_{ij} can be defined as $\text{NIE}(t) = \mathbb{S}_{11}(t) - \mathbb{S}_{10}(t)$.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

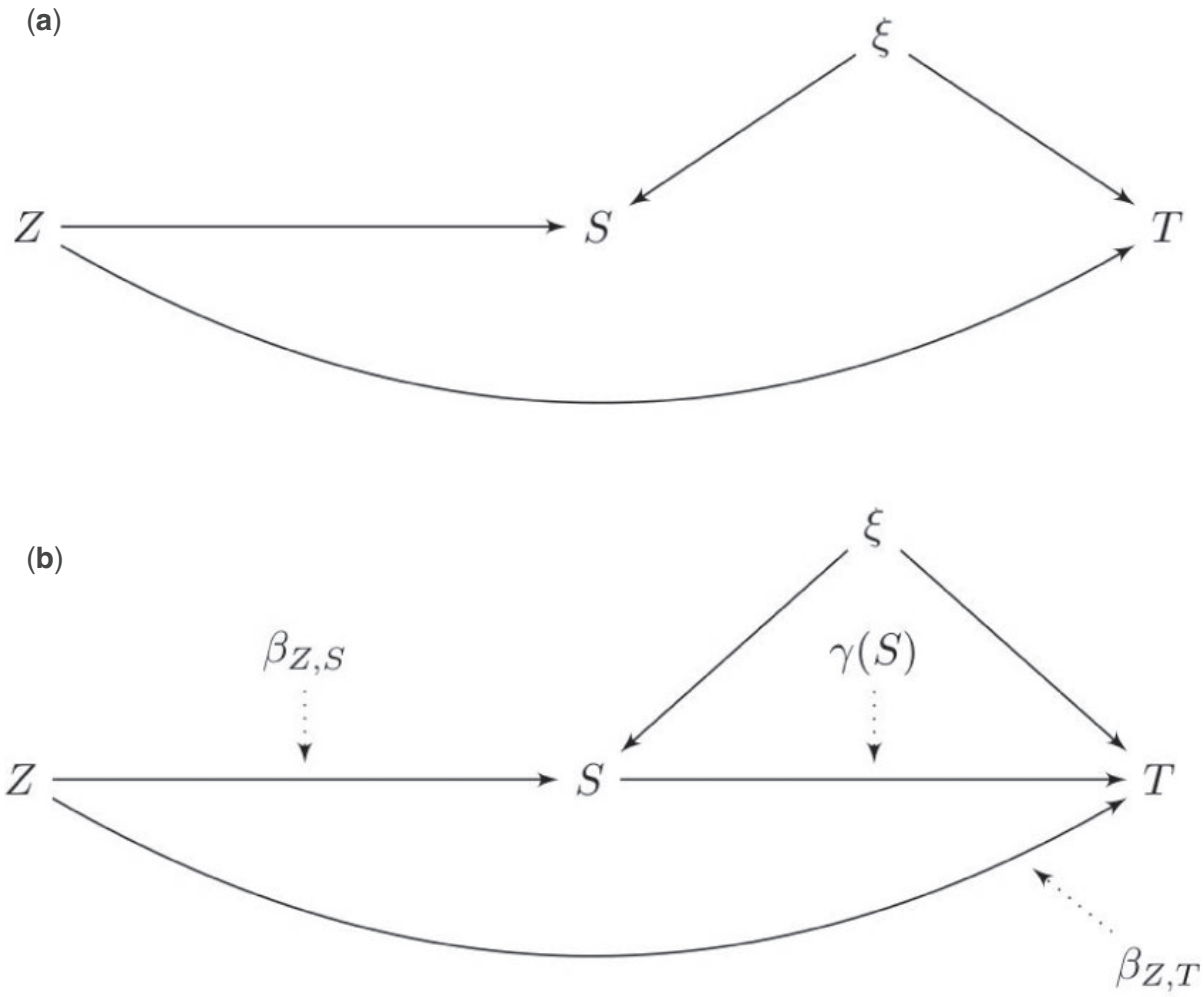


Fig. 1. (a) Joint model by Sofeu *and others* (2019): given the random effects (ξ), S and T are independent, and thus there is conditionally no direct effect of Z through S. (b) A generalization of (a) where S and T are not conditionally independent upon ξ but a direct link from S to T remains through the function $\gamma(S)$ as in (2.2).

The quantity $\mathrm{NIE}(t)$ is the difference between the survival of a subject with treatment Z_{ij} set to $Z_{ij} = 1$ for both T_{ij} and S_{ij} and the survival of a subject with treatment set to $Z_{ij} = 1$ for T_{ij} but set to $Z_{ij} = 0$ for S_{ij}.

In the same way, the *natural direct effect* (NDE) is defined as $\mathrm{NDE}(t) = \mathbb{S}_{10}(t) - \mathbb{S}_{00}(t)$ and the total effect (TE) as $\mathrm{TE}(t) = \mathbb{S}_{11}(t) - \mathbb{S}_{00}(t)$. The proportion of treatment effect (PTE) is then defined as:

$$\mathrm{PTE}(t) = \frac{\mathbb{S}_{11}(t) - \mathbb{S}_{10}(t)}{\mathbb{S}_{11}(t) - \mathbb{S}_{00}(t)} = \frac{\mathrm{NIE}(t)}{\mathrm{TE}(t)}. \tag{2.1}$$

The quantity $\mathrm{PTE}(t)$ represents the proportion of the treatment effect on T_{ij} that is due to the effect of the treatment on S_{ij}. A high value for $\mathrm{PTE}(t)$ means that a large fraction of the treatment effect on T is due to its effect on S_{ij}, making S_{ij} an interesting surrogate.

In the following section, we present a joint model for two time-to-event endpoints that also accounts for meta-analytic data, and we show how the functions $\mathbb{S}_{z'z}(t)$ (and hence $\mathrm{PTE}(t)$) can be computed from this model.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

2.2. *Frailty model for two time-to-event endpoints*

As in the previous section, T_{ij} and S_{ij} are the uncensored final and surrogate endpoint, respectively, and X_{ij} is a vector of baseline covariates. The joint distribution of (T_{ij}, S_{ij}) must be carefully described because the occurrence of T_{ij} may preclude the occurrence (or observation) of S_{ij}. Moreover, these endpoints are not independent at the individual level: subjects who are susceptible to shorter surrogate times S_{ij} may also be susceptible to shorter survival times T_{ij}. That is, some subjects are "more frail" than others, and a natural association between (T_{ij}, S_{ij}) at the individual level should be considered. Furthermore, we assume a meta-analytic setting in which the outcomes of subjects from the same cluster (a clinical center or a trial) are also associated. Therefore, two types of associations must be considered: at the individual and trial level. These two associations can be modeled with a joint model that uses shared random effects at each level. Sofeu *and others* (2019) have proposed such a model for validating surrogate endpoints using meta-analytic data. However, their model assumes independence between S_{ij} and T_{ij} conditionally on the vector of random effects ξ_{ij} and therefore no direct effect of S_{ij} on T_{ij} given ξ_{ij} is possible (see Figure 1(a)). For this model, $\text{NIE}(t \mid \xi_{ij}) = 0$ and thus $\text{PTE}(t \mid \xi_{ij}) = 0$. To allow a direct effect of S_{ij} on T_{ij} and to have a more comprehensive modeling, we decided to use the joint model given by Equation (2.2) below. In this model, risks of the final and surrogate endpoint are modeled assuming proportional hazards. A direct effect of S_{ij} on T_{ij} is enabled by a modification of the hazard function of T_{ij} once the surrogate endpoint S_{ij} occurs. This modification is accounted for by a function $\gamma(S_{ij})$ that depends on S_{ij}, see Figure 1(b). For an individual j from trial i, this model is defined as follows:

$$\begin{cases} \lambda_{S_{ij}(z)}(t) = \lambda_{0,S}(t) \exp\left(\omega_{ij} + u_i + z(\nu_{S,i} + \beta_{Z,S}) + \beta_S^{\mathsf{T}} X_{ij}\right) \\ \lambda_{T_{ij}(z,s)}(t) = \lambda_{0,T}(t) \exp\left(\zeta\omega_{ij} + \alpha u_i + z(\nu_{T,i} + \beta_{Z,T}) + \beta_T^{\mathsf{T}} X_{ij} + \gamma(s) I(s \le t)\right). \end{cases} \tag{2.2}$$

Here, ω_{ij} is an individual-level frailty, u_i a trial-level frailty representing the potential heterogeneity between patients across trials and $(\nu_{S,i}, \nu_{T,i})$ trial-level frailties allowing for a potential heterogeneity between treatment effects across trials. These frailties are assumed to be jointly Gaussian,

$$\begin{pmatrix} \omega_{ij} \\ u_i \\ \nu_{S,i} \\ \nu_{T,i} \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma) \quad \text{where} \quad \mu = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \theta^2 & 0 & 0 & 0 \\ 0 & \rho^2 & 0 & 0 \\ 0 & 0 & \sigma^2_{\nu_S} & \sigma_{\nu_{S,T}} \\ 0 & 0 & \sigma_{\nu_{S,T}} & \sigma^2_{\nu_T} \end{pmatrix}. \tag{2.3}$$

The parameters (ζ, α) are power parameters that allow the frailties to affect hazard functions differently. The model parameters are estimated by maximizing a penalized likelihood,

$$L_{\text{pen}}(\phi) = L(\phi) - \kappa_T \int_0^\infty \left(\lambda''_{0,T}(t)\right)^2 \mathrm{d}t - \kappa_S \int_0^\infty \left(\lambda''_{0,S}(t)\right)^2 \mathrm{d}t, \tag{2.4}$$

where ϕ is the vector of all parameters of the model and $L(\phi)$ is the likelihood of the data. The parameters κ_T and κ_S are chosen by cross-validation and penalize the likelihood by the integral of the second derivative of the baseline hazards to ensure smooth estimates of these functions. Details on likelihood construction and estimation can be found in Appendix A of the Supplementary material at *Biostatistics* online.

In model given by Equation (2.2), a direct effect of $S_{ij}(z')$ on $T_{ij}(z, s)$ is introduced through the term $\gamma(s) I(s \le t)$. The indicator $I(s \le t)$ ensure that the risk of occurrence of $T_{ij}(z, s)$ only depends of s from the time of occurrence s.

The strength of an occurrence of $S_{ij}(z')$ on the risk of occurrence of $T_{ij}(z, s)$ is taken into account through the function $\gamma(s)$ with $\gamma(s) = 0$ indicating no effect.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

The parameters $\beta_{Z,S}$ and $\beta_{Z,T}$ represents the direct effect of the treatment on $S_{ij}(z')$ and $T_{ij}(z,s)$, respectively. To have an indirect effect of the treatment on $T_{ij}(z,S_{ij}(z'))$ through $S_{ij}(z')$, both $\beta_{Z,S}$ and the function $\gamma(s)$ should be different from 0. Figure 5 in Appendix F of the Supplementary material at *Biostatistics* online shows an example of how the occurrence of $S_{ij}(z')$ can change the risk of occurrence of $T_{ij}(z,S_{ij}(z'))$.

Following model (2.2) and assuming that $T_{ij} = T_{ij}(z,S_{ij}(z))$ and $S_{ij} = S_{ij}(z)$, we have $T_{ij}(1,s) \perp\!\!\!\perp S_{ij}(0) \mid (X_{ij},\xi_{ij})$, where $\xi = (\omega,u,\nu_S,\nu_T)$ is the vector of random effects. Thus, the cross-world assumption that the counterfactual $T(z,s)$ is independent of $S(z')$ holds conditionally on X and ξ.

As pointed out in the previous section, this model requires a less stringent assumption than assumption (A5), since it does not require that (Z_{ij},X_{ij}) capture all the confounding between $S_{ij}(z')$ and $T_{ij}(z,S_{ij}(z'))$, but rather that this confounding should be captured by (Z_{ij},X_{ij},ξ_{ij}). Thus, we make a somewhat less stringent assumption:

(A5') $T_{ij}(z',s) \perp\!\!\!\perp S_{ij}(z) \mid (Z_{ij}=z,X_{ij},\xi_{ij}), \forall z,z',s$

Similarly, conditioning on X_{ij} in assumptions (A2)–(A4) is replaced in this model by a conditioning on (X_{ij},ξ_{ij}).

From (2.2), we can derive the hazard function for the observed data as:

$$\lambda_{S_{ij}}\left(t \mid \xi_{ij},Z_{ij},X_{ij}\right) = \lambda_{0,S}(t)\exp\left(\omega_{ij}+u_i+Z_{ij}(\nu_{S,i}+\beta_{Z,S})+\beta_S^{\mathsf{T}}X_{ij}\right)$$

and

$$\lambda_T\left(t \mid \xi_{ij},Z_{ij},S_{ij},X_{ij}\right) = \lambda_{0,T}(t)\exp\left(\zeta\omega_{ij}+\alpha u_i+Z_{ij}(\nu_{T,i}+\beta_{Z,T})+\beta_T^{\mathsf{T}}X_{ij}+\gamma(S_{ij})I\left(S_{ij}\le t\right)\right)$$

Under the assumptions (A1)–(A5') and (A6), we can write the "counterfactual" survival function $\mathbb{S}_{z'z}(t)$ as:

$$\mathbb{S}_{z'z}(t) = \int_{\mathcal{X}}\int_0^{\infty}\int_{\mathbb{R}^4}\mathcal{S}_T(t \mid s,z',\xi,x)f_S(s \mid z,\xi,x)f(\xi)f_X(x)\mathrm{d}s\mathrm{d}\xi\,\mathrm{d}x, \tag{2.5}$$

where the function $\mathcal{S}_T(t \mid s,z,\xi,x)$ is the conditional survival function of T obtained from $\lambda_T(t \mid \xi,z,s,x)$ and $f_S(s \mid z',\xi,x)$ is the conditional density function of S derived from $\lambda_S(t \mid \xi,z,x)$. The sets $\mathcal{X}$ and $\mathbb{R}^4$ are the supports of the covariates X and ξ, respectively. From the computation of $\mathbb{S}_{z'z}(t)$, we can derive $\mathrm{PTE}(t)$ according to (2.1) by considering appropriate values for z and z'.

A confidence interval for $\mathrm{PTE}(t)$ can be computed using either nonparametric or parametric bootstrap. The latter proceeds by taking the (estimated) covariance matrix of the maximum likelihood estimator (MLE) $\hat{\phi}$, $\hat{\Sigma}_{\hat{\phi}}$. Using the asymptotic normality of the MLE, we can generate a large number of vector of parameters $\phi_i \sim \mathcal{N}\left(0,\hat{\Sigma}_{\hat{\phi}}\right)$. For each simulated ϕ, we compute a function $\mathrm{PTE}_i(t)$. The empirical standard deviation of the vector $(\mathrm{PTE}_i(t))$ can then be used to computed a confidence interval for $\mathrm{PTE}(t)$ as $\mathrm{PTE}(t) \pm 1.96 \times \mathrm{se}$.

Furthermore, as described in Sofeu *and others* (2019), this model can be used to calculate the coefficient of determination of the trial-level random effects, R^2_{trial}, defined as $\frac{\sigma^2_{\nu_{S,T}}}{\sigma^2_{\nu_S}\sigma^2_{\nu_T}}$ and the Kendall's τ which measures the individual-level association between S and T.

3. Simulations

We conducted a simulation study to assess the performance of this model and the associated estimation procedure. We simulated scenarios with different numbers of trials, different variances for the random

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

effects ω and u and scenarios in which there was an indirect treatment effect on the final endpoint through the surrogate and scenarios in which there was no indirect effect.

3.1. *Simulation scheme*

Observations were generated from (2.2), where T was generated using a piecewise-exponential distribution.

We first generated the frailties according to (2.3) and the surrogate S according to the model (2.2) with a Weibull baseline hazard function for $\lambda_{0,S}(t)$. Then, a piecewise exponential model is applied to T, which consists of two parts. First, on $[0, S]$, the hazard of T conditionally on the random effects is given as in (2.2) with a constant baseline hazard. Then, on $[S, \infty[$, the hazard is simply that of the first interval multiplied by $e^{\gamma(S)}$. The procedure for generating T is detailed in Appendix B of the Supplementary material at *Biostatistics* online. After generating S and T, we applied a constant censoring and retrieved the associated censored follow-up times and event indicators. This resulted in a censoring rate of about 40% for S and T. The simulation scheme is fully detailed in Appendix C of the Supplementary material at *Biostatistics* online.

Several scenarios were considered for which the number of trials, K, was set to 20 or 50 and the function $\gamma(s)$ was equal to 0 or $\gamma(s) = e^{-cs^2}$ for $c = 0.001$. For the case where $\gamma(s) = 0$, there is no direct effect of S on T and thus no indirect effect of Z on T through S. In this case, by definition, we have $\text{PTE}(t) = 0$. The case where $\gamma(s) = e^{-cs^2}$ defines a small mediation where $\text{PTE}(t)$ is approximately equal to 0.15 at maximum follow-up. This function was chosen to yield a positive direct effect of S on T that smoothly decreases towards 0 over the time scale of the simulations.

We also examined scenarios with different values for the variances θ^2 and ρ^2 of the random effects ω_{ij} and u_i respectively. A first scenario with high variances where $\theta^2 = \rho^2 = 1$ and a second with lower variances where $\theta^2 = \rho^2 = 0.64$. Patients were randomly assigned to one of the Kth trials with equal probability K^{-1}. All patients in the same trial had the same frailties $(u_i, \nu_{S,i}, \nu_{T,i})$ to create a cluster effect across trials. For all these scenarios, the total number of patients was set at 1500. Each scenario was repeated 300 times.

To determine the true value of $\text{PTE}(t)$ for each scenario, we simulated data sets in which treatment was fixed for each endpoint. In this way, we were able to use a nonparametric estimate of the values of the survival functions $\mathbb{S}_{z'z}(t)$ associated with the data set in which the treatment values for S and T were fixed at z and z', respectively. We evaluated $\text{PTE}(t)$ for t ranging between 10 and 45 with a step of 1. These calculations were used in the results to compute the true value of $\text{PTE}(t)$ at three time points (20, 30 and 40) and in the graphical representation of the estimated $\text{PTE}(t)$ versus to the true $\text{PTE}(t)$.

3.2. *Results*

Estimations were evaluated for each parameter of the model in terms of bias, mean standard error, empirical standard deviation, and 95% coverage rate. For $\text{PTE}(t)$, we focused on three specific time points (20, 30, and 40). We also evaluated the estimations of the NIE and NDE at these time points as well as estimations of R^2_{trial} and the Kendall's τ.

Simulations results are shown in Table 1 for the cases with $\theta^2 = \rho^2 = 1$ and in Table 2 for the cases with $\theta^2 = \rho^2 = 0.64$. In Figure 2, we plotted the estimated $\text{PTE}(t)$ from each simulated data set with their average at each time compared to the true value of the function $\text{PTE}(t)$.

From Tables 1 and 2, it can be seen that increasing the number of trials improves the estimation of the parameters in terms of bias and coverage rate. This difference mainly affects the variances of the trial-level random effects (parameters $\sigma^2_{\nu_S}$, $\sigma^2_{\nu_T}$, and $\sigma_{\nu_{S,T}}$) and the parameters $\beta_{Z,S}$ and $\beta_{Z,T}$ of the treatment effects on S and T.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

Table 1. *Simulations results for scenarios where* $\theta^2 = \rho^2 = 1$

		K=20 trials				K= 50 trials			
Parameter	Value	Est.[†]	SE.[‡]	SD.[§]	CR.[¶]	Est.	SE.	SD.	CR.
				$\gamma(s) = 0$					
PTE(20)	0.000	−0.000	0.214	0.138	0.976	0.004	0.116	0.107	0.957
PTE(30)	0.000	−0.002	0.186	0.157	0.945	−0.004	0.115	0.114	0.955
PTE(40)	0.000	−0.007	0.171	0.159	0.961	−0.017	0.115	0.108	0.965
NIE(20)	0.000	0.001	0.019	0.018	0.963	0.002	0.018	0.017	0.935
NIE(30)	0.000	0.002	0.021	0.022	0.940	0.000	0.020	0.020	0.945
NIE(40)	0.000	0.001	0.022	0.022	0.958	−0.002	0.022	0.020	0.967
NDE(20)	0.121	0.164	0.029	0.041	0.594	0.164	0.025	0.032	0.562
NDE(30)	0.184	0.181	0.031	0.045	0.812	0.188	0.027	0.035	0.889
NDE(40)	0.195	0.190	0.032	0.048	0.835	0.204	0.029	0.036	0.846
$\beta_{Z,S}$	−1.000	−0.760	0.211	0.348	0.673	−1.029	0.196	0.194	0.950
$\beta_{Z,T}$	−1.000	−1.046	0.209	0.264	0.929	−1.106	0.209	0.208	0.957
θ^2	1.000	1.096	0.507	0.498	0.929	1.065	0.550	0.584	0.887
ρ^2	1.000	1.832	0.496	1.323	0.673	1.096	0.357	0.402	0.935
$\sigma^2_{\nu_S}$	0.700	1.107	0.368	0.598	0.780	0.803	0.305	0.323	0.947
$\sigma^2_{\nu_T}$	0.700	0.682	0.307	0.434	0.775	0.819	0.337	0.384	0.930
$\sigma_{\nu_{S,T}}$	0.420	0.485	0.229	0.348	0.838	0.512	0.224	0.260	0.935
ζ	1.000	1.067	0.397	0.498	0.919	1.160	0.488	0.575	0.932
α	1.000	1.003	0.167	0.206	0.911	1.052	0.199	0.199	0.957
τ	0.395	0.416	—	0.072	0.757	0.370	—	0.051	0.935
R^2_{trial}	0.360	0.396	0.151	0.207	0.817	0.430	0.164	0.168	0.910
				$\gamma(s) > 0$					
PTE(20)	0.081	0.136	0.235	0.168	0.933	0.141	0.119	0.118	0.929
PTE(30)	0.130	0.147	0.201	0.281	0.954	0.144	0.118	0.118	0.948
PTE(40)	0.133	0.172	0.182	0.213	0.942	0.150	0.118	0.122	0.959
NIE(20)	0.011	0.025	0.022	0.023	0.896	0.027	0.022	0.022	0.904
NIE(30)	0.027	0.030	0.024	0.024	0.948	0.030	0.024	0.024	0.953
NIE(40)	0.029	0.032	0.025	0.026	0.933	0.032	0.025	0.026	0.945
NDE(20)	0.123	0.146	0.027	0.040	0.764	0.158	0.027	0.031	0.721
NDE(30)	0.182	0.157	0.029	0.041	0.782	0.171	0.029	0.034	0.882
NDE(40)	0.189	0.160	0.030	0.044	0.758	0.176	0.030	0.034	0.901
$\beta_{Z,S}$	−1.000	−0.971	0.209	0.291	0.870	−0.977	0.181	0.178	0.948
$\beta_{Z,T}$	−1.000	−0.951	0.195	0.258	0.882	−0.988	0.177	0.172	0.942
θ^2	1.000	1.043	0.447	0.436	0.934	1.025	0.472	0.501	0.901
ρ^2	1.000	1.444	0.398	0.634	0.804	1.231	0.362	0.412	0.942
$\sigma^2_{\nu_S}$	0.700	0.772	0.290	0.453	0.870	0.806	0.295	0.396	0.904
$\sigma^2_{\nu_T}$	0.700	0.595	0.254	0.556	0.565	0.783	0.283	0.305	0.934
$\sigma_{\nu_{S,T}}$	0.420	0.423	0.196	0.387	0.737	0.545	0.216	0.257	0.929
ζ	1.000	1.068	0.341	0.360	0.952	1.071	0.356	0.355	0.951
α	1.000	1.055	0.155	0.167	0.964	1.017	0.151	0.147	0.962
τ	0.395	0.411	—	0.051	0.861	0.387	—	0.046	0.926
R^2_{trial}	0.360	0.473	0.174	0.215	0.831	0.486	0.157	0.155	0.888

[†]Mean of the estimated parameters.
[‡]Mean of the estimated standard errors.
[§]Empirical standard deviation.
[¶]Coverage rate of the 95% confidence interval.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

Table 2. *Simulations results for scenarios where* $\theta^2 = \rho^2 = 0.64$

Parameter	Value	K=20 trials Est.†	SE.‡	SD.§	CR.¶	K=50 trials Est.	SE.	SD.	CR.
				$\gamma(s) = 0$					
PTE(20)	0.000	−0.032	0.333	0.588	0.939	−0.001	0.110	0.107	0.949
PTE(30)	0.000	0.100	0.458	1.585	0.952	0.007	0.109	0.099	0.968
PTE(40)	0.000	−0.016	0.266	0.138	0.963	0.003	0.109	0.102	0.971
NIE(20)	0.000	−0.000	0.017	0.018	0.939	0.001	0.017	0.017	0.946
NIE(30)	0.000	0.001	0.019	0.019	0.939	0.002	0.020	0.019	0.952
NIE(40)	0.000	−0.002	0.020	0.020	0.959	0.001	0.021	0.021	0.965
NDE(20)	0.121	0.180	0.027	0.063	0.371	0.170	0.026	0.030	0.507
NDE(30)	0.184	0.205	0.030	0.069	0.531	0.193	0.028	0.032	0.898
NDE(40)	0.195	0.223	0.031	0.073	0.517	0.211	0.030	0.036	0.861
$\beta_{Z,S}$	−1.000	−0.894	0.190	0.312	0.762	−0.988	0.167	0.154	0.952
$\beta_{Z,T}$	−1.000	−1.084	0.204	0.380	0.715	−1.025	0.182	0.193	0.958
θ^2	0.640	0.644	0.338	0.330	0.919	0.668	0.363	0.392	0.899
ρ^2	0.640	0.887	0.256	0.513	0.795	0.855	0.255	0.317	0.928
$\sigma^2_{\nu_S}$	0.700	0.890	0.310	0.437	0.859	0.623	0.229	0.258	0.817
$\sigma^2_{\nu_T}$	0.700	0.769	0.302	0.546	0.725	0.812	0.291	0.345	0.910
$\sigma_{\nu_{S,T}}$	0.420	0.395	0.203	0.442	0.634	0.388	0.180	0.215	0.857
ζ	1.000	1.170	0.519	0.635	0.930	1.189	0.541	0.723	0.920
α	1.000	1.051	0.171	0.203	0.950	1.043	0.173	0.211	0.942
τ	0.315	0.312	—	0.064	0.862	0.309	—	0.052	0.920
R^2_{trial}	0.360	0.335	0.132	0.233	0.708	0.346	0.146	0.174	0.891
				$\gamma(s) > 0$					
PTE(20)	0.081	0.063	0.303	1.865	0.874	0.119	0.099	0.107	0.918
PTE(30)	0.130	0.182	0.179	0.125	0.943	0.133	0.097	0.094	0.952
PTE(40)	0.133	0.171	0.164	0.129	0.937	0.139	0.097	0.096	0.952
NIE(20)	0.011	0.033	0.022	0.025	0.821	0.025	0.020	0.022	0.879
NIE(30)	0.027	0.038	0.024	0.025	0.925	0.031	0.023	0.022	0.960
NIE(40)	0.029	0.037	0.025	0.026	0.937	0.034	0.024	0.023	0.949
NDE(20)	0.123	0.157	0.029	0.048	0.613	0.179	0.027	0.029	0.429
NDE(30)	0.182	0.169	0.030	0.047	0.777	0.200	0.029	0.032	0.873
NDE(40)	0.189	0.175	0.031	0.051	0.739	0.208	0.031	0.034	0.873
$\beta_{Z,S}$	−1.000	−1.022	0.185	0.376	0.634	−0.979	0.175	0.172	0.944
$\beta_{Z,T}$	−1.000	−0.899	0.195	0.275	0.800	−1.052	0.184	0.158	0.977
θ^2	0.640	0.667	0.323	0.365	0.922	0.677	0.343	0.325	0.924
ρ^2	0.640	1.006	0.276	0.502	0.722	0.767	0.230	0.268	0.929
$\sigma^2_{\nu_S}$	0.700	0.756	0.264	0.529	0.694	0.803	0.270	0.298	0.935
$\sigma^2_{\nu_T}$	0.700	0.960	0.335	0.544	0.831	0.874	0.294	0.329	0.938
$\sigma_{\nu_{S,T}}$	0.420	0.530	0.200	0.421	0.694	0.553	0.209	0.265	0.881
ζ	1.000	1.095	0.396	0.469	0.934	1.116	0.428	0.469	0.932
α	1.000	1.036	0.144	0.167	0.922	1.048	0.158	0.150	0.966
τ	0.315	0.332	—	0.062	0.800	0.311	—	0.051	0.912
R^2_{trial}	0.360	0.422	0.121	0.232	0.675	0.448	0.141	0.176	0.805

†Mean of the estimated parameters.
‡Mean of the estimated standard errors.
§Empirical standard deviation.
¶Coverage rate of the 95% confidence interval.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

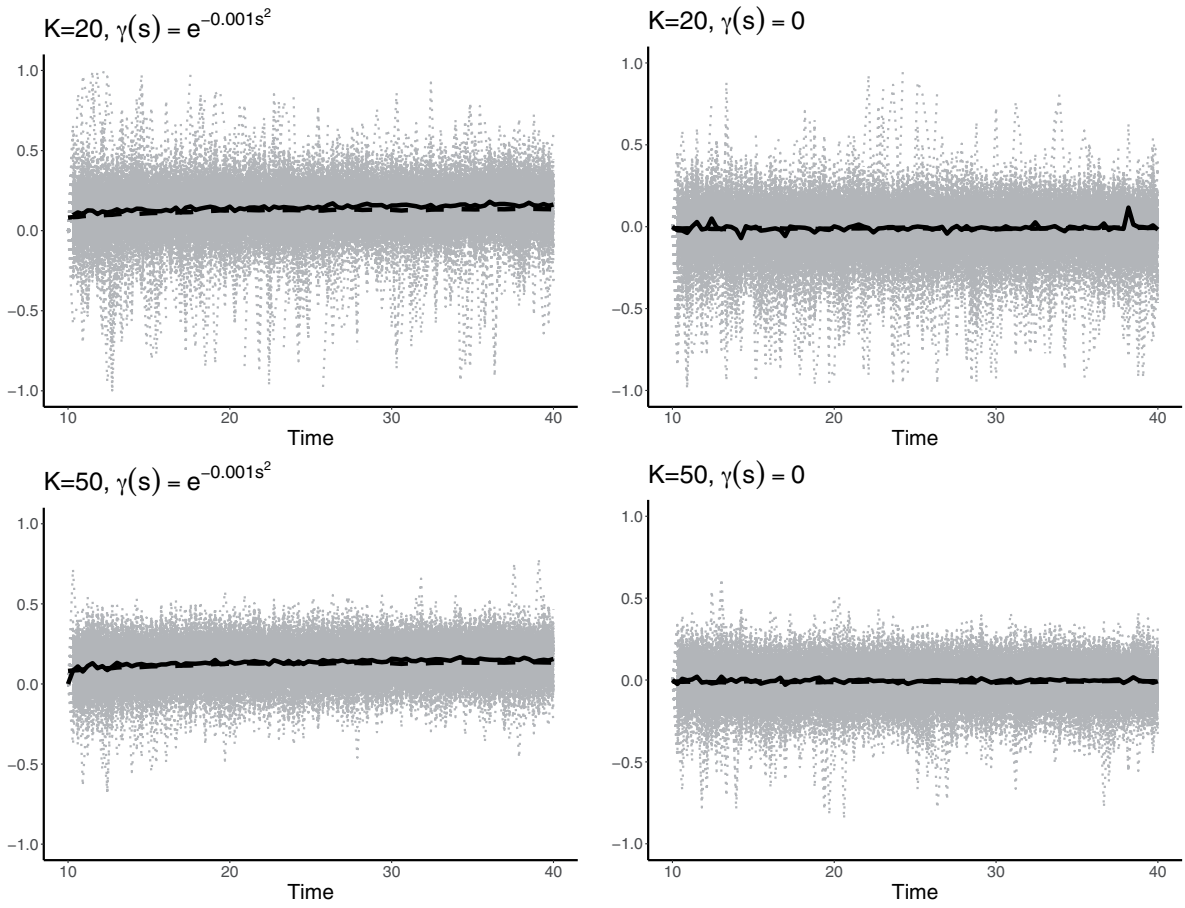


Fig. 2. Estimated PTE(t) for scenarios where $\theta^2 = \rho^2 = 1$ (dotted gray curves), their average (solid black curve), and the true curve (black dashed curve).

For trial-level random effects variances, the estimations are biased upward while their associated standard errors are biased downward, but increasing the number of trials reduces these biases. This bias is also reflected in the estimation of R^2_{trial} which is biased and results in a low coverage rate.

In addition, increasing the number of trials reduces the differences between the estimated standard errors of the different parameters and their empirical standard deviation (columns SE and SD respectively, in the tables). This improvement leads to a better coverage rate of the confidence interval for these parameters.

As for the estimation of the function PTE(t) and the NDE and NIE, from Tables 1 and 2, we see that both the NIE and PTE(t) as well as their standard errors are well estimated. Their confidence intervals also have coverage rate close to the nominal 95%. However, the estimation of natural direct effect (NDE) is biased which results in a low coverage rate. For the NDE, increasing the number of centers seems to moderately improve its estimation and the coverage rate, except for earlier timepoints ($t = 20$) where there is no clear improvement.

Figure 2 shows the estimations of PTE(t) for the cases where $\theta^2 = \rho^2 = 1$. Results for $\theta^2 = \rho^2 = 0.64$ are given in Figure 4 in Appendix F of the Supplementary material at *Biostatistics* online. These figures show the 300 estimated PTE(t) (one per simulated data set) in the dotted curves, their average at each time (solid black curve), and the true PTE(t) (dashed black curve). From these figures, we see that increasing the number of trials (K) from 20 (top panel) to 50 (bottom panel) reduces both the bias of the estimations of PTE(t) and its variance.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

These figures also show that the bias of the estimator of PTE(t) is smaller for larger values of t. This could be due to the fact that for shorter times t the difference of survivals $\mathbb{S}_{11}(t) - \mathbb{S}_{00}(t)$ at the denominator of PTE(t) is likely close to zero (since both survivals are close to one) resulting in a less stable estimation of PTE(t).

This suggests that the estimation of PTE(t) should only be done for sufficiently large values of t. As can be seen in Figure 2, we only plotted the estimate of PTE(t) for $t \geq 10$ since the results for $t < 10$ were not relevant due to the very large instability of the estimations of PTE(t). This also suggests that before computing PTE(t), one should plot the function $\mathbb{S}_{11}(t)$ and $\mathbb{S}_{00}(t)$ to see if there are some time points where the two curves cross or are too close to avoid computing the function PTE(t) for those time points.

In all of the above scenarios, simulations included unmeasured confounders (simulated by the individual random effects) that were well specified with respect to the model: both the confounders at the simulation and at the estimation stages are Gaussian. We investigated scenarios in which these confounders were misspecified. First, with uniformly distributed confounders over [0, 1] and second, with binary confounders. The results of these additional simulations can be found in Appendix D of the Supplementary material at *Biostatistics* online. These results suggest that the approach is quite robust when there are (at least time-fixed) unmeasured confounders since the effects on the bias in estimating PTE(t) are limited. Moreover, we also investigated a case with a smaller number of trials ($K = 10$) whose results are given in Table 3. From these results, we can see that the method performs poorer with a lower number of centers, especially regarding the estimation of the standard errors of the parameters of the model as well as for PTE(t), suggesting that this method should be applied on studies with a sufficient number of trials/clusters.

4. Application

We applied this approach to the meta-analysis gastric comparing adjuvant chemotherapy with surgery alone in patients with resectable gastric cancer (Paoletti *and others*, 2010). We also used this method in a second application with data from a trial in second-line treatment of metastatic colorectal cancer (Peeters *and others*, 2014). The analysis and results of this second application can be found in Appendix F of the Supplementary material at *Biostatistics* online.

4.1. *Time-to-relapse as a surrogate of overall survival in gastric cancer*

We studied time-to-relapse (TTR) as a surrogate for overall survival (OS). OS was defined as the time from randomization to death from any cause, whereas TTR was defined as the time from randomization to disease recurrence or the occurrence of second cancer, whichever occurred first.

4.1.1. *Data description* The data set contains information on 3288 patients from 14 clinical trials. Of these 3288 patients, 1634 (49%) received adjuvant chemotherapy before surgery, and the remaining 1654 (51%) underwent surgery without prior adjuvant chemotherapy. The total number of deaths was 1705 (51.8%) and ranged from 13.4% to 84% in the 14 trials. A total of 1352 TTR events (41%) were observed, ranging from 10% to 73% across the trials. The median follow-up time was 4.78 years in the adjuvant chemotherapy group (range = [0.005, 24.88]) and 4.40 years in the control group (range = [0.005, 23.96]). Kaplan–Meier estimates of survival functions for both endpoints are provided in Figure 6 in Appendix G of the Supplementary material at *Biostatistics* online.

4.1.2. *Data analysis and results* We applied the model defined by (2.2) to this data set. The baseline hazard functions, $\lambda_{0,T}(t)$ and $\lambda_{0,S}(t)$, were estimated using splines with seven knots. For the estimation of the function $\gamma(s)$, cubic B-splines with one interior knot at the median of the observed surrogate times S

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

Table 3. *Simulations results with $K = 10$ centers and $\gamma(s) > 0$*

Parameters	Value	Est.[†]	SE.[‡]	SD.[§]	CR.[¶]
PTE(20)	0.081	0.180	0.719	0.194	0.970
PTE(30)	0.130	0.149	0.471	0.212	0.960
PTE(40)	0.133	0.120	0.698	0.398	0.980
NIE(20)	0.011	0.032	0.025	0.026	0.880
NIE(30)	0.027	0.036	0.028	0.029	0.900
NIE(40)	0.029	0.038	0.029	0.030	0.940
NDE(20)	0.123	0.160	0.036	0.065	0.550
NDE(30)	0.182	0.167	0.039	0.065	0.770
NDE(40)	0.189	0.173	0.040	0.068	0.740
$\beta_{Z,S}$	−1.000	−0.973	0.283	0.426	0.798
$\beta_{Z,T}$	−1.000	−0.939	0.286	0.383	0.862
θ^2	1.000	1.048	0.535	0.667	0.898
ρ^2	1.000	1.087	0.422	1.074	0.615
$\sigma^2_{\nu_S}$	0.700	0.783	0.435	0.560	0.798
$\sigma^2_{\nu_T}$	0.700	0.690	0.414	0.583	0.734
$\sigma_{\nu_{S,T}}$	0.420	0.436	0.303	0.421	0.817
ζ	1.000	1.154	0.498	0.564	0.881
α	1.000	1.047	0.232	0.274	0.899
τ	0.395	0.340	—	0.102	0.780
R^2_{trial}	0.360	0.515	0.197	0.239	0.872

[†]Mean of the estimated parameters.
[‡]Mean of the estimated standard errors.
[§]Empirical standard deviation.
[¶]Coverage rate of the 95% confidence interval.

were used. The results for the estimation of the main parameters of the model are given in Table 4. In this table, the estimations of PTE(t) at three time points (2, 3, and 8 years) are also reported. The graphical representation of the estimation of $\gamma(s)$ is given in Figure 3 and the estimated function PTE(t) is given in Figure 4(a). The estimated natural direct, indirect, and total effects associated with PTE(t) are shown in Figure 4(b).

From Table 4, we can see that there is a significant treatment effect on the surrogate ($\beta_{Z,S} = -0.47$), suggesting that adjuvant chemotherapy prolongs TTR. The estimation of $\beta_{Z,T}$, which is not significantly different from 0, suggests that the treatment has no direct effect on overall survival (see Figure 1(b)). In addition, the variances of the random effects ω_{ij} and u_i (θ^2 and ρ^2, respectively), which are significantly different from zero, suggest an association between S and T at the individual level and an heterogeneity among of the patients between the trials. This suggests that these frailties capture some of the effects of confounding factors that were not measured. However, the nonsignificant variances of ν_S, ν_T, and $\nu_{S,T}$ do not indicate heterogeneity in treatment effects between trials.

Furthermore, as shown in Figure 3, the estimation of $\gamma(s)$ suggests that there is a strong effect of TTR on OS. The estimated function being positive shows that the occurrence of a TTR event increases the risk of occurrence of death. Because the function is increasing, a late occurrence of a TTR event has a greater effect on increasing the risk of OS than an earlier occurrence of a TTR event.

Overall, $\hat{\gamma}(s)$ and $\hat{\beta}_{Z,S}$ being significantly different from zero indicates that there is a mediated effect of the treatment on OS through TTR and $\hat{\beta}_{Z,T}$ being close to zero implies that most of the total effect is

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

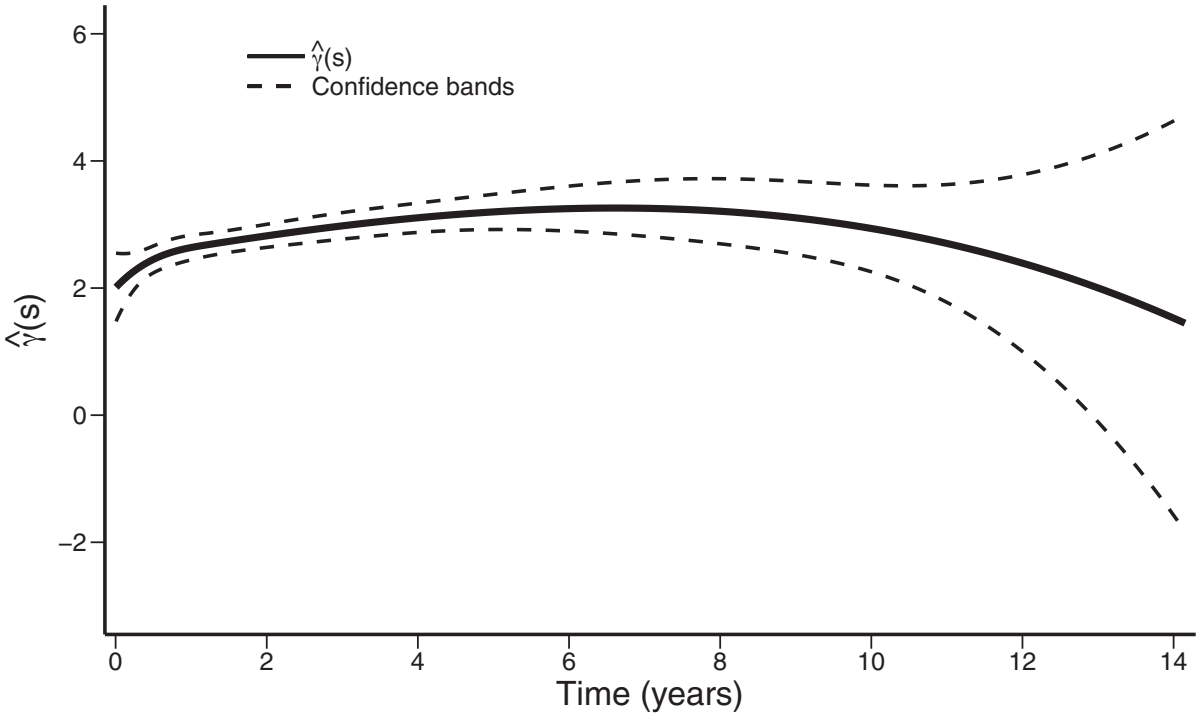


Fig. 3. Estimated $\hat{\gamma}(s)$ from the gastric application with its 95% pointwise confidence bands.

Table 4. *Estimated parameters of model (2.2) for the application on gastric cancer*

Parameter	Est.	Std. Err.	p-value
PTE(2)	0.853	0.283	—
PTE(3)	0.875	0.254	—
PTE(8)	0.839	0.155	—
$\beta_{Z,S}$	−0.453	0.104	<0.001
$\beta_{Z,T}$	−0.057	0.091	0.532
θ^2	3.807	0.317	<0.001
ρ^2	1.035	0.172	<0.001
$\sigma^2_{v_S}$	0.011	0.023	0.633
$\sigma^2_{v_T}$	0.018	0.027	0.495
$\sigma_{v_{S,T}}$	−0.014	0.011	0.212
ζ	0.681	0.043	<0.001
α	0.758	0.063	<0.001

this mediated effect. This can be seen in Figure 4(a), which shows that the estimated function $\hat{\text{PTE}}(t)$ is very close to 1.

Since $\hat{\beta}_{Z,S} < 0$ and $\hat{\gamma}(s) > 0$, patients treated with adjuvant chemotherapy have better survival, so their risk of death is likely to be increased at a later time point than patients in the control group, implying that the overall mediated treatment effect has a protective effect towards OS. This is shown in Figure 4(b), where the natural direct effect is not significantly different from zero over time, whereas both the natural indirect effect and the overall effect are significantly different from zero. From these results, it appears

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

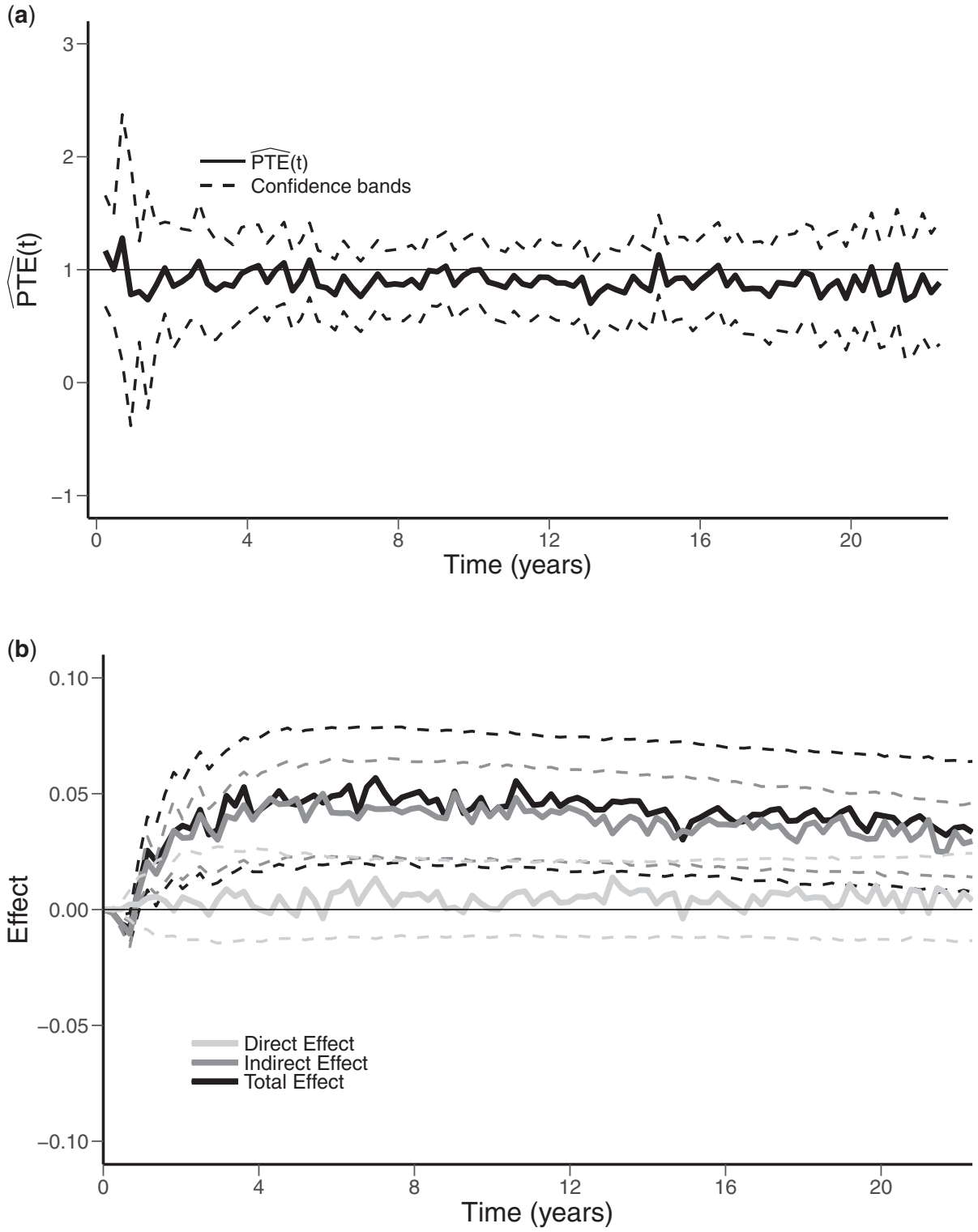


Fig. 4. Results from the application on gastric cancer. (a) Estimated PTE(t) and its 95% pointwise confidence bands. (b) Estimated natural effects and their 95% pointwise confidence bands.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

that there is a significant overall effect of treatment on OS and that most of this effect comes from its effect on TTR. TTR is therefore likely to be a good surrogate for OS in the setting of adjuvant chemotherapy in patients with resectable gastric cancer.

These results had to be taken with caution as the limited number of trials (14) may cause bias in the estimated parameters of the model (and especially their standard errors). However, the high value of the estimated PTE(t) suggests that even if these results may be biased, it is likely that the true value PTE(t) is also high and that TTR is still a good surrogate for OS.

5. Discussion

In this article, we have proposed an approach to mediation analysis in the context of surrogate validation when both the surrogate and the final endpoint are time-to-event. This approach can be decomposed into two parts. First, using the counterfactual framework of causal inference, we defined a surrogacy measure that corresponds to the proportion of the treatment effect on the final endpoint that results from its effect on the surrogate. To calculate this proportion, we need to estimate the distribution of counterfactual outcomes, which leads us to the second part of this article, which deals with the statistical model used in this context.

To causally interpret the estimated direct and indirect effects of the treatment, this approach relies on several assumptions. In particular, assumption (A5') is crucial. To ensure that this assumption holds, the baseline covariates X and the random effects vector ξ must capture all of the confounding between S and T. It should be noted that there are situations where this is unlikely, for example, when there is time-varying confounding. A generalization of this method would include both time-varying covariates and time-varying random effects but our approach is currently limited to settings of time-fixed confounding.

Problems can arise with the function PTE(t) when both survival values appearing in the denominator are equal. This is the case, for example, when the survival functions $\mathbb{S}_{11}(t)$ and $\mathbb{S}_{00}(t)$ of the experimental and control groups cross. In this case, at this particular time t, there is no overall effect of treatment on the final endpoint, and PTE(t) is not defined. It may therefore be helpful to first examine the estimated Kaplan–Meier curves in each group to determine whether this is likely to be the case.

We used a counterfactual approach to define the survival functions $\mathbb{S}_{z'z}(t)$, which are then used to define the proportion PTE(t). To be able to identify these counterfactual survival functions from an observed data set, we made several assumptions. In particular, the sequential ignorability assumption warrants some attention and its plausibility might be investigated using sensitivity analyses.

The surrogacy measure proposed in this article is the function PTE(t). For clinical purposes, it may be more meaningful to assess surrogacy only at some relevant time points, for example, at 12 or 24 months. In this case, the function PTE(t) can be calculated only for these time points. However, the entire function PTE(t) is also of interest, as it can show a trend for surrogacy over time.

We would like to point out that while this approach to mediation analysis and the associated definition of indirect and direct effects is based on the counterfactual outcomes, other approaches were also proposed for time-to-event outcomes (Strohmaier *and others*, 2015; Aalen *and others*, 2020). These methods use dynamic modeling and do not use counterfactual outcomes to derive a causal definition of direct and indirect effects but are based on modeling structural equations using additive hazards in a continuous-time manner.

We proposed a method to validate a surrogate endpoint by studying the proportion of the treatment effect on the final endpoint that goes through the surrogate. The indirect treatment effect through the surrogate is composed of the direct treatment effect on the surrogate and the effect of the surrogate on the final endpoint. If at least one of these effects is absent, the indirect effect of treatment on the final endpoint is zero. However, as noted by VanderWeele (2013), in the case where there is no direct effect of the surrogate on the final endpoint, the surrogate could still be a good surrogate if, for example, there is a strong association between the surrogate and the final endpoint due to confounders. Model (2.2) can also

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

be used to study the association between the surrogate and the final endpoint through the R^2_{trial} and the Kendall's τ as described in Sofeu *and others* (2019). These quantities quantify the association between the two endpoints at the individual and trial levels. In particular, the R^2_{trial} is the coefficient of determination of the trial-level random effects. These quantities have been proposed for surrogate validation (Buyse *and others*, 2000). However, because they only measure association, these quantities do not provide information regarding the proportion of treatment mediated. Kendall's τ, which measures the individual-level association between S and T, would not be able to discriminate which proportion of this association is due to confounding or due to a direct link of S on T. Although a strong mediated effect may be sufficient for surrogacy, it is not a necessary condition for an endpoint to be a good surrogate. This model allows the use of both approaches, which complement each other, to evaluate a surrogate by either mediated effects or association.

As the simulations with $K = 10$ trials suggest, this method should be employed with a sufficient number of trials. However, other units of clustering could be used, such as centers of care for example.

Because the proposed approach focuses on the indirect effect of a treatment on a final endpoint through a surrogate, it may not be appropriate in cases where the surrogate is a composite endpoint that includes the final endpoint as an event. In such a case, the occurrence of the final endpoint leads to the occurrence of the surrogate and it is therefore difficult to define what would be a direct effect of the surrogate on the final endpoint since both endpoints would occur at exactly the same time. This is the case for progression-free survival when death is also accounted as a progression. The proposed approach is not appropriate for this situation, and it would be interesting to see if other approaches can be used to derive an appropriate definition of a direct and indirect effect for composite endpoints.

6. Software

The proposed method is available in the `jointSurroPenal` function of the R package `frailtypack` (Król *and others*, 2017) which is available on the R-CRAN at https://CRAN.R-project.org/package=frailtypack. The data set should have the following format:

```
head(data)
patientID    timeT    timeS statusT statusS trt trialID
        1 9.057946 2.217739       1       1   0       1
        2 2.986813 1.389263       1       1   0       1
        3 8.874237 8.874237       1       0   1       1
        4 3.245388 1.809671       1       1   1       1
        5 4.448964 2.603604       1       1   0       1
```

Where `timeT, timeS, statusT,` and `statusS` are respectively the time for the final and surrogate endpoints as well as their censoring indicators. The `R` command to call model (2.2) with one interior knot to estimate the function $\gamma(s)$ is:

```
require(frailtypack)
result<-jointSurroPenal(data,mediation=TRUE,g.nknots=1,
                        rt.ntimes=30,rt.nmc=10000,rt.boot=T,
                        rt.nboot=2000,rt.boot.nmc=10000)
```

After the execution of the program, summary of the estimations and related figures (Kaplan–Meier estimates, baseline hazards, $\gamma(s)$, PTE(t), and natural effects) can be obtained from the commands `summary(result)` and `plot(result)`.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

Supplementary material

Supplementary material is available online at http://biostatistics.oxfordjournals.org.

Acknowledgments

Computations for this work were performed using HPC resources from the Mesocentre de Calcul Intensif Aquitain (MCIA). This publication is based on data obtained from www.projectdatasphere.org, which is maintained by Project Data Sphere. Neither Project Data Sphere nor the owner(s) of any information from the web site have contributed to, approved or are in any way responsible for the contents of this publication. The authors would like to thank the referees and associate editor whose comments greatly helped improve this manuscript.

Conflict of Interest: None declared.

Funding

This work was supported within the framework of the PIA3 (Investment for the future), project reference 17-EURE-0019.

References

Aalen, O. O., Stensrud, M. J., Didelez, V., Daniel, R., Røysland, K. and Strohmaier, S. (2020). Time-dependent mediators in survival analysis: modeling direct and indirect effects with the additive hazards model. *Biometrical Journal* **62**, 532–549.

Andrews, R. M. and Didelez, V. (2020). Insights into the cross-world independence assumption of causal mediation analysis. *Epidemiology* **32**, 209–219.

Buyse, M. and Molenberghs, G. (2005). *The Evaluation of Surrogate Endpoints*. (Vol. 427). T. Burzykowski (Ed.). New York: Springer.

Burzykowski, T., Buyse, M., Piccart-Gebhart, M. J., Sledge, G., Carmichael, J., Luck, H.-J., Mackey, J. R., Nabholtz, J.-M., Paridaens, R., Biganzoli, L. *and others*. (2008). Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *Journal of Clinical Oncology* **26**, 1987–1992.

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D. and Geys, H. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* **1**, 49–67.

Buyse, M., Molenberghs, G., Paoletti, X., Oba, K., Alonso, A., Van der Elst, W. and Burzykowski, T. (2016). Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. *Biometrical Journal* **58**, 104–132.

Conlon, A. S. C., Taylor, J. M. G. and Elliott, M. R. (2014). Surrogacy assessment using principal stratification when surrogate and outcome measures are multivariate normal. *Biostatistics* **15**, 266–283.

Daniels, M. J. and Hughes, M. D. (1997). Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* **16**, 1965–1982.

Didelez, V. (2019). Defining causal mediation with a longitudinal mediator and a survival outcome. *Lifetime Data Analysis* **25**, 593–610.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

Freedman, L. S., Graubard, B. I. and Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine* **11**, 167–178.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945–960.

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

HUANG, Y.-T. (2020). Causal mediation of semicompeting risks. *Biometrics*.

IMAI, K., KEELE, L. AND YAMAMOTO, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* **25**, 51–71.

JOFFE, M. M. AND GREENE, T. (2009). Related causal frameworks for surrogate outcomes. *Biometrics* **65**, 530–538.

KRÓL, A., MAUGUEN, A., MAZROUI, Y., LAURENT, A., MICHIELS, S. AND RONDEAU, V. (2017). Tutorial in joint modeling and prediction: a statistical software for correlated longitudinal outcomes, recurrent events and a terminal event. *Journal of Statistical Software* **81**, 1–52. https://doi.org/10.18637/jss.v081.i03

LEE, D.-W., JANG, M.-J., LEE, K.-H., CHO, E. J., LEE, J.-H., YU, S. J., KIM, Y. J., YOON, J.-H., KIM, T.-Y., HAN, S.-W. *and others*. (2016). TTP as a surrogate endpoint in advanced hepatocellular carcinoma treated with molecular targeted therapy: meta-analysis of randomised controlled trials. *British Journal of Cancer* **115**, 1201–1205.

LIN, S.-H., YOUNG, J. G., LOGAN, R. AND VANDERWEELE, T. J. (2017). Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. *Statistics in Medicine* **36**, 4153–4166.

PAOLETTI, X., OBA, K., BURZYKOWSKI, T., MICHIELS, S., OHASHI, Y., PIGNON, J.-P., ROUGIER, P., SAKAMOTO, J., SARGENT, D., SASAKO, M. *and others*. (2010). Benefit of adjuvant chemotherapy for resectable gastric cancer: a meta-analysis. *JAMA* **303**, 1729–1737.

PEETERS, M., PRICE, T. J., CERVANTES, A., SOBRERO, A. F., DUCREUX, M., HOTKO, Y., ANDRÉ, T., CHAN, E., LORDICK, F., PUNT, C. J. A. *and others*. (2014). Final results from a randomized phase 3 study of folfiri±panitumumab for second-line treatment of metastatic colorectal cancer. *Annals of Oncology* **25**, 107–116.

PRENTICE, R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine* **8**, 431–440.

ROBINS, J. M. AND GREENLAND, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, **3**, 143–155.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688.

SOFEU, C. L., EMURA, T. AND RONDEAU, V. (2019). One-step validation method for surrogate endpoints using data from multiple randomized cancer clinical trials with failure-time endpoints. *Statistics in Medicine* **38**, 2928–2942.

STROHMAIER, S., RØYSLAND, K., HOFF, R., BORGAN, Ø., PEDERSEN, T. R. AND AALEN, O. O. (2015). Dynamic path analysis—a useful tool to investigate mediation processes in clinical survival trials. *Statistics in Medicine* **34**, 3866–3887.

TCHETGEN TCHETGEN, E. J. (2011). On causal mediation analysis with a survival outcome. *The International Journal of Biostatistics* **7**, Article 33. DOI:10.2202/1557-4679.1351.

VANDENBERGHE, S., DUCHATEAU, L., SLAETS, L., BOGAERTS, J. AND VANSTEELANDT, S. (2018). Surrogate marker analysis in cancer clinical trials through time-to-event mediation techniques. *Statistical Methods in Medical Research* **27**, 3367–3385.

VANDERWEELE, T. J. (2013). Surrogate measures and consistent surrogates. *Biometrics* **69**, 561–565.

ZHENG, C. AND LIU, L. (2021). Quantifying direct and indirect effect for longitudinal mediator and survival outcome using joint modeling approach. *Biometrics* **78**, 1233–1243.

ZIGLER, C. M. AND BELIN, T. R. (2012). A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint. *Biometrics* **68**, 922–932.

[*Received November 22, 2021; revised October 24, 2022; accepted for publication October 25, 2022*]

Downloaded from https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxac044/6832427 by Universite de Bordeaux user on 18 November 2022

Supporting information for "Time-to-event surrogate endpoint validation using mediation analysis and meta-analytic data" by Q. Le Coënt, C. Legrand and V. Rondeau.

Web Appendix A: Likelihood construction and estimation

For an individual j from trial i, let T^*_{ij} and S^*_{ij} be the true, uncensored, final and surrogate endpoints. Let T_{ij} and S_{ij} be the censored observed outcomes, δ_{ij} and d_{ij} their associated censoring indicators. That is, if C_{ij} is the censoring time we have,

$$T_{ij} = \min\left(T^*_{ij}, C_{ij}\right)$$
$$S_{ij} = \min\left(S^*_{ij}, T_{ij}\right)$$
$$\delta_{ij} = \mathbb{1}_{T^*_{ij} \leq C_{ij}}$$
$$d_{ij} = \mathbb{1}_{S^*_{ij} \leq T_{ij}}$$

Let ϕ be the vector of all parameters of the model. Assuming an independent and non-informative censoring the individual contribution to the likelihood of

subject j from trial i given $(\omega_{ij}, u_i, \nu_{S_i}, \nu_{T_i}, Z_{ij}, X_{ij})$ is:

$$\begin{aligned} L_{ij}(\phi) &= \lambda_{T_{ij}}(T_{ij} \mid S_{ij}, d_{ij}, \omega_{ij}, u_i, \nu_{T,i}, \nu_{S,i}, Z_{ij}, X_{ij}; \phi)^{\delta_{ij}} \\ &\quad \times \exp\left(- \int_0^{T_{ij}} \lambda_{T,ij}(t \mid S_{ij}, d_{ij}, \omega_{ij}, u_i, \nu_{T,i}, \nu_{S,i}, Z_{ij}, X_{ij}; \phi) \mathrm{d}t \right) \\ &\quad \times \lambda_{S_{ij}}(S_{ij} \mid \omega_{ij}, u_i, \nu_{T,i}, \nu_{S,i}, Z_{ij}, X_{ij}; \phi)^{d_{ij}} \\ &\quad \times \exp\left(- \int_0^{S_{ij}} \lambda_{S,ij}\left(t \mid \omega_{ij}, u_i, \nu_{T,i}, \nu_{S,i}, Z_{ij}, X_{ij}; \phi\right) \mathrm{d}t \right) \end{aligned}$$

where

$$\begin{aligned} \lambda_{T_{ij}}(t \mid S_{ij}, d_{ij}, \omega_{ij}, u_i, \nu_{T,i}, \nu_{S,i}; \phi) = \lambda_{0,T}(t) \exp\Big(& \zeta\omega_{ij} + \alpha u_i + Z_{ij}\left(\nu_{T_i} + \beta_{Z,T}\right) \\ & + \beta_T^{\mathsf{T}} X_{ij} + d_{ij}\gamma(S_{ij}) \Big) \end{aligned}$$

and

$$\begin{aligned} \lambda_{S_{ij}}(t \mid \omega_{ij}, u_i, \nu_{T,i}, \nu_{S,i}, Z_{ij}, X_{ij}; \phi) = \lambda_{0,S}(t) \exp\Big(& \omega_{ij} + u_i + Z_{ij}\left(\nu_{S_i} + \beta_{Z,S}\right) \\ & + \beta_S^{\mathsf{T}} X_{ij} \Big). \end{aligned}$$

Once we have the conditional contribution for a subject, we can get the contribution of a trial by taking the product of individual contributions integrated over the distribution of ω. Hence for trial i we have,

$$L_i(\phi \mid \mu_i, \nu_{T,i}, \nu_{S,i}) = \prod_{j=1}^{n_i} \int_{\mathbb{R}} L_{ij}(\phi \mid \omega_{ij}, \mu_i, \nu_{T,i}, \nu_{S,i}) f_\omega(\omega_{ij}) \mathrm{d}\omega_{ij}$$

where f_ω is chosen as the density of a centered gaussian distribution with variance θ^2. The whole (marginal) likelihood data is then,

$$L(\phi) = \prod_{i=1}^{K} \int_{\mathbb{R}^3} L_i(\phi \mid \mu_i, \nu_{T,i}, \nu_{S,i}) f(\mu_i, \nu_{T,i}, \nu_{S,i}) \mathrm{d}\mu_i \mathrm{d}\nu_{T,i} \mathrm{d}\nu_{S,i}$$

where by assumption $f(\mu_i, \nu_{T,i}, \nu_{S,i})$ is the density of a centered tridimensional gaussian vector whose covariance matrix is given by:

$$\begin{pmatrix} \rho^2 & 0 & 0 \\ 0 & \sigma_S^2 & \sigma_{ST} \\ 0 & \sigma_{ST} & \sigma_T^2 \end{pmatrix}$$

The baseline hazard functions $\lambda_{0,S}(t)$ and $\lambda_{0,T}(t)$ are estimated using M-splines with associated parameters $\eta_S = (\eta_{S,1}, \ldots, \eta_{S,\kappa_1})$ and $\eta_T = (\eta_{T,1}, \ldots, \eta_{T,\kappa_1})$ where κ_1 is the number of knots used in the spline basis. The function $\gamma(s)$ is estimated using a cubic B-splines basis, with associated parameters $\beta_\gamma = (\beta_{\gamma,1}, \ldots, \beta_{\gamma,\mathfrak{K}_2})$ where $\mathfrak{K}_2$ is the number of splines used which is 4 plus the number of innerknots. The penalized likelihood to maximize is:

$$L_{\text{pen}}(\phi) = L(\phi) - \kappa_T \int_0^\infty \left(\lambda_{0,T}^{''}(t)\right)^2 \mathrm{d}t - \kappa_S \int_0^\infty \left(\lambda_{0,S}^{''}(t)\right)^2 \mathrm{d}t \qquad (1)$$

where $\lambda_{0,T}^{''}(t)$ and $\lambda_{0,S}^{''}(t)$ are the second order derivatives of the baseline hazard functions of S and T. The penalization terms κ_T and κ_S can be chosen using cross-validation on reduced models (proportional-hazard models based on one outcome only). The estimation of ϕ is then carried out by maximizing $L_{\text{pen}}(\phi)$ using Marquardt algorithm. The marginalization of the conditional individual likelihood over the random effect is performed using a Monte-Carlo method for the trial-level random effects and a Gauss-Hermite quadrature for the individual random effects.

Web Appendix B: Piecewise Exponential distribution and simulation

A piecewise exponential distribution is characterized by a piecewise-constant hazard function. Althought an exponential distribution is fully characterized by only one parameter, λ say, a piecewise distribution needs to be characterized by the differents constants risks $\lambda_1, \ldots, \lambda_p$ as well as the breakpoints where the hazard function changes. In our case there is only one cutpoint at (S, the value of the surrogate, and thus two constant risks. We now describe how to generate value from a piecewise constant distribution given the breakpoints and the risks. In order to generate a random variable , T say, whose distribution has a piecewise constant hazard, one can use inverse transform sampling by inverting the associated survival function. Given the survival function $S(t)$ of

the variable T, we can draw observations $T_1, \ldots, T_n$ by inverting $S(t)$ and using uniform variables U. Indeed let $Y = S^{-1}(U)$ where $U \sim \mathcal{U}_{[0,1]}$, then,

$$\begin{aligned} P(Y \leq t) &= P(S^{-1}(U) \leq t) \\ &= P(U \geq S(t)) \\ &= 1 - S(t)) \end{aligned}$$

Thus Y has $S(t)$ as survival function and is then distributed as T. In order to sample T we need to be able to compute both $S(t)$ et $S^{-1}(t)$. Let $(a_i)_{i\in\{0,\ldots,K\}} \in R_+^{K+1}$ the breakpoints of the hazard function of T and $(\alpha_i)_{i\in\{1,\ldots,K+1\}}$ the constant risks associated with the intervals $[a_{i-1}, a_i[$. The piecewise constant hazard function is

$$\lambda_0(t) = \sum_{i=1}^{K} \alpha_i \mathbb{1}_{t\in[a_{i-1},a_i[}$$

and the survival function is then:

$$\begin{aligned} S(t) &= \exp\Big\{ -\int_0^t \lambda_0(u)\mathrm{d}u \Big\} \\ &= \exp\Big\{ -\int_0^t \sum_{i=1}^{K} \alpha_i \mathbb{1}_{u\in[a_{i-1},a_i[}\mathrm{d}u \Big\} \end{aligned}$$

Our aim is to be able to numerically evaluate S^{-1}. First we can rewrite S as

$$\begin{aligned} S(t) &= \exp\Big\{ -\int_0^t \sum_{i=1}^{K} \alpha_i \mathbb{1}_{u\in[a_{i-1},a_i[}\mathrm{d}u \Big\} \\ &= \exp\Big\{ -\sum_{i=1}^{K} \alpha_i \int_0^t \mathbb{1}_{u\in[a_{i-1},a_i[}\mathrm{d}u \Big\} \\ &= \exp\Big\{ -\sum_{i=1}^{K} 0 \vee (t - a_{i-1} \wedge a_i - a_{i-1}) \Big\}. \end{aligned}$$

For $t \in [a_{i-1}, a_i[$ we have,

$$\sum_{i=1}^{K} \Big\{ 0 \vee (t - a_{i-1} \wedge a_i - a_{i-1}) \Big\} = \sum_{j=1}^{i-1} \Big\{ \alpha_j (a_j - a_{j-1}) \Big\} + \alpha_i (t - a_{i-1})$$

Thus, for $t \in [a_{i-1}, a_i[$,

$$t = a_{i-1} - \frac{1}{\alpha_i} \Big(\log\big(S(t)\big) + \sum_{j=1}^{i-1} \alpha_j (a_j - a_{j-1}) \Big) \tag{2}$$

For a given value of $S(t)$, and knowing both $(a_i)_{i\in\{0,\dots,K\}}$ and $(\alpha_i)_{i\in\{1,\dots,K+1\}}$, we can determine the interval $[a_{i-1}, a_i[$ where t lies. Indeed, if $t \in [a_{i-1}, a_i[$,

$$a_{i-1} \leq t \leq a_i$$
$$a_{i-1} \leq a_{i-1} - \frac{1}{\alpha_i}\Big(\log\big[S(t)\big] + \sum_{j=1}^{i-1} \alpha_j(a_j - a_{j-1})\Big) \leq a_i$$

Thus we get two conditions on $S(t)$:

$$\log\big(S(t)\big) \leq -\sum_{j=1}^{i-1} \alpha_j(a_j - a_{j-1}) \quad (C_1)$$
$$\log\big(S(t)\big) \geq -\sum_{j=1}^{i} \alpha_j(a_j - a_{j-1}) \quad (C_2)$$

Those two conditions allow us to determine the interval $[a_{i-1}, a_i[$ in which t lies, given $S(t)$, the breakpoints and the risks. Once t has been located in the proper interval, Equation 2 is used to compute t. That is from the value of $S(t)$ we have a process to compute the value of t. We thus have a method to evaluate $S^{-1}(t)$ and can use to simulate T through reverse sampling.

Web Appendix C: Simulation scheme

Let be K be the number of trials and n the total number of subjects in the dataset. The different steps for simulating the datasets according to the parameters of the model are:

1. Generate $\nu = \left(\begin{pmatrix}\nu_{S_i}\\ \nu_{T_i}\end{pmatrix}\right)_{1\leq i\leq K}$ where $\begin{pmatrix}v_{S_i}\\ v_{T_i}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \Sigma = \begin{pmatrix}\sigma^2_{v_S} & \sigma_{V_{ST}}\\ \sigma_{V_{ST}} & \sigma^2_{V_T}\end{pmatrix}\right)$ with $\sigma^2_{V_S} = \sigma^2_{V_T} = 0.7$ and $\sigma_{V_{ST}} = 0.42$
2. Generate $u = (u_i)_{1\leq i\leq K}$ where $u_i \sim \mathcal{N}(0, \rho^2)$
3. For $i = 1, \dots, n$:
   (a) Draw k uniformly from $\{1, \dots, K\}$

(b) Draw $\omega \sim \mathcal{N}(0, \theta^2)$

(c) Draw $Z \sim \mathcal{B}(p = 0.5)$

(d) Compute $\zeta_S = \exp(\omega + u_k + v_{S_k} Z + \beta_{zs} Z)$

(e) Compute $\zeta_T = \exp(\zeta\omega + \alpha u_k + v_{T_k} Z + \beta_{zt} Z)$

(f) Draw $S^* \sim \mathcal{W}(a = \gamma_S, b = \frac{1}{\rho_S \zeta_S^{-\gamma_S}})$ with $\gamma_S = 1.8$ and $\rho_S = 0.0045$

(g) Draw $T^* \sim \mathcal{PE}\left(\rho_T \zeta_T^{-\gamma_T}, S^*, \gamma(S^*)\right)$ with $\rho_T = 0.02$

(h) Take censoring time $C = 50$

(i) Take $T = T^* \wedge C$, $\delta = \mathbb{1}_{T* \leq C}$

(j) Take $S = T \wedge S^*$ and $\delta_S = \mathbb{1}_{S^* \leq T}$

(k) Return the values of $(T, S, \delta, \delta_S, Z, k)$

Web Appendix D: Simulations with unmeasured confounding

We investigated cases with unmeasured confounding. Two cases were studied : the first one included a uniformly distributed confounder, $U \sim \mathcal{U}[0,1]$, and the second one included a binary confounder $\mathbb{P}(U=1)=0.5$. We set the number of centers to $K=30$, we used $\gamma(s)=\exp\left(-0.001s^2\right)$ and $\theta^2=\rho^2=0.64$. Both scenario was repeated 200 times. The results are given in Table 1 and Figure 1. From Table 1 results we see that the variances of the trial-level random effects and especially ρ^2 and $\sigma_{\nu_{S,T}}$ tend to be overestimated and the treatment effects on both S and T are underestimated. However the impact of these bias seems to be lessened on the estimation of the function PTE(t) according to Figure 1. From this figure we see that the estimation of PTE(t) is less biased for larger t which, as stated in the main article, can be explained by the smaller total effect for small t which induces a higher instability for PTE(t). Thus the random effects in the model can be useful in the presence of unmeasured confounders as to strenghten the estimation of PTE(t).

Table 1: Simulations results with unmeasured confounding

		$U \sim \mathcal{U}[0,1]$				$U \sim \mathcal{B}(0.5)$			
Parameter	Value	Est.*	SE.†	SD.‡	CR.§	Est.	SE.	SD.	
PTE(20)	0.131	0.136	0.137	0.131	0.974	0.134	0.142	0.117	0.948
PTE(30)	0.146	0.154	0.128	0.116	0.967	0.149	0.132	0.123	0.953
PTE(40)	0.152	0.164	0.127	0.105	0.978	0.157	0.123	0.108	0.976
NIE(20)	0.024	0.026	0.023	0.022	0.971	0.026	0.022	0.020	0.967
NIE(30)	0.031	0.031	0.024	0.022	0.960	0.031	0.024	0.025	0.925
NIE(40)	0.033	0.033	0.024	0.021	0.978	0.034	0.025	0.023	0.962
NDE(20)	0.163	0.159	0.032	0.044	0.835	0.167	0.032	0.041	0.854
NDE(30)	0.179	0.168	0.035	0.047	0.838	0.177	0.035	0.046	0.873
NDE(40)	0.185	0.169	0.035	0.046	0.831	0.177	0.036	0.045	0.887
$\beta_{Z,S}$	-1.000	-0.828	0.172	0.215	0.736	-0.903	0.175	0.219	0.831
$\beta_{Z,T}$	-1.000	-0.834	0.172	0.215	0.753	-0.916	0.166	0.218	0.848
θ^2	0.640	0.679	0.328	0.365	0.898	0.665	0.311	0.346	0.878
ρ^2	0.640	0.937	0.251	0.405	0.762	0.930	0.127	0.424	0.751
ζ	1.000	1.148	0.394	0.546	0.917	1.146	0.393	0.473	0.949
α	1.000	1.018	0.142	0.162	0.919	1.042	0.141	0.164	0.911
$\sigma^2_{\nu_S}$	0.700	0.760	0.261	0.375	0.821	0.793	0.146	0.377	0.844
$\sigma^2_{\nu_T}$	0.700	0.921	0.291	0.422	0.894	0.776	0.119	0.393	0.776
$\sigma_{\nu_{S,T}}$	0.420	0.522	0.185	0.312	0.823	0.476	0.151	0.321	0.781
τ	0.315	0.327	—	0.055	0.838	0.329	—	0.058	0.819
R^2_{trial}	0.360	0.422	0.126	0.184	0.696	0.396	0.131	0.211	0.688

*Mean of the estimated parameters

†Mean of the estimated standard errors

‡Empirical standard deviation

§Coverage rate

The black dashed curve is true value of PTE(t) to be estimated, dotted gray curves represent the estimations for each simulated dataset and the solid black curve is their average at each time point

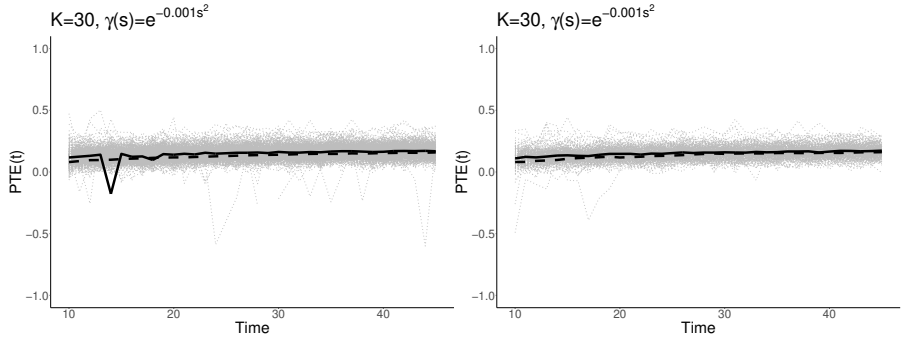


Figure 1: Results of the estimation of the function PTE(t) in the case of uniform unmeasured confounder (left panel) of binary (right panel).

Web Appendix E: Time-to-progression as a surrogate of overall survival in colorectal cancer

The second application was based on data from a randomized phase 3 multicenter trial evaluating the addition of panitumumab to a Folfiri-based chemotherapy in second-line metastatic colorectal cancer The included patients had all experienced progression after or during treatment with Fluoropyrimidine-based therapy. We examined time-to-progression (TTP), defined as the time from randomization to disease progression, as a putative surrogate for overall survival (OS).

Data description

The original study included data from 1186 randomized patients from a phase 3 multicenter trial and we had the data from 926 randomly selected patients. These 926 patients were from 151 centers. We used these centers as the basis for the meta-analytic approach. In this dataset, of the 926 patients, 461 were randomized to the experimental group receiving Panitumumab in addition to a Folfiri-based chemotherapy and 465 received Folfiri alone. The median follow-up time was 13.8 months (range 0.25-47 months) in the experimental group

and 12.3 months (range 0.4-50 months) in the control group. There were 417 observed deaths (90%) and 328 progressions (71%) in the experimental arm and 419 observed deaths (90%) and 319 progressions (69%) in the control arm. Kaplan-Meier estimates of survival for both endpoints are provided in Figure 2.

Data analysis and results

Estimations of the parameters of the model defined are given in Table 2.

Table 2: Estimated parameters for the application on second-line metastatic colorectal cancer.

Parameter	Est.	Std. Err.	p-value
$\beta_{Z,S}$	-0.266	0.135	0.048
$\beta_{Z,T}$	0.014	0.110	0.901
θ^2	1.617	0.453	< 0.001
ρ^2	0.089	0.082	0.280
$\sigma^2_{\nu_S}$	0.098	0.124	0.429
$\sigma^2_{\nu_T}$	0.001	0.006	0.937
$\sigma_{\nu_{S,T}}$	-0.007	0.043	0.877
ζ	0.797	0.194	< 0.001
α	0.723	0.468	0.122

There is no evidence of a direct effect of treatment on OS, as $\beta_{Z,T}$ is not significantly different from zero. Nevertheless, there is a direct effect of treatment on TTP, since $\beta_{Z,S}$ is significantly different from zero. The sign of $\beta_{Z,T}$ implies that the risk of TTP occurrence is reduced by the addition of Panitumumab to a Folfiri-based chemotherapy. Regarding the random effects variances, there is an association between TTP and OS at the individual level ($\theta^2 \neq 0$), but otherwise no evidence of heterogeneity in patients or treatment effects on TTP and OS between centers.

The estimation of $\gamma(s)$, PTE(t) and their associated natural effects can be found in Figure 3. From this figure, we can see that the estimated function is close to 0 overall. Thus, even if we have a direct effect of treatment on TTP, if

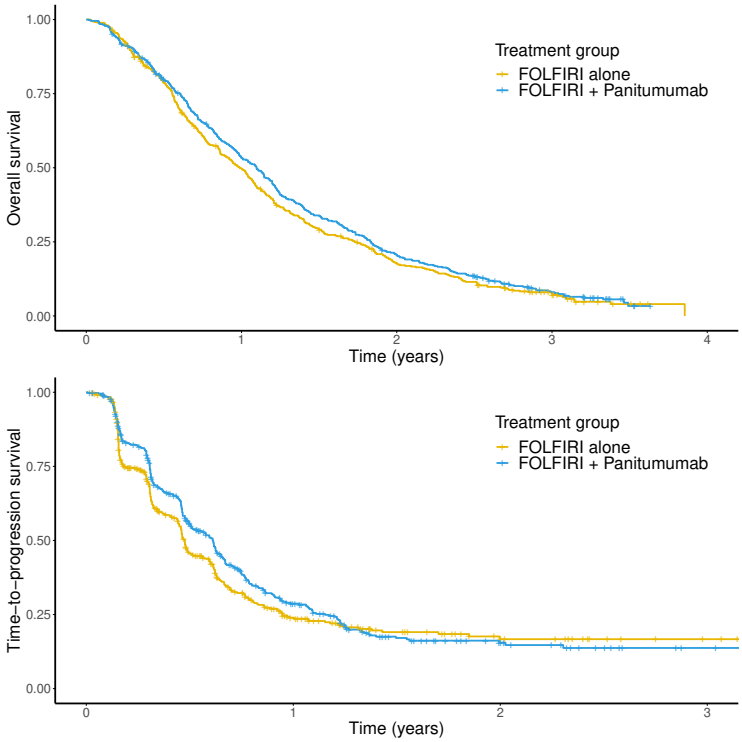
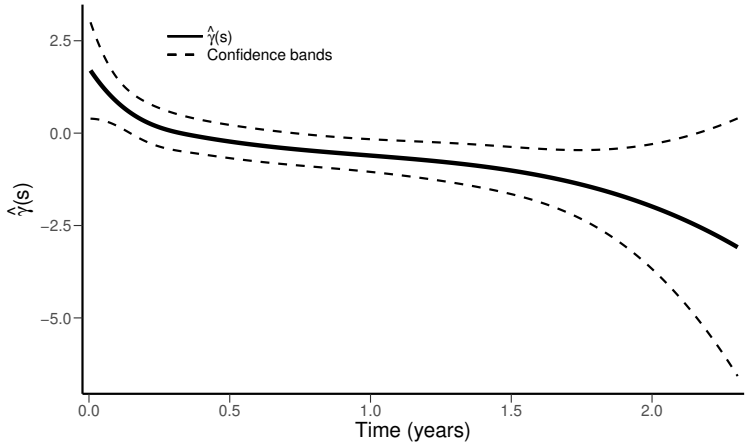
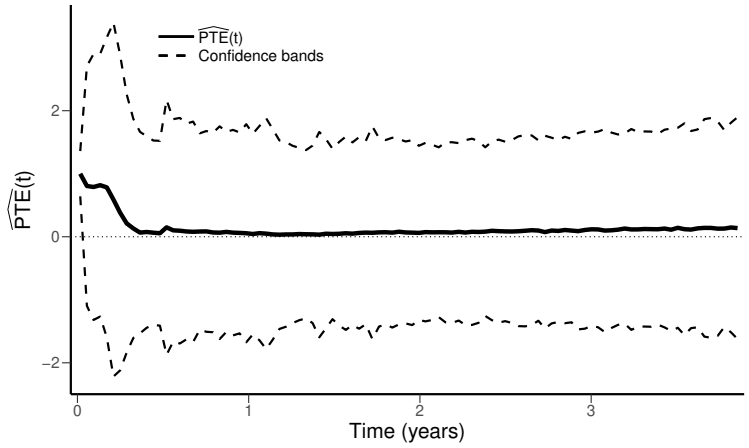


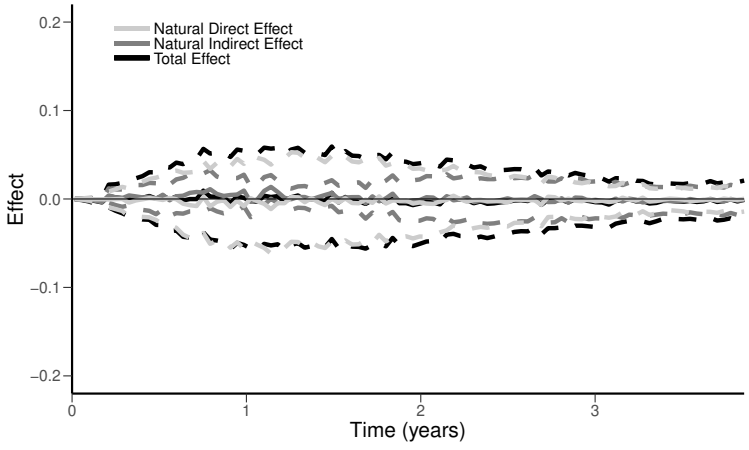
Figure 2: Kaplan-Meier estimates of overall survival (upper panel) and time-to-progression (bottom panel) by treatment group for the second-line colorectal cancer application.



(a) Estimated $\hat{\gamma}(s)$



(b) Estimated $\widehat{PTE}(t)$



(c) Estimated natural effects

Figure 3: Results of the application on colorectal cancer.

there is no effect of TTP on OS, this cannot lead to an indirect effect of treatment on OS through its effect on TTP. The graphical representation of the estimations of the natural direct, indirect and total effects shows that there is no evidence of a direct effect of the treatment on OS (which was expected from the estimated value of $\beta_{Z,T}$) but also no evidence of an indirect effect of the treatment. In the case of the absence of a total effect it may not be relevant to even consider computing the function PTE(t), but we have done so for illustrative purposes. Since the definition of PTE(t) is based on the ratio of the indirect effect over total effect, we may end up with a very erratic PTE(t) if the total effect is close to zero, which was the case here. Therefore we chose to report the average of all bootstrap estimators as the central estimate of PTE(t). We see that it is close to zero, and the wide confidence bands suggest that PTE(t) is indeed highly variable. Since there is no total effect of treatment on OS in this case, it is not even relevant to examine TTP as a potential surrogate.

Web Appendix F: Addtional figures

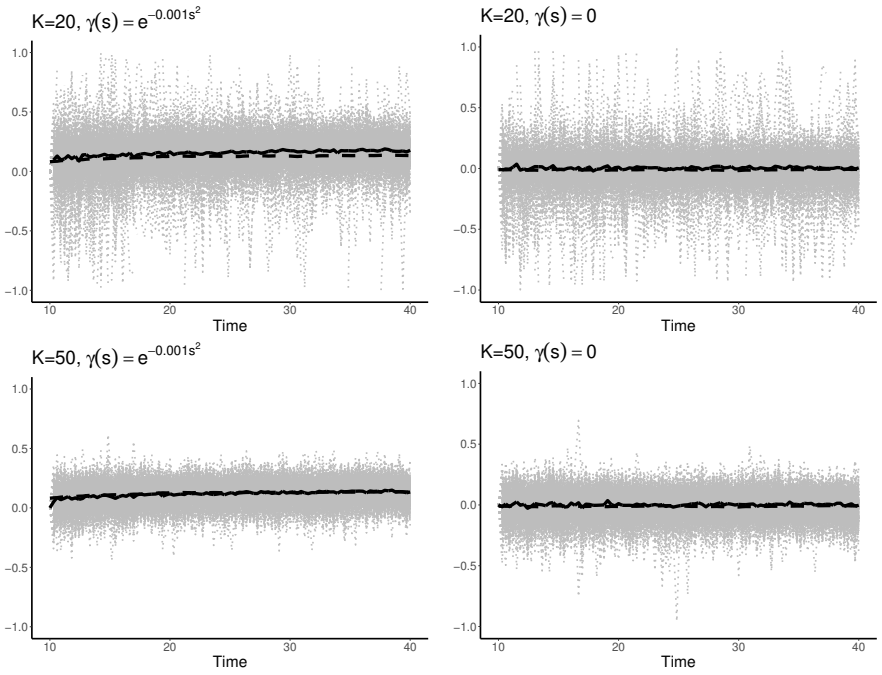


Figure 4: Computed PTE(t) for scenarios where $\theta^2 = \rho^2 = 0.64$ (dotted gray curves), their average (solid black curve) and the true curve (black dashed curve).

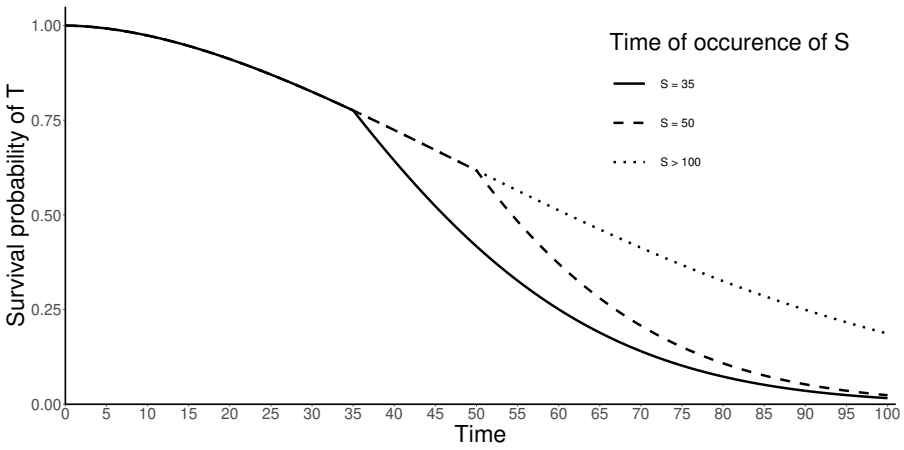


Figure 5: Example of the influence of S on the survival of T based on simulated data. In this case, ealier occurrences of S reduces the survival of T.

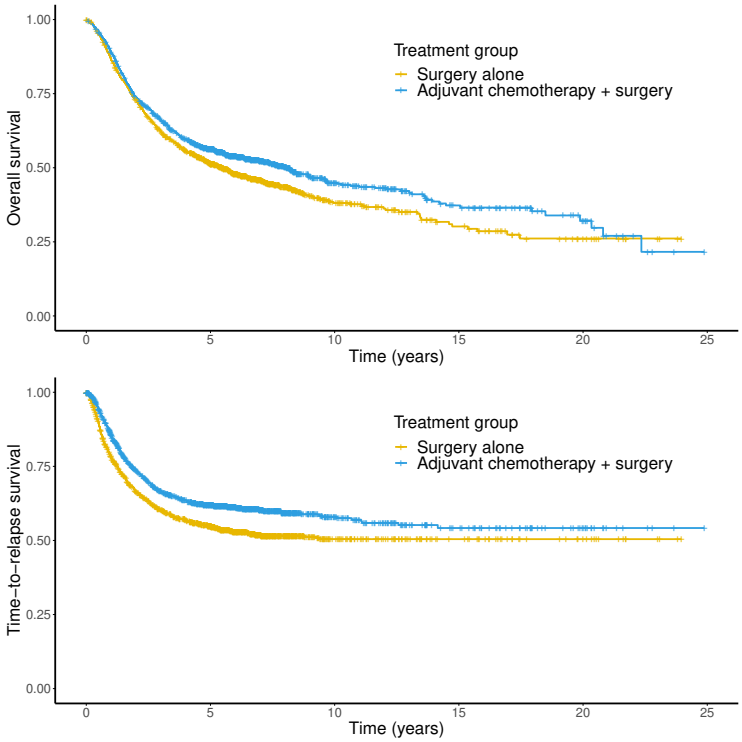


Figure 6: Kaplan-Meier estimates of overall survival (upper panel) and time-to-relapse survival (bottom panel) by treatment group for the resectable gastric cancer application.

III.3. | ADDTIONAL REMARKS

Composite endpoints

When both the surrogate and final endpoints are time-to-event, the surrogate might be a composite endpoint including the occurrence of the final endpoint as a composite event. For example in oncology the time-to-progression is defined as the time between treatment randomization and cancer progression or death whichever occurs first. Therefore, death is a composite event of the progression. From a mediation analysis viewpoint this raises a difficulty since the final endpoint always triggers the surrogate, which creates a "backward" causality. Moreover in our proposed approach, the occurrence of both the surrogate and the final endpoint at the exact same time could be interpreted as follows from a mediation perspective. In our proposed model, the risk of occurrence of T depends on the occurrence of S, therefore when both S and T occurs at the same time it means that the occurrence of S has instantaneously triggered the occurence of T which in turns means that the risk of occurrence of T has suddently taken an infinite value and thus $\gamma(S) = \infty$. Therefore, our method is not suited for surrogate endpoints that are composite endpoint including the final endpoint as an event. While this may appears as a limitation since many surrogates (or potential surrogate) used in practice are such composite endpoints, we may argue that using a composite endpoint as a surrogate is somewhat in contradiction with the purpose of a surrogate. If almost all the surrogate events are due to occurrences of the final endpoint, then the surrogate will be a good surrogate since highly correlated with the final endpoint, but of no use in practice to help reduce the duration of a trial. The use of a composite endpoint as a surrogate could still be of interest if the proportion of the surrogate events due to the final endpoint occurrence remains reasonable.

Software implementation

The proposed approach have been implemented in the R package `frailtypack` currently available on the CRAN. It is proposed as an extension of the function `jointSurroPenal` of the package. The function $\gamma(S)$ in the model is estimated using B-splines.

IV

Validation of a longitudinal marker as a surrogate using mediation analysis and joint modeling: evolution of PSA as a surrogate of the DFS

IV.1 Introduction 105
IV.2 Main article 106
IV.3 Additional remarks 134

IV.1. | Introduction

A second work of this thesis aimed at adapting the method proposed in the first work in the case where the time-to-event surrogate is replaced by longitudinal continuous surrogate. A novel joint model was developed in which a linear mixed model for the surrogate is defined, while the submodel for the time-to-event is still a proportional hazards model. Individual-level random effects are included in the longitudinal submodel to take into account the association between the repeated measurements while the trial-level random effects are used to take into account the heterogeneity of the treatment effects on the two endpoints. The two submodels are linked together through a link function, and we focused specifically on two types of link: the "current-level" and the "current-slope" links. This model is defined for counterfactual outcomes, from which a definition of the direct and indirect treatment effects on the final endpoint are derived. From these definitions, the proportion of the treatment effect on the final endpoint through the longitudinal surrogacy is taken as a measure of surrogacy. Given the time-to-event nature of the final endpoint, the natural effects are defined on a survival scale. As for the approach presented in Chapter III, identifiability assumptions are made in order to

give a causal interpretation of the estimated natural effects and how the random effects may help relax some stringent assumptions regarding the absence of unmeasured confounders. We also discuss the interpretation of the two types of link function from a mediation viewpoint. As previously, the joint model is estimated by maximazing a penalized log-likelihood using the Levenberg-Marquardt algorithm. The computation of the likelihood requires numerous integral approximation. For the integral over individual-level random effects we used a pseudo-adaptive Gauss-Hermite quadrature while for the integral over the trial-level random effects we used a Monte-Carlo approach. Moreover, the likelihood requires the evualution of the survival function, which itself require an integration over time. For this integral we used a Gauss-Kronrod quadrature rule. In order to reduce the computation time, the outer intregal (over the trial-level random effects) was implemented using `OpenMP` to support parallel computing through multithreading. We conducted a simulation study to evaluate the performances of our approach. We also applied it to a dataset of a multicenter phase III randomized trial in prostate cancer evaluating the benefit of a 24 months extension of a homornal therapy after radiotherapy. We investigated the evolution of PSA over time as a surrogate endpoint of the disease-free survival using the proportion of treatment effect as a measure of surrogacy. We studied both the current-level and current-slope link function. Results show that for the current-level case, a significant proportion of the treatment effect on the surrogate is mediated by the PSA (from 40% to 70%), while for the current-slope link this proportion was much less (around 3%). This indicates that the current-level of the PSA might be an interesting surrogate of the disease-free survival in this setting. This work is still under preparation as there are issues in the estimations of the standard-errors of the parameters of the model which are severly biased downward, resulting in very low coverage rate of the confidence intervals and were are investigating the causes of these bias.

IV.2. | MAIN ARTICLE

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Validation of longitudinal marker as a surrogate using mediation analysis and joint modeling: evolution of PSA as a surrogate of the DFS.

Quentin Le Coent*[1] | James Dignam[2] | Catherine Legrand[3] | Virginie Rondeau[1]

[1]Department of Biostatistics, Bordeaux Population Health Research Center, INSERM U1219, Université de Bordeaux, Bordeaux, France
[2]Department of Public Health Sciences, University of Chicago Biological Sciences, Chicago, IL, USA
[3]ISBA/LIDAM, UCLouvain, Louvain-la-Neuve, Belgium

Correspondence
Quentin Le Coënt, Bordeaux Population Health Center, INSERM U1219, Université de Bordeaux, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France. Email: quentin.le-coent@u-bordeaux.fr

Present Address
Present address

Abstract

Longitudinal biomarkers constitute a broad class of potential surrogates endpoints. Several approaches have been proposed for surrogate validation but available methods for validating a longitudinal biomarker as a surrogate of time-to-event endpoint such as death are limited. In this work we propose a method for validating a longitudinal outcome as a surrogate of a time-to-event endpoint using a combination of joint modeling and mediation analysis. The proportion of the total treatment on the time-to-event true endpoint due to its effect on the biomarker is used as a surrogacy measure. This method can incorporate meta-analytic data using a joint model with random effects at both the individual and trial levels. From this model, the indirect treatment effect through the surrogate as well as the direct and total treatment effects are derived using a mediation formula. A simulation study was designed to evaluate the performance of this approach. We applied this method to a multi-centric study on prostate cancer to investigate the use of prostate-specific antigen level as a surrogate for disease-free survival.

KEYWORDS:
Surrogacy; Biomarkers; Joint modeling; Mediation analysis

1 | INTRODUCTION

The use of surrogate endpoints in clinical trials is an important stake as it reduces both the duration and the cost of these trials compared to standard endpoints such as overall survival (OS), disease-free survival (DFS) or progression-free survival (PFS).

Longitudinal biomarkers are important class of potential surrogates because they assess the evolution of a disease over time. However, when evaluating a treatment, efficacy on clinical endpoints such as OS (which we will call final endpoint from now

on) will always be the main objective. For example, tumor size or prostate-specific antigen levels may indicates how a disease evolves and thus if the treatment is effective. In order for a trial conducted solely on the surrogate to be able to remains valid regarding the final endpoint, the surrogate must have been previously statistically validated.

Several methods have been proposed in this regard, following the seminal work of Prentice and the introduction of the well-known Prentice criteria[1] to validate a surrogate if the treatment effect on the final endpoint is null when conditioned on the surrogate. Soon after, Freedman[2] proposed the proportion of treatment on the final endpoint explained by the surrogate as a measure of surrogacy. Since then, the methods developed for surrogate validation, can be categorized in two main approaches[3]: methods validating surrogates through associations or based on identifications of the effects between the surrogate and the final endpoint.

For the first approach, the most commonly used method is the meta-analytic approach to assess the surrogacy by looking at the association between the surrogate and the final endpoint at two levels[4,5], considering first the association between the surrogate and the final endpoint at the individual level and then the association between the treatment effects on the surrogate and the final endpoint at the trial level. A putative surrogate will be validated if both associations are strong enough. Different models were proposed in that framework based on the nature of the surrogate and the final endpoint. Renard[6] was the first to adapt the two-stage approach of Buyse[5] to extend it to the case of the validation of a longitudinal potential surrogate biomarker for a time-to-event final outcome. In their work, the link between the surrogate and the final endpoint is made through shared random effects which are then used to derive the association between both outcomes at the trial level. The use of meta-analytic or (or multi-centric) is important as it brings heterogeneity in the validation process and thus strengthens it.

In the second approach, methods were proposed which put focus on identification of effects rather than associations: a surrogate will be validated if the proportion of the treatment effect on the final endpoint that goes through its effect on the surrogate is sufficiently large. Developments of this second approach have been made using mediation analysis. Frangakis[7] was the first one to use the causal inference framework to study a potential surrogate using principal stratification to define causal effects within subgroups of subjects having the same surrogates outcomes under the two treatment arms. Since then, causal approaches has been undertaken by several authors[8,9,10]. However, a limitation of these approaches is that they often use single-trial data. Also, the methods developed for estimating the proportion of treatment effect explained by a surrogate are often limited to cross-sectional biomarkers[2,11,12].

Joint models were proposed, outside the context of surrogate marker validation, to analyse the relationship between a biomarker process and a time-to-event outcome[13,14,15]. Liu[16] used such a joint model to examine the causal mechanism in the joint analysis of a longitudinal biomarker and survival and studied specifically the role played by the random effects. In this article we propose a joint model for the validation of a potential surrogate marker using mediation analysis where the potential surrogate is a longitudinal biomarker and the final endpoint is the time to a clinical event. A specificity of our approach is that it

is based on clustered data (meta-analytic or multicentric trial data) which brings heterogeneity in data of the validation process and thus strengthens it. Le Coent[17] used such a joint modeling approach to derive the proportion of treatment effect when both the surrogate and the final endpoint are right-censored time-to-event in a mediation analysis framework.

Recently, Zheng and Liu[18] proposed an approach to derive the indirect effect of the treatment on the final endpoint through the surrogate using a joint model combined with mediation analysis. In this approach the link between both outcomes is made by adjusting for the current level of the biomarker in the hazard function of the final endpoint. This link can be used to derive the effect of biomarker on the final endpoint and then the indirect treatment effect on the final endpoint through the surrogate biomarker. However, this approach does not consider the validation of a potential surrogate based on meta-analysis as the proposed approach was developed for the analysis of a single clinical trial.

Therefore, we extend the approach of Zheng and Liu[18] and define a new joint model for a longitudinal biomarker and a time-to-event endpoint that also takes into account clustered data from a meta-analysis (or a multicentric clinical trial). Our model includes random effects both at the individual level and at the trial level. Individual-level random effects represent individual deviation from the population in terms of intercept and slope in the longitudinal submodel while trial-level random effects are used to account for the heterogeneity of the treatment effects on both outcomes across trials (or centers). The association between the biomarker process and the final outcome is modeled through a link function in the survival submodel. The choice of the link function depends on the underlying biomarker process; examples of potential links that we will discuss in more details include the current-level or the current-slope of the biomarker process.

We will then demonstrate how we can use this new joint model to derive a decomposition of the total treatment effect on the final endpoint as an indirect effect through the biomarker and a direct effect. The (time-dependent) proportion of the indirect over the total effect of the treatment on the final endpoint can then be used as a measure of surrogacy. Our model will allow to estimate these quantities while acknowledging the meta-analytic nature of the data at hand and thus leading to a meta-analytic validation of surrogate markers.

This article is organised as follows. In Section 2, we introduce the proposed joint model and define the related direct and indirect treatment effect as well as the proposed surrogacy measure. In Section 3, we present a simulation study to investigate the performances of this method regarding the estimation of the parameters of the model as well as the measure of surrogacy. In Section 4 we apply this method to a multi-center study on prostate cancer to evaluate the prostate-specific antigen (PSA) evolution over time as a surrogate for disease-free survival. Finally, in Section 5, we conclude and provide some comments and further developments on the proposed methodology.

2 | METHODS

In the following we focus on the setting of a meta-analysis. Let K be the number of trials, K_i the number of patients in the ith trial and n_{ij} the number of repeated observations for the jth patient in the ith trial. For patient j in the ith trial, let Z_{ij} be the binary treatment assignment of subject ij where $Z_{ij} = 1$ denotes the experimental arm and X_{ij} a vector of covariates.

2.1 | Potential outcomes and mediation analysis

Let $M_{ij}^z(t)$ the potential biomarker process at time t for subject ij had he been assigned to treatment z. The whole potential biomarker path is denoted by $M_{ij}^z := \{M_{ij}^z(t), 0 \leq t \leq \tau\}$ where τ is the maximum follow-up time.

Hence subject ij has two potential biomarker paths, M_{ij}^1 and M_{ij}^0, of which only one can be observed.

Given treatment z and path m we define the potential event time T_{ij}^{zm} defined as the time-to-event for subject ij had he been assigned to treatment z and biomarker path m. In the same fashion we denote by C_{ij}^{zm} the potential censoring time. In most cases one can only control the value z of the treatment assigned to a subject but not the biomarker path m. Given the survival context, we observe $T_{ij}^* = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I_{T_{ij} \leq C_{ij}}$ be the follow-up time and censoring indicator of subject ij. From a survival analysis perspective we are interested by the survival function associated with the potential outcomes T_{ij}^{zm}:

$$\mathbb{S}^{zz'}(t) = \mathbb{P}(T_{ij}^{zM^{z'}} \geq t), \quad 0 \leq t \leq \tau.$$

From these functions we can derive the total treatment effect (TTE) on the survival scale as:

$$\text{TTE}(t) = \mathbb{S}^{11}(t) - \mathbb{S}^{00}(t),$$

which is the difference of survival when the treatment is set to 1 for both T and M versus when the treatment is set to 0. This total treatment effect can be decomposed into a natural direct effect (NDE) and an natural indirect effect (NIE):

$$\begin{aligned}\text{TTE}(t) &= \mathbb{S}^{11}(t) - \mathbb{S}^{00}(t) \\ &= \underbrace{\mathbb{S}^{11}(t) - \mathbb{S}^{10}(t)}_{\text{NIE}(t)} + \underbrace{\mathbb{S}^{10}(t) - \mathbb{S}^{00}(t)}_{\text{NDE}(t)}.\end{aligned}$$

Finally, the measure of surrogacy is the proportion of treatment effect (PTE), defined as the ratio of indirect effect over the total effect:

$$\text{PTE}(t) := \frac{\text{NIE}(t)}{\text{TTE}(t)}.$$

A surrogate will be validated if the value PTE(t) is large enough. While there is no acknowledged threshold, a value of PTE(t) larger than 0.8 could be considered enough for a surrogate to be validated. Moreover, PTE(t) is a function of time, but for practical purposes it could be more relevant to only consider specific time-points to validate a surrogate.

We now introduce a joint model adapted for the context of meta-analyses from which we can derive an expression to compute the functions $\mathbb{S}^{zm}(t)$ from the observed data.

2.2 | Joint model

As enonced previously, $M_{ij}(t)$ denotes the true value of the biomarker at time t. Let $\tilde{M}_{ij}(t)$ be the observed biomarker for suject ij. Using the notations defined before, the joint model for M_{ij}^{z} and T_{ij}^{zm} is

$$\begin{cases} \tilde{M}_{ij}^{z}(t) = \theta_{ij}' f(t) + \beta_M' X_{ij}^{M}(t) + (\beta_{Z,M} + \nu_{M,i})z + \varepsilon_{ij}(t) \\ \lambda_{ij}^{zM^z}(t) = \lambda_0(t) \exp\left(\left(\beta_{Z,T} + \nu_{T,i}\right) z + \beta_T' X_{ij}^{T}(t) + \eta' h\left(M^z\right) \right), \end{cases} \tag{1}$$

where $M_{ij}^{z}(t) = \theta_{ij}' f(t) + \beta_M' X_{ij}^{M}(t) + (\beta_{Z,M} + \nu_{M,i})z$ is the true, error-free, biomarker and $\varepsilon_{ij}(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is the independent measurement error. The random effects $(\nu_{M,i}, \nu_{T,i})$ are trial-level effects taking into account the heterogeneity of the treatment effect across trial and are assumed to be jointly gaussian:

$$(\nu_{M,i}, \nu_{T,i})' \sim \mathcal{N}\left(\mu_\nu, \Sigma_\nu\right), \quad \mu_\nu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; \Sigma_\nu = \begin{pmatrix} \sigma_{\nu_M}^2 & \sigma_{\nu_{M,T}} \\ \sigma_{\nu_{M,T}} & \sigma_{\nu_T}^2 \end{pmatrix}.$$

The vector $f(t)$ represents the temporal evolution of the biomarker and may be composed of several components. The associated vector $\theta_{ij} = \beta + \omega_{ij}$ is the sum of fixed effects (β) and individual random-effects (ω_{ij}) assiociated with each component of $f(t)$. For example if $f(t) = (1, t)'$ then $\theta_{ij} = \left(\beta_0 + \omega_{ij0}, \beta_1 + \omega_{ij1}\right)'$. The vector ω is assumed to be gaussian with mean 0 and unstructured covariance matrix Σ_ω.

The function $\lambda_0(t)$ is the baseline hazard function that is estimated in a parametric but flexible way using cubic M-splines [19]. The parameters $\beta_{Z,M}$ and $\beta_{Z,T}$ are the fixed treatment effects on the biomarker and the time-to-event respectively. The terms $X_{ij}^{M}(t)$ and $X_{ij}^{T}(t)$ are (possibly time-dependent) covariates included on the biomarker and time-to-event submodel respectively, with their associated vector of coefficients β_M and β_T. These covariates are not necessarily mutually exclusive and can share the same components. The function $h(\cdot)$ is the link function between the longitudinal marker and the event process. Common choices of link functions are $h(\cdot) = (\omega_0, \omega_1)$ (shared random effects), $h(\cdot) = M(t)$ (current value of the biomarker) and $h(\cdot) = \frac{\mathrm{d}M(t)}{\mathrm{d}t}$ (current slope). Parameters η are association parameters between the biomarker and the event time process. From a mediation analysis point of view, a mediated effect of the treatment on the final endpoint through the biomarker is possible only if the quantities in the link function depends on the treatment. In the following we focus mainly on the current value and current slope link functions. Figure 1(b) represents the case with a current-level link function and the associated parameters of model 1 while Figure 1(c) concerns the case of the current-slope link in which the arrow between Z and $\frac{\mathrm{d}M(t)}{\mathrm{d}t}$ only exists if the treatment as an effect on the slope for example by including and interaction between time and treatment (whose effect is represented by the

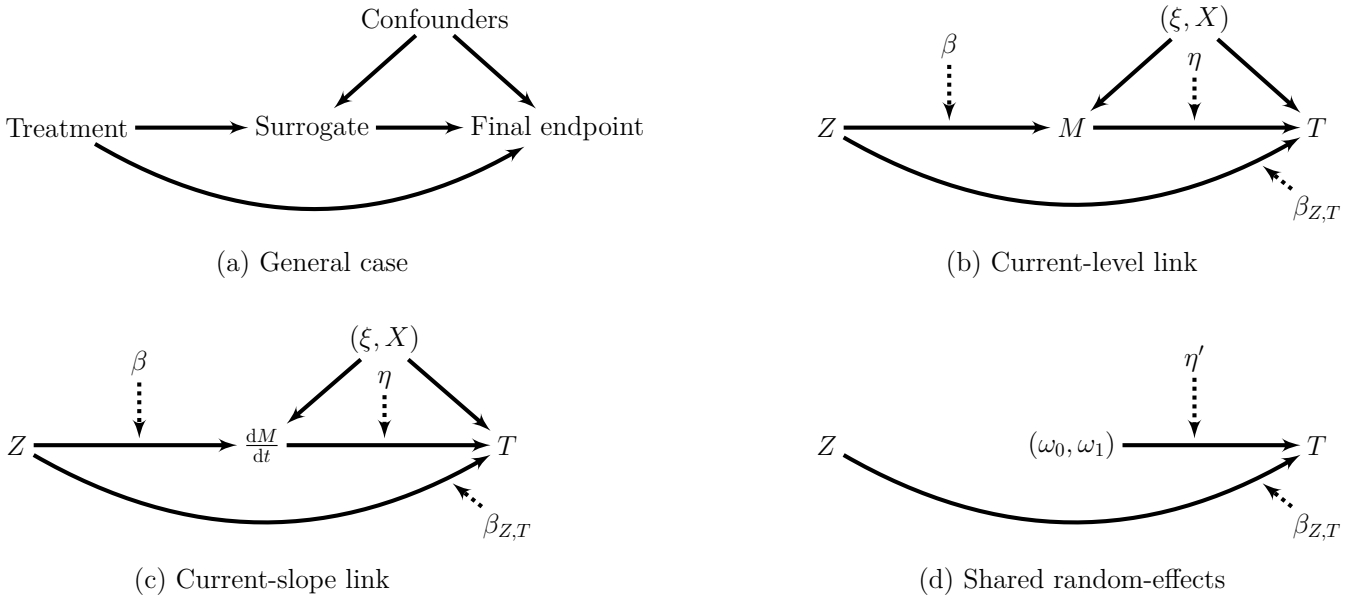


Figure 1 Pathways of mediation of the treatment on the final endpoint T in: the general setting (a), the "current-level" link function (b), the "current-slope" link (c) and the "shared random-effects" case (c) in the joint model 1. The solid lines represent the link between the outcomes while the dotted lines represent the strength of these links and the paramters in the model that control this strength.

term $\beta_{Z\times t}$ in Figure 1(c)). The case with shared random effects does not allow a mediated effect since the individual random effects are independent of the treatment (Figure 1(d)).

Letting ϕ being the whole vector of paramaters of the model, i.e.,

$$\phi = (\lambda_{0,T}, \eta, \sigma_\varepsilon, \Sigma_\omega, \Sigma_\nu, \beta, \beta_T, \beta_M, \beta_{Z,T}, \beta_{Z,M}),$$

Estimator of ϕ, $\hat{\phi}$, is obtained by maximizing the penalized likelihood using the LevenbergMarquardt algorithm[20,21]:

$$L_{\text{pen}}(\phi) = L(\phi) - \kappa \int_0^\infty \left(\lambda''_{0,T}(t)\right)^2 \mathrm{d}t, \tag{2}$$

where $L(\phi)$ is the full likelihood of the data and $\kappa \int_0^\infty \left(\lambda''_{0,T}(t)\right)^2$ is a penalization term on the second order derivative of the baseline hazard function resulting in a smooth estimate. The larger κ is, the smoother the estimated function $\hat{\lambda}_0(t)$ will be. This parameter can be chosen using cross-validation. Details on the likelihood construction and computation are provided in Web Appendix A. Standard errors of the parameters of the model through the inverse of Hessian matrix of the penalized likelihood evaluated at $\hat{\phi}$. In the following we note by ξ the vector of the random effects in the model (both individual and trial-level random effects). The main difficulty in the maximization of the (log) likelihood comes from the integration over the random-effects. We have chosen a Monte Carlo approach for integrating over the trial-level random effects and a pseudo-adaptive gauss-hermite rule for integrating over the individual-level random effects[22].

2.3 | Derivation of mediation measurements

In order to derive quantities related to mediations measurements and the potential outcomes distributions, we rely on some assumptions[23].

First the stable unit treatment value assumption (SUTVA) requires that each version of the treatment is well defined and that there is no interference between an individual treatment assignment and another's outcome. The first part SUTVA is reasonable for randomized clinical trials in which each version of the treatment is defined without ambiguity. The second part of the SUTVA may not hold in particular settings, even in randomized clinical trials. For example in infectious diseases, if a patient is in contact with another whose likelihood of getting infected depends on the treatment, then the potential outcome of the this patients not only depends on the treatment he/she receives but also on the treatment the other patient receive. However for other settings such as oncology in our case the second part of the SUTVA is likely to be met.

We also require the consistency assumption which states that, for subject ij, the biomarker process M_{ij} equals the potential process associated with the treatment actually assigned to this patient, $M_{ij} = M_{ij}^{Z_{ij}}$. In the same way, the event time T_{ij} equals $T_{ij}^{Z_{ij}M_{ij}}$ as well as the censoring time C_{ij} is $C_{ij}^{Z_{ij}M_{ij}}$.

We want to point out that generally, the consistency assumption requires that the biomaker process M is measured without error. However, this is likely not to be the case and in a joint perspective we are interested by modeling the true biomaker process which is the observed process plus an independent error term. Hence we do not require the "no measurement-error" assumption but rather assume that the model for the true process is correctly specified. We also make an adapted version of the sequential ignorability assumption[24,17]:

- $(T_{ij}^{zm}, M_{ij}^{z}, C_{ij}^{zm}) \perp Z_{ij} \mid (X_{ij}, \xi_{ij}), \forall z, z', m,$
- $(T_{ij}^{zm}, C_{ij}^{zm} \perp M_{ij}^{z} \mid (Z_{ij}, X_{ij}, \xi_{ij}), \forall z, z', m.$

This version the sequential ignorability states that all the confounding between the biomarker and the time-to-event should be captured by either the covariates and/or the random effects. The use of random effects thus make this assumption more flexible are they can capture the confounding not taken into account by the covariates. Recall from Section 2.1 that we are interested in computing $\mathbb{S}^{zz'}(t)$ which is the survival function of T for Z set to z for M and to z' for T. Using the SUTVA and the consistency assumption we have, for covariates x[25],

$$\mathbb{S}^{zz'}(t \mid x) = \int_{\xi} S_T\left(t \mid h(M^{z'}), \xi, Z = z, X = x\right) f_{\xi}(\xi) \mathrm{d}\xi, \tag{3}$$

where $S_T\left(t \mid h(M^{z'}), \xi, Z = z, X = x\right)$ is the survival function of T conditional on $h(M)$, the treatment Z, covariates X and random effects ξ:

$$S_T\left(t \mid h(M^{z'}), \xi, Z = z, X = x\right) = \exp\left[-\int_0^t \lambda_{0,T}(u)\exp\left((\beta_{Z,T} + \nu_T)z + \beta_T' X^T(u) + h(M^{z'})\right)\mathrm{d}u\right].$$

The function $f_\xi(\xi)$ in Formula 3 is the joint density of the random-effects.

Formula 3 involves an integration over the distribution of the random-effects ξ which are assumed gaussian that can be carried out using either a Monte-Carlo method or gaussian quadrature. It should be noted that according to the model 1, the quantity involved in $h(\cdot)$ only depends on the covariates and the random effects. Hence, as oppposed to Zheng and Liu [18] who assume that the error term ϵ is included in the argument of the link function h, we do not need to integrate over the distribution of ϵ once we have conditioned on z, x and ξ.

From $\mathbb{S}^{zz'}(t \mid x)$ we can derive the direct and indirect effect of the treatment for a given level of covariates x as:

$$\mathrm{NIE}(t \mid x) = \mathbb{S}^{11}(t \mid x) - \mathbb{S}^{10}(t \mid x),$$
$$\mathrm{NDE}(t \mid x) = \mathbb{S}^{10}(t \mid x) - \mathbb{S}^{00}(t \mid x).$$

In order to obtain these effects for the overall population we can integrate over the covariates to yield

$$\mathrm{NIE}(t) = \int_X \left(\mathbb{S}^{11}(t \mid x) - \mathbb{S}^{10}(t \mid x)\right) f_X(x)\mathrm{d}x,$$
$$\mathrm{NDE}(t) = \int_X \left(\mathbb{S}^{10}(t \mid x) - \mathbb{S}^{00}(t \mid x)\right) f_X(x)\mathrm{d}x.$$

This integration can be approximated by computing $\mathrm{NIE}(t \mid x)$ and $\mathrm{NDE}(t \mid x)$ for each individual in the dataset and averaging them all. In the case were the covariate are categorical we only need to compute the NIE and NDE for each category and then take their sum weighted by the proportion of each category in the data. From the estimated parameters $\hat{\phi}$ of the model we can thus derive estimates of the indirect and direct treatment effect, $\widehat{\mathrm{NIE}}(t)$ and $\widehat{\mathrm{NDE}}(t)$ respectively, and compute an estimation of $\mathrm{PTE}(t)$:

$$\widehat{\mathrm{PTE}}(t) = \frac{\widehat{\mathrm{NIE}}(t)}{\widehat{\mathrm{NIE}}(t) + \widehat{\mathrm{NDE}}(t)}.$$

Estimated standard error and confidence bands for the function $\hat{\mathrm{PTE}}(t)$ can be obtained using parametric boostrap based on the asymptotic gaussian distribution of the maximum likelihood estimator $\hat{\phi}$. By sampling ϕ from this distribution and computing $\mathrm{PTE}(t)$ for each draw we can obtain percentile-based confidence bands, e.g by using the 2.5^{th} and 97.5^{th} percentiles of the obtained $\mathrm{PTE}(t)$ at each time t.

3 | SIMULATIONS

3.1 | Scenarios and data generation

We ran a simulations study to evaluate the finite sample performances of our proposed approach . The parameters estimation of the proposed model, direct effect, indirect effect and PTE(t) were investigated in terms of bias, mean estimated standard error, standard deviation and coverage rate of the 95% confidence intervals. Since the direct effect, indirect effect and PTE(t) are functions of time we chose three timepoints at which they are evaluated. A graphical representation of the estimations of these quantities as functions of time is also provided. We focused on cases where the link function $h(\cdot)$ was either the "current-level" or the "current-slope" associations. We also considered a case where there is no link between the longitudinal biomarker and the time-to-event process. For each scenario we simulated datasets with a number of trials equal to either 20 or 50. For each scenario the number of subjects in the dataset was fixed and set to 800. Each patient was randomly assigned to one of the K trials (uniform draw). The treatment Z was randomly generated with $\mathbb{P}(Z = 1) = \mathbb{P}(Z = 0) = \frac{1}{2}$. The assumed simulated model is given by:

$$\begin{cases} \tilde{M}_{ij}^{z}(t) = \theta_{ij}^{'} f(t) + +(\beta_{Z,M} + \nu_{M,i})z + \beta_2 z \times t + \beta_M^{\prime} X_{ij}^{M} + \varepsilon_{ij}(t) \\ \lambda_{ij}^{zM^z}(t) = \lambda_0(t) \exp\left(\left(\beta_{Z,T} + \nu_{T,i}\right) z + \beta_T^{\prime} X_{ij}^{T} + \eta^{\prime} h\left(M^z\right) \right), \end{cases} \tag{4}$$

where,

$$f(t) = \begin{pmatrix} 1 \\ t \end{pmatrix}, \quad \theta_{ij} = \begin{pmatrix} \beta_0 + \omega_{ij0} \\ \beta_1 + \omega_{ij1} \end{pmatrix}.$$

Parameters β_0, β_1 and β_2 are respectively the fixed intercept, slope and the effect of the treatment on the slope (see used values in Tables 1). The individual-level random effects ω_{ij} were generated from a bivariate gaussian distribution with mean 0 and covariance matrix

$$\begin{pmatrix} \omega_{ij0} \\ \omega_{ij1} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_\omega), \quad \Sigma_\omega = \begin{pmatrix} \sigma_{\omega_1}^2 & \sigma_{\omega_{12}} \\ \sigma_{\omega_{12}} & \sigma_{\omega_2}^2 \end{pmatrix}.$$

The trial level random effects were simulated from a centered bivariate gaussian distribution with mean 0 and covariance matrix

$$\begin{pmatrix} \nu_{M,i} \\ \nu_{T,i} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_\nu), \quad \Sigma_\nu = \begin{pmatrix} \sigma_{\nu_M}^2 & \sigma_{\nu_{M,T}} \\ \sigma_{\nu_{M,T}} & \sigma_{\nu_T}^2 \end{pmatrix}.$$

The covariates X_{ij}^1 and X_{ij}^2 included one binary variable X with $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 0.5$. The link functions used were defined as follow:

- For the "current-level" scenarios: $h(M_{ij}^z) = \mathbb{E}(M_{ij}^z(t) \mid \omega_{ij0}, \omega_{ij1}, \nu_{M,i}) = \beta_0 + \omega_{ij0} + \left(\beta_1 + \omega_{ij1}\right) t + (\beta_{Z,M} + \nu_{T,i})z + \beta_2 z \times t + \beta_M^{\prime} X_{ij}$,
- For the "current-slope" scenarios: $h(M_{ij}^z) = \beta_1 + \omega_{ij1} + \beta_2 z$.

Follow-up times and censoring indicators were simulated using a permutation algorithm[26] resulting in approximately 20% of censoring. The average number of repeated measurements per subject was ≈ 6.2. Regarding the surrogacy measure, we focused on two time points t_1 and t_2 corresponding to approximately 25% and 50% of survival. The true values for PTE$(\cdot)$, NIE$(\cdot)$ and NDE$(\cdot)$ were estimated by simulating datasets in which the treatment was fixed to either 1 or 0 for T and M in order to estimate the functions $S^{11}(\cdot)$, $S^{00}(\cdot)$ and $S^{10}(\cdot)$. For example $S^{10}(t)$ was estimated by simulating a dataset where the treatment was fixed to 1 for T and to 0 for M.

3.2 | Simulations results

Preliminary results regarding the scenarios with a "current-slope" link are shown in Table 1. While the estimations of the model parameters are quite correct in terms of bias, these results show a severe underestimation of the standard-deviations of the model estimates, especially for the parameters of the longitudinal submodel. These underestimations result in poor coverage rates of the 95% confidence intervals and reflect on the surrogacy measures derived from the model. This suggests that there is an issue in either the estimating process or the data simulation process. We are currently investigating the source of this issue by either modifying the data generation process or the methods used in the computation of the likelihood but we haven't yet find a solution.

4 | APPLICATION

We applied the proposed approach to a multicenter study on locally advanced prostate cancer[27]. We investigated the evolution of PSA over time as a surrogate of the disease-free survival. In this application the centers (institutions) were used as clustering units for the meta-analytic approach.

Armstrong et al[28] investigated the PSA as a surrogate of the overall survival in the context of metastatic prostate cancer. In this article, the authors used the proportion of treatment effect on the overall survival going through the treatment effect on a PSA decline $\geq$ 30% at 3-months. They concluded that this proportion is approximately two third. However their analysis is limited by the use of a single-time value (decline/non decline) and the PTE is estimated from fitting two separate Cox models on time-to-death (one using only the the treatment arm and the other adjusting on both the treatment arm and the surrogate). The proportion of treatment effect is taken as one minus the ratio of coefficient of the treatment effects in the second (adjusted model) versus that in unadjusted model. Since the surrogate is a posttreatment variable likely to be influenced by the treatment, raw adjustment on the surrogate creates a selection bias by not taking into account potential confounders between the surrogate and the final endpoint which may result in a lack of causal interpretability. Moreover, taking estimated regression coefficients as measures of "effects" renders difficult the interpretation of the proportion of treatment effect on other scale, such as the survival scale.

Data description

The trial evaluated the benefit of adding a long-term androgen deprivation therapy (LTAD) for patients with T2c-T4 prostate cancer undergoing radiotherapy. Patients were randomized to either radiotherapy with a short-term androgen deprivation therapy (STAD arm) or to radiotherapy with an addionnal 24 months of ADT (LTAD arm). The final endpoint is the disease-free survival defined as the time to occurence of a local progression, distant metastasis, a biochemical failure or death, whichever occured first. Biochemical failure was defined as either three consecutive PSA rises after a post-treatment nadir, receiving additionnal ADT or a PSA greater than 4ng/mL, whichever occured first[27]. For practical purposes, we focused on an analysis censored at 5 years even if the original dataset some patients had a longer follow-up.

Out of 1423 patients, 703 were randomized to the STAD arm and the remaining 716 patients were included in the LTAD arm. This study was a multicentric trial with 117 centers that we used as the basis for our meta-analytic approach. The number of observed events was 495 in the LTAD arm and 563 in the SDAT arm. Table 2 provides a description of the patients characteristics and follow-up in each treatment arm. Previous results showed that at 5 years, the LTAD arm was superior to STAD regarding the disease-free survival (46.4% vs 28.1%, $p < 0.001$)[29,27]. Kaplan-Meier estimates of the disease-free survival for each treatment arm are available in Figure 1 of the Web Appendix C.

Data analysis and results

According to previous longitudinal analyses of PSA, we used a log transformation of the PSA in the longitudinal submodel instead of the raw value of PSA[30,31,32,33]. This is mainly due to the “exponential” increase of the PSA. Figure 2 shows the average (across patients) trajectory of the log-transformed PSA over time for the STAD arm (dashed curve) and the LTAD arm (dotted curve). The shape indicates a sharp decrease in PSA after baseline (due to patients being treated with radiotherapy) and an almost linear increase of the PSA during the following years. From this figure we can see that the differences in trajectories between both group may mainly lies in different levels of PSA rather than in different slopes, suggesting that the "current-level" association may be more relevant from a surrogacy point of view than the slope (since the treatment may have no effect on the slope).

To take into account the specific evolution of PSA over time we included two functions of time to modelize both the sharp decrease of the PSA following baseline, $f_1(t)$, and the linear increase in the long-term, $f_2(t)$. According to previous analyses[30,33] we used functions of the following form: $f_1(t) = (1+t)^{\alpha} - 1$ and $f_2(t) = \frac{t^{\lambda+1}}{(1+t)^{\lambda}}$. The parameters α and λ were estimated using profile likelihood on mixed models on the PSA evolution through a grid search which yielded $\alpha = -11.4$ and $\lambda = -0.3$. The resulting estimated curve is shown in Figure 2 (solid curve). Prior analysis on a reduced model based only on the longitudinal PSA suggested that of the three main components of the PSA trajectories (intercept, $f_1(t)$ and $f_2(t)$), there was an individual

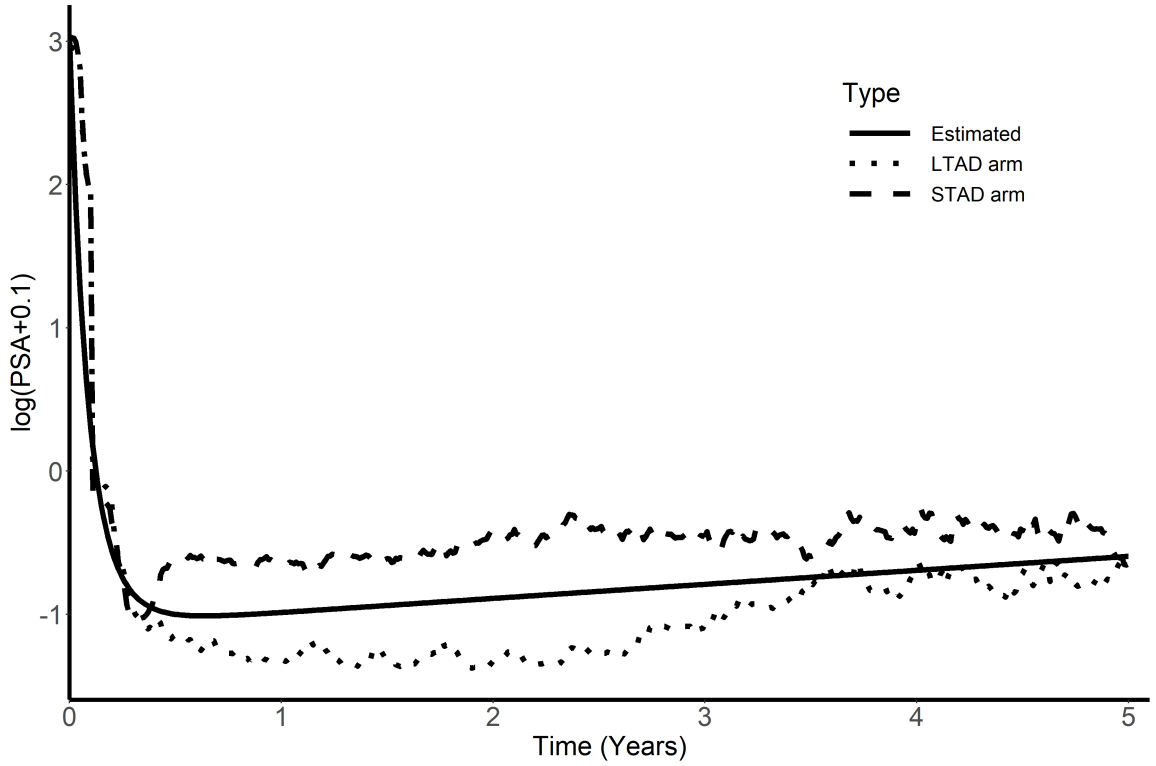


Figure 2 Average PSA trajectory (in $\log(\cdot + 0.1)$ scale) in both treatment arms (dotted and dashed lines) and the estimated curve based on a mixed model (solid line).

heterogeneity only on the intercept and $f_2(t)$. As results, in the joint model we only included individual-random effect for these two components. We used the following joint model:

$$\begin{cases} \log\left(PSA + 0.1\right) = \theta_{ij}^{'} f(t) + \beta_M' X_{ij}^M(t) + (\beta_{Z,M} + \nu_{M,i}) z + \varepsilon_{ij}(t) \\ \lambda_{ij}^{zM^z}(t) = \lambda_0(t) \exp\left(\left(\beta_{Z,T} + \nu_{T,i}\right) z + \beta_T' X_{ij}^T(t) + \eta' h\left(M^z\right)\right). \end{cases} \tag{5}$$

The covariates X_{ij}^M and X_{ij}^T included are the Gleason score and the clinical stage. Here we have $f(t) = \left(1, f_1(t), f_2(t)\right)'$. We estimated two models: one with a "current-level" link function and another with a "current-slope" link. In the latter case, from a mediation viewpoint, in order to enable a potential mediated effect the slope must be dependent on the treatment Z. Therefore, we added an interaction term (in both models) between the treatment and the long-term increase of the PSA represented by $f_2(t)$. Hence, $\theta_{ij}' f(t)$ is as follows:

$$\theta_{ij}' f(t) = \begin{pmatrix} \beta_0 + \omega_{ij0} \\ \beta_1 \\ \beta_2 + \omega_{ij2} + \beta_3 Z_{ij} \end{pmatrix}' \times \begin{pmatrix} 1 \\ f_1(t) \\ f_2(t) \end{pmatrix},$$

whith $(\omega_{ij0}, \omega_{ij2})' \sim \mathcal{N}\left(0, \Sigma_\omega\right)$ the individual-level random effects.

The two link functions used are:

$$h\left(M_{ij}^{Z_{ij}}\right) = \theta_{ij}' f(t) + \beta_M' X_{ij}^M(t) + (\beta_{Z,M} + \nu_{M,i}) Z_{ij} \quad \text{current-level case,}$$
$$h\left(M_{ij}^{Z_{ij}}\right) = \beta_1 f_1(t) + (\beta_2 + \omega_{ij2} + \beta_3 Z_{ij}) f_2'(t) \quad \text{current-slope case.}$$

Estimated results are shown in Table 3. For the evaluation of PTE(t) we selected three time points, respectively at 1, 3 and 5 years. From Table 3 we see that in both the current-level and current-slope cases there is a significant treatment effect on both the longitudinal evolution of the PSA and the final endpoint. The treatment significantly reduces the average level of PSA ($\beta_{Z,M}$) but also its long-term increase (parameter β_3). The estimated variances of trial-level random effects (σ_M and σ_T) suggest that there is no heterogeneity of the treatment effects on both endpoints accross centers. This is not surprising since there centers were all part of the same clinical trial and thus likely followed all the same treatment procedures and follow-ups, vanishing out any heterogeneity of the treatment effects. From both the current-level and current-slope link, the association (η) between the longitudinal evolution of PSA over time and the disease-free survival process is significant. Therefore, a significant effect of the treatment on PSA and a significant effect of the PSA on the DFS (either by its current value or slope) suggests that there is a mediated effect. However, the magnitude of this mediated effect has to be compared to the direct treatment effect on the event time process (parameter $\beta_{Z,T}$). Moreover, even if the association is significant in both cases, the association in the current-value case is much stronger than in the current-slope case ($\eta = 1.336$ versus 0.106). Therefore, the mediated effect is more important for the current-level link which can be seen in the estimated PTE. For the current-level link, PTE ranges from 0.472 to 0.762 while for the current-slope it ranges from 0.024 to 0.034. The evolution of PTE over time is given in the top panel of Figure 3, with the PTE for the current level and slopes links on the left and right respectively. The associated estimated natural direct, indirect and total effects are represented in the bottom panel. The confidences bands for these functions were obtained using parametric bootstrap with 2000 sampling. These results suggest that the slope (even if significantly associated with the time-to-event process) is a poor surrogate as it mediates less than 3% of the total treatment effect on the disease-free survival. On the contrary the current-level seems to be a more interesting surrogate which mediates a much larger proportion of the treatment effect, with a trend over time suggesting that this proportion increases. This increase over time might be explained by the significant interaction between the treatment and the long-term increase of the PSA, $f_2(t)$. Since $f_2(t)$ is increasing, this interaction have more weight over time which might explain the increasing PTE over time seen in Figure 3. These results suggests that the current-level of PSA might be an interesting surrogate of the disease-free survival, especially for long-term benefit. One possible extension of our approach is to combine the current-level and the current-slope as surrogates of the time-to-disease. This approach would combine both, using two associations parameters η_1 and η_2, but might complicate the computation and the interpretation of the results in terms of surrogacy and mediation. In a joint model combining both the current level and slope interpreting which component influences the proportion of treatment effect might be difficult, especially if there is a trending

evolution over time. Moreover, for practical purposes we limited our analysis up to a 5 years follow-up, but it would be of course interesting to investigate further the long time surrogacy of the evolution of the PSA on the disease-free survival.

5 | DISCUSSION

In this paper we proposed an approach to validate a longitudinal biomarker as a surrogate of a time-to-event using mediation analysis and joint modeling. An advantage of the proposed method is that it can be applied to meta-analytic data while most of the developments for surrogate validation using mediation analysis were restriced to unclustered data. The use of meta-analysic data introduces heterogeneity (and especially heterogeneity of the treatment effects) within the validation process and helps strenghthen it.

When studying a longitudinal biomarker as a surrogate of a time-to-event, the trajectory of the biomarker process must be carefully modeled and the nature of the association between the two outcomes should be well thought. Especially, one should consider which component of the trajectory is the most relevant from a surrogacy point of view. For example, is a sudden increase of the biomarker more informative than its raw value? If so, including the slope of the biomarker in the link function may be more relevant than its current level. In this article we considered two options: either the value itself or the slope but a combination of the two is also possible. These are single-time values (the slope or the value at time t) but an aggregate over time may also be relevant. For example the cumulative biomarker values $\int_0^t M(u)\mathrm{d}u$ have been previously proposed as an association term in joint modeling [15].

Regarding the trajectory of the biomarker over time, we used a simple linear mixed model but other approaches based on mechanistics model have been proposed in the context of joint modeling [34].

However, we have to be aware that from a mediation analysis perspective the association term included in the link function should depends on the treatment. For example, a comon used link function in joint models is the shared random-effects link where the individual random intercept and slope (ω_0 and ω_1) in the notations of our model) are included in the survival submodel. These random effect are independent of the treatment, thus no matter the treatment effect on the biomarker, there will be no mediated treatment effect on the event time process through the random effects. Similarly, in order for a mediation to be possible in the case where the slope is included as the association term, the slope should depends on the treatment effect. The magnitude of this effect will yield either a strong or weak mediated effect, but not allowing the association term to be influenced by the treatment will lead, by model definition, to the absence of mediation.

In this article we mainly focused on deriving the proportion of treatment effect on the final endpoint going through its effect on the biomarker using mediation analysis but as stated in the introduction other approaches have been proposed [35]. These approaches use meta-analytic data to derive association between the surrogate and the final endpoint at the individual-level as

well as at the trial level through the use of random effects. It should be noted that even if we focused on mediation, these two-level assiotations can also be computed from our proposed model. The trial level association can be derived from the covariance matrix of the trial level random effects Σ_ν as the coefficient of determination $R^2_{\text{trial}} = \frac{\sigma^2_{\nu_{M,T}}}{\sigma^2_{\nu_M}\sigma^2_{\nu_T}}$. One possibility to look at the individual-level assiocation is to compute the Kendall's τ between the biomarker M and the event-time T. In the case of shared-random effects the Kendall's τ only depends on the distribution of the individual-level random effects and can be computed[36]. However, when the link function includes either M or its slope the derivation of a formula for the Kendall's τ may be more complicated as this link should be taken into account. Besides these measures of associations, another measure of surrogacy proposed is the surrogate threshold effect. Its purpose is to predict the treatment effect on the final endpoint from its estimated effect on the surrogate[37]. In the setting of shared-random or correlated random effects, the surrogate and final endpoint are assumed independent given the random effects. This can be translated as removing the arrow from M to T in Figure 1a. In that case the total treatment effect on the final endpoint is $\beta_T + \nu_T$ and one needs only to focus on predicting it from the observed effect of the treatment on the surrogate. However, when there is a direct link between the biomaker and the final endpoint, predicting the total effect of the treatment on the final endpoint requires to also take into account the effect of the biomaker on the final endpoint which renders the computation more complicated.

In our application, DFS is a composite endpoint defined as the earliest of either death, biochemical failure, distant metastasis or local recurrence. In particular, the definition of a biochemical failure is based on the evolution of the PSA, which creates de facto a link between DFS and the biomarker. It should be assessed if studying surrogacy in cases where the final endpoint definition is partly based on the biomarker is relevant. In our case it should also be assessed if the use of biochemical failures as composite events has more influence in a model based on the current level or the current slope of the PSA. In the definition of a biochemical failure, both the current value of the PSA (value greater than 4ng/mL) and slope (three consecutive rises) intervene so that it seems likely that the use of biochemical failures influences both cases.

Finally, in this article we focused on validating a single biomaker as a surrogate using a joint model, but joint modeling has been extended in order to take into account the effect of several biomarkers on a time-to-event[38]. It would be an interesting to extend this approach in the case of several biomarkers.

ACKNOWLEDGEMENTS

This PhD program is supported within the framework of the PIA3 (Investment for the future). Project reference 17-EURE-0019. Computations for this work were performed using HPC resources from the Mesocentre de Calcul Intensif Aquitain (MCIA).

References

1. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine* 1989; 8(4): 431–440.

2. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Statistics in medicine* 1992; 11(2): 167–178.

3. Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics* 2009; 65(2): 530–538.

4. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in medicine* 1997; 16(17): 1965–1982.

5. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000; 1(1): 49–67.

6. Renard D, Geys H, Molenberghs G, et al. Validation of a longitudinally measured surrogate marker for a time-to-event endpoint. *Journal of Applied Statistics* 2003; 30(2): 235–247.

7. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; 58(1): 21–29.

8. Zigler CM, Belin TR. A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint. *Biometrics* 2012; 68(3): 922–932.

9. Conlon AS, Taylor JM, Elliott MR. Surrogacy assessment using principal stratification when surrogate and outcome measures are multivariate normal. *Biostatistics* 2014; 15(2): 266–283.

10. Vandenberghe S, Duchateau L, Slaets L, Bogaerts J, Vansteelandt S. Surrogate marker analysis in cancer clinical trials through time-to-event mediation techniques. *Statistical methods in medical research* 2018; 27(11): 3367–3385.

11. Parast L, Cai T, Tian L. Evaluating surrogate marker information using censored data. *Statistics in medicine* 2017; 36(11): 1767–1782.

12. Wang X, Parast L, Tian L, Cai T. Model-free approach to quantifying the proportion of treatment effect explained by a surrogate marker. *Biometrika* 2020; 107(1): 107–122.

13. Wulfsohn MS, Tsiatis AA. A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics* 1997; 53(1): 330–339.

14. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-dependent data: an overview. *Statistica Sinica* 2004; 14(3): 809–834.

15. Rizopoulos D. *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press . 2012.

16. Liu L, Zheng C, Kang J. Exploring causality mechanism in the joint analysis of longitudinal and survival data. *Statistics in medicine* 2018; 37(26): 3733–3744.

17. Le Coënt Q, Legrand C, Rondeau V. Time-to-event surrogate endpoint validation using mediation analysis and meta-analytic data. *Biostatistics* 2022. kxac044.

18. Zheng C, Liu L. Quantifying direct and indirect effect for longitudinal mediator and survival outcome using joint modeling approach. *Biometrics* 2021; 78: 1233–1243.

19. Joly P, Commenges D, Letenneur L. A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics* 1998; 54(1): 185–194.

20. Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* 2007; 8(4): 708–721.

21. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics* 1963; 11(2): 431–441.

22. Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis* 2012; 56(3): 491–501.

23. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science* 2010; 25(1): 51–71.

24. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis.. *Psychological methods* 2010; 15(4): 309-334.

25. Tchetgen EJT. On causal mediation analysis with a survival outcome. *The international journal of biostatistics* 2011; 7(1).

26. Sylvestre MP, Abrahamowicz M. Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in medicine* 2008; 27(14): 2618–2634.

27. Horwitz EM, Bae K, Hanks GE, et al. Ten-year follow-up of radiation therapy oncology group protocol 92-02: a phase III trial of the duration of elective androgen deprivation in locally advanced prostate cancer. *Journal of Clinical Oncology* 2008; 26(15): 2497–2504.

28. Armstrong AJ, Garrett-Mayer E, Ou Yang YC, et al. Prostate-specific antigen and pain surrogacy analysis in metastatic hormone-refractory prostate cancer. *Journal of Clinical Oncology* 2007; 25(25): 3965–3970.

29. Hanks GE, Pajak TF, Porter A, et al. Phase III trial of long-term adjuvant androgen deprivation after neoadjuvant hormonal cytoreduction and radiotherapy in locally advanced carcinoma of the prostate: the Radiation Therapy Oncology Group Protocol 92–02. *Journal of Clinical Oncology* 2003; 21(21): 3972–3978.

30. Ferrer L, Rondeau V, Dignam J, Pickles T, Jacqmin-Gadda H, Proust-Lima C. Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in medicine* 2016; 35(22): 3933–3948.

31. Sène M, Bellera CA, Proust-Lima C. Shared random-effect models for the joint analysis of longitudinal and time-to-event data: application to the prediction of prostate cancer recurrence. *Journal de la Société Française de Statistique* 2014; 155(1): 134–155.

32. Taylor JM, Park Y, Ankerst DP, et al. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* 2013; 69(1): 206–213.

33. Proust-Lima C, Taylor JM. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* 2009; 10(3): 535–549.

34. Król A, Tournigand C, Michiels S, Rondeau V. Multivariate joint frailty model for the analysis of nonlinear tumor kinetics and dynamic predictions of death. *Statistics in Medicine* 2018; 37(13): 2148–2161.

35. Burzykowski T, Buyse M, Molenberghs G. *The evaluation of surrogate endpoints*. Springer . 2005.

36. 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. *Statistics in medicine* 2019; 38(16): 2928–2942.

37. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* 2006; 5(3): 173–186.

38. Rustand D, Niekerk vJ, Krainski ET, Rue H, Proust-Lima C. Fast and flexible inference approach for joint models of multivariate longitudinal and survival data using Integrated Nested Laplace Approximations. *arXiv preprint arXiv:2203.06256* 2022.



Table 1 Simulations results

		K=20 trials				K=50 trials			
Parameter	Value	Est.*	SE.†	SD.‡	CR.§	Est.	SE.	SD.	CR.
Current-slope									
Surrogacy measures									
PTE(t_1)	0.090	0.427	1.114	1.717	0.620	0.177	0.540	2.716	0.653
PTE(t_2)	0.663	0.487	2.243	8.521	0.808	0.614	1.818	16.221	0.884
NIE(t_1)	0.000	-0.003	0.011	0.003	0.388	-0.030	0.009	0.028	0.381
NIE(t_2)	-0.007	-0.020	0.007	0.018	0.572	-0.001	0.006	0.015	0.683
NDE(t_1)	0.000	-0.006	0.034	0.066	0.572	-0.078	0.030	0.046	0.582
NDE(t_2)	-0.003	0.000	0.023	0.045	0.748	-0.011	0.019	0.027	0.744
Longitudinal submodel									
Intercept (β_0)	1.000	0.936	0.034	0.151	0.628	0.964	0.035	0.131	0.719
Slope (β_1)	0.300	0.288	0.026	0.106	0.496	0.286	0.027	0.131	0.568
$\beta_{Z,M}$	1.000	0.934	0.084	0.217	0.532	0.976	0.074	0.155	0.678
Trt x time (β_2)	0.300	0.370	0.043	0.217	0.280	0.331	0.047	0.211	0.442
β_M	0.100	0.186	0.034	0.166	0.616	0.156	0.036	0.133	0.643
Survival submodel									
Association (η)	0.300	0.320	0.081	0.152	0.684	0.309	0.083	0.138	0.779
$\beta_{Z,T}$	0.300	0.230	0.145	0.214	0.772	0.294	0.121	0.149	0.894
β_T	-0.300	-0.300	0.083	0.082	0.948	-0.292	0.084	0.095	0.915
Random effects									
Trial-level									
σ_{v_M}	0.400	0.316	0.154	0.262	0.748	0.271	0.134	0.178	0.673
σ_{v_T}	0.400	0.401	0.036	0.187	0.320	0.386	0.033	0.133	0.322
$\sigma_{v_{M,T}}$	0.280	0.250	0.031	0.177	0.316	0.224	0.026	0.113	0.276
Individual-level									
σ_{ω_1}	0.562	0.597	0.035	0.050	0.824	0.605	0.036	0.051	0.754
σ_{ω_2}	0.562	0.431	0.041	0.174	0.424	0.431	0.041	0.179	0.342
$\sigma_{\omega_{12}}$	-0.113	-0.098	0.030	0.045	0.736	-0.098	0.030	0.044	0.769
Residuals (σ_ε)	0.500	0.491	0.006	0.011	0.612	0.492	0.006	0.009	0.653

*Mean of the estimated parameters

†Mean of the estimated standard errors

‡Empirical standard deviation

§Coverage rate

Table 2 Patients characteristics (based on a 5 years follow-up)

Variable	LTAD (N=716)	STAD (N=703)
Age	69.82 (43, 88)	69.58 (43, 87)
iPSA*	2.98 (-0.69, 5.21)	3.04 (-1.50, 5.52)
Gleason		
< 7	291 (41%)	290 (41%)
7	251 (35 %)	226 (32%)
> 7	174 (24 %)	187 (27%)
Clinical stage		
T2	331 (46%)	323 (46%)
T3-T4	385 (54%)	380 (54%)
Number of PSA measures	10.03 (1, 18)	8.50 (1, 18)
Mean time between two measures †	0.37 (0, 2.89)	0.38 (0, 3.51)
Median follow-up	5.0 (0.0, 5.0)	3.27 (0.0, 5.0)
DFS events	328 (46%)	432 (61%)
Cause of DFS events		
Death	79 (24%)	66 (15%)
Local progression	60 (18%)	87 (20%)
Distant metastasis	13 (4%)	12 (3%)
Biochemical failure	176 (54%)	267 (62%)

Continuous variables: Mean (Minimum, Maximum)

Categorical variables: Count (Percentage)

* PSA at study entry (in the $\log(\cdot + 0.1)$ scale)

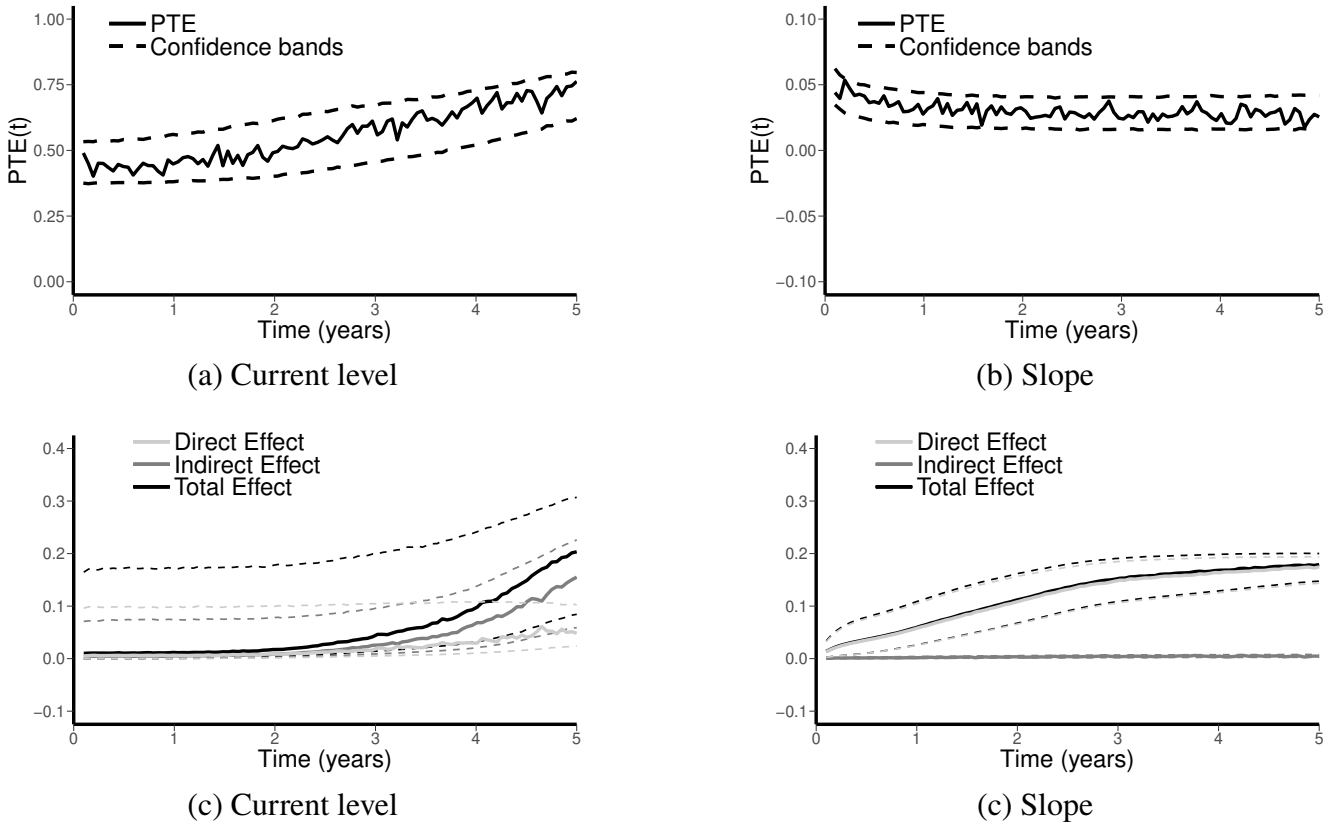
† Time expressed in years

Table 3 Results of the application on prostate cancer.

	Current-level			Current-slope		
	Est.	Std. Err.	p-value.	Est.	Std. Err.	p-value.
Mediation measures						
PTE(1)	0.472	0.038		0.034	0.008	
PTE(3)	0.559	0.056		0.024	0.005	
PTE(5)	0.762	0.045		0.026	0.007	
NIE(1)	0.005	0.018		0.002	0.001	
NIE(3)	0.026	0.023		0.004	0.001	
NIE(5)	0.155	0.041		0.005	0.001	
NDE(1)	0.006	0.029		0.062	0.018	
NDE(3)	0.020	0.027		0.149	0.019	
NDE(5)	0.048	0.019		0.174	0.013	
Longitudinal submodel						
Intercept (β_0)	3.282	0.060	< 0.001	3.313	0.060	< 0.001
$f_1(t)$ (β_1)	4.277	0.021	< 0.001	4.269	0.021	< 0.001
$f_2(t)$ (β_2)	0.590	0.029	< 0.001	0.481	0.030	< 0.001
Treatment ($\beta_{Z,M}$)	-0.257	0.035	< 0.001	-0.297	0.037	< 0.001
Treatment $\times$ $f_2(t)$ (β_3)	-0.192	0.018	< 0.001	-0.141	0.018	< 0.001
Gleason†						
$= 7$	0.055	0.037	0.137	0.064	0.038	0.088
> 7	0.162	0.040	< 0.001	0.168	0.041	< 0.001
Clinical stage‡						
T3-T4	0.126	0.032	< 0.001	0.122	0.033	< 0.001
Survival submodel						
Association (η)	1.336	0.064	< 0.001	0.106	0.004	< 0.001
Treatment ($\beta_{Z,M}$)	-0.360	0.100	< 0.001	-0.495	0.077	< 0.001
Gleason†						
$= 7$	0.117	0.094	0.214	0.087	0.088	0.325
> 7	0.166	0.098	0.090	0.354	0.091	< 0.001
Clinical stage‡						
T3-T4	0.083	0.080	0.307	0.190	0.075	0.010
Random effects						
Individual-level						
σ_{ω_0}	0.300	0.017		0.310	0.018	
σ_{ω_1}	0.079	0.004		0.069	0.004	
$\sigma_{\omega_{02}}$	-0.010	0.006		-0.011	0.006	
Trial-level						
σ_{ν_M}	0.084	0.021		0.147	0.013	
σ_{ν_T}	0.008	0.003		0.005	0.003	
$\sigma_{\nu_{M,T}}$	-0.015	0.002		-0.025	0.002	
Residuals (σ)	0.605	0.004		0.607	0.004	

† Reference group: Gleason < 7

‡ Reference group: Clinical stage T2



(a) Current level (b) Slope

(c) Current level (c) Slope

Figure 3 Results from the application on prostate cancer. Top panel is the estimated PTE for the "current-level" (left) and "current-slope" links. Bottom panel displays the associated estimated natural direct, indirect and total effects. Dashed curves are the 95% pointwise confidence bands. Note that for readability the scales between figures (a) and (b) are different.

Supporting information for "Validation of longitudinal marker as a surrogate using mediation analysis and joint modeling: evolution of PSA as a surrogate of the DFS" by Q. Le Coënt, J. Dignam, C. Legrand and V. Rondeau.

Web Appendix A: Likelihood construction

The joint model is given by,

$$\begin{cases} \tilde{M}_{ij}^{z}(t) = \theta_{ij}^{\mathsf{T}} f(t) + \beta_M^{\mathsf{T}} X_{ij}^1(t) + (\beta_{Z,M} + \nu_{M,i})z + \varepsilon_{ij}(t) \\ \lambda_{T_{ij}^{zm}}(t) = \lambda_0(t) \exp\left(\beta_T^{\mathsf{T}} X_{ij}^2(t) + (\beta_{Z,T} + \nu_{T,i})z + \eta^{\mathsf{T}} h(m)\right) \end{cases} \tag{1}$$

where random effects (ν_{1i}, ν_{2i}) are trial-level effects taking into account the heterogeneity of the treatment effect across trial and are assumed to be jointly gaussian:

$$(\nu_{M,i}, \nu_{T,i})^{\mathsf{T}} \sim \mathcal{N}(\mu_\nu, \Sigma_\nu), \quad \mu_\nu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \Sigma_\nu = \begin{pmatrix} \sigma^2_{\nu_M} & \sigma_{\nu_{TM}} \\ \sigma_{\nu_{TM}} & \sigma^2_{\nu_T} \end{pmatrix}$$

The function $f(t)$ represents the temporal evolution of the biomarkers and may be composed of several components. The parameter θ is the sum of fixed effects and individual random effects assiocated with each component of $f(t)$, $\theta_{ij} = \beta + \omega_{ij}$. For example if $f(t) = \begin{pmatrix} 1 \\ t \end{pmatrix}$ then $\theta_{ij} = \begin{pmatrix} \beta_0 + \omega_{ij0} \\ \beta_1 + \omega_{ij1} \end{pmatrix}$ The vector ω is supposed to be gaussian with mean 0 and covariance matrix Σ_ω. The error terms ε_{ij} are supposed independent and $\varepsilon_{ij}(t) \sim \mathcal{N}(0, \sigma_\varepsilon)$. The baseline hazard function $\lambda_0(t)$ is estimated using cubic M-splines. If K knots are used, there are two boundary knots and $k-2$ inner knots, and $k+2$ parameters $(\Xi_1, \ldots, \Xi_{K+2})$. Let G be the number of trials, $(n_1, \ldots, n_G)$ the number of patients in each trials and n_{ij} the number of measurements for patient j from trial i.

For individual ij the time-to-event and censored times are denoted T_{ij} and C_{ij}. The observed longitudinal biomarker is $\tilde{M}_{ij} = \{\tilde{M}_{ijk} = \tilde{M}_{ij}(t_{ijk}), 1 \leq k \leq n_{ij}\}$. The follow-up time is $T^*_{ij} = \min(T_{ij}, C_{ij})$ with its associated indicator $\delta_{ij} = \mathbb{1}_{T_{ij} \leq C_{ij}}$.

The observed data for patient ij are therefore,

$$\mathcal{O}_{ij} = \left(T^*_{ij}, \delta_{ij}, M_{ij}, Z_{ij}, X_{ij}\right).$$

Let ϕ be the vector of all the parameters of the model:

$$\phi = (\lambda_{0,T}, \eta, \sigma_\varepsilon, \Sigma_\omega, \Sigma_\nu, \beta_T, \beta_M, \beta, \beta_{Z,T}, \beta_{Z,M})$$

The complete likelihood of the data is the product of the contribution of each trial:

$$L(\phi) = \prod_{i=1}^{G} L_i(\phi)$$

where $L_i(\phi)$ is the contribution of the ith trial. Let ν_i be the trial-level random effects, $\nu_i = (\nu_{T,i}, \nu_{M,i})$ Given, ν_i the patients from trial i are independent. Hence,

$$\begin{aligned} L_i(\phi) &= \int_{\mathbb{R}^2} L_i(\phi, \nu_i) d\nu_i \\ &= \int_{\mathbb{R}^2} L_i(\phi \mid \nu_i) f(\nu_i) d\nu_i \\ &= \int_{\mathbb{R}^2} \left(\prod_{j=1}^{n_i} L_{ij}(\phi \mid \nu_i) \right) f(\nu_i) d\nu_i \end{aligned}$$

where $L_{ij}(\phi \mid \nu_i)$ is the individual contribution of subject j from trial j. We have,

$$\begin{aligned} L_{ij}(\phi \mid \nu_i) &= p_\phi(T_{ij}^*, \delta_{ij}, \tilde{M}_{ij} \mid \nu_i) \\ &= \int_{\mathbb{R}^2} p_\phi(T_{ij}^*, \delta_{ij}, \tilde{M}_{ij}, \omega_{ij} \mid \nu_i) \mathrm{d}\omega_{ij} \\ &= \int_{\mathbb{R}^2} p_\phi(T_{ij}^*, \delta_{ij}, \tilde{M}_{ij} \mid \nu_i, \omega_{ij}) f(\omega_{ij}) \mathrm{d}\omega_{ij} \\ &= \int_{\mathbb{R}^2} p_\phi(T_{ij}^*, \delta_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij},) p_\phi\left(\tilde{M}_{ij} \mid \nu_i, \omega_{ij}\right) f(\omega_{ij}) \mathrm{d}\omega_{ij} \end{aligned}$$

Assuming that the censoring time C is independent of T given M, X and Z and non informative we have

$$\begin{aligned} p_\phi\left(T_{ij}^*, \delta_{ij} \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}\right) &= \left(\lambda(T_{ij}^* \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}) S(T_{ij}^* \mid \nu_i, \omega_{ij}, \tilde{M}_{ij})\right)^{\delta_{ij}}) S(T_{ij}^* \mid \nu_i, \omega_{ij}, \tilde{M}_{ij})^{1-\delta_{ij}} \\ &= \lambda(T_{ij}^* \mid \nu_i, \omega_{ij}, \tilde{M}_{ij})^{\delta_{ij}} S(T_{ij}^* \mid \nu_i, \omega_{ij}, \tilde{M}_{ij}) \\ &= \left[\lambda_0(T_{ij}^*) \exp\left((\beta_{Z,T} + \nu_{T,i}) Z_{ij} + \beta_T^\intercal X_{ij}(t) + h\left(M_{ij}\right)\right)\right]^{\delta_{ij}} \times \\ &\quad \exp\left(- \int_0^{T_{ij}^*} \lambda_0(u) \exp\left((\beta_{Z,T} + \nu_{T,i}) Z_{ij} + \beta_T^\intercal X_{ij}(u) + h\left(M_{ij}^*\right)\right) \mathrm{d}u \right) \end{aligned}$$

where the integral

$$\int_0^{Y_{ij}} \lambda_0(u) \exp\left((\beta_{Z,T} + \nu_{T,i}) Z_{ij} + \beta_T^\intercal X_{ij}(u) + h\left(M_{ij}\right)\right) \mathrm{d}u$$

needs to be computed over the trajectory of $h\left(M_{ij}\right)$. The longitudinal part of the individual contribution given the random effects is

$$p_\phi\left(\tilde{M}_{ij} \mid \nu_i, \omega_{ij}\right) = \frac{1}{(2\pi\sigma^2)^{n_{ij}/2}} \exp\left\{ - \frac{\|\tilde{M}_{ij} - \theta_{ij}^\intercal f(t) + \beta_M^\intercal X_{ij}^M + (\beta_{Z,M} + \nu_{M,i}) Z_{ij}\|^2}{2\sigma^2} \right\}$$

Finally the full likelihood can be written as

$$\boxed{L(\phi) = \prod_{i=1}^{G} \int_{\mathbb{R}^2} \left\{ \prod_{j=1}^{K_i} \left[\int_{\mathbb{R}^2} \left(p_\phi\left(Y_{ij}, \delta_{ij} \mid M_{ij}, \nu_i, \omega_{ij}\right) \times \prod_{k=1}^{n_{ij}} p_\phi\left(M_{ijk} \mid \nu_i, \omega_{ij}\right) f(\omega_{ij}) \right) \mathrm{d}\omega_{ij} \right] \right\} f(\nu_i) \mathrm{d}\nu_i}$$

Since $\lambda_{0,T}$ is estimated using splines, a penalized likelihood may instead be maximized:

$$l_{\text{pen}}(\phi) = l(\phi) - \kappa \int_0^\infty \frac{\mathrm{d}^2 \lambda_{0,T}(u)}{\mathrm{d}^2 u} \mathrm{d}u.$$

In practice maximization of $l_{\text{pen}}(\phi)$ is carried out using the Marquardt-Levenberg algorithm (Marquardt, 1963). The inverse of the Hessian of $l_{\text{pen}}(\phi)$ evaluated at the maximum likelihood estimator $\hat{\phi}$ can be used as an estimate of the covariance matrix of $\hat{\phi}$.

Numerical approximations

Since the integral involved in the computation of The inner integral over ω_{ij} is computed using Gauss-Hermite quadrature : let $\omega_{ij} = (\omega_{ij0}, \ldots, \omega_{ijp})$

$$\begin{aligned} I_\omega &= \int_{\mathbb{R}^p} \left(p_\phi \left(Y_{ij}, \delta_{ij} \mid M_{ij}, \nu_i, \omega_{ij} \right) \times \prod_{k=1}^{n_{ij}} p_\phi \left(M_{ijk} \mid \nu_i, \omega_{ij} \right) f(\omega_{ij}) \right) \mathrm{d}\omega_{ij} \\ &\approx \sum_{l1=1}^{q} \cdots \sum_{lp=1}^{q} \left(p_\phi \left(Y_{ij}, \delta_{ij} \mid M_{ij}, \nu_i, (\mathcal{X}_{l1}, \ldots, \mathcal{X}_{lp}) \right) \times \prod_{k=1}^{n_{ij}} p_\phi \left(M_{ijk} \mid \nu_i, (\mathcal{X}_{l1}, \ldots, \mathcal{X}_{lp}) \right) \right. \\ &\left. \times f\left((\mathcal{X}_{l1}, \ldots, \mathcal{X}_{lp}) \right) \mathcal{W}_{l1} \ldots \mathcal{W}_{lp} \mu\left((\mathcal{X}_{l1}, \ldots, \mathcal{X}_{lp}) \right) \right) \end{aligned}$$

where $\mathcal{X}$ and $\mathcal{W}$ are the nodes and weights of the Gauss-Hermite quadrature and $\mu(\cdot)$ is a normalizing function. A more accurate approximation can be achieved using pseudo-adaptive quadrature Rizopoulos, 2012.

The outer integral over the trial-level random effects distribution is carried out through Monte-Carlo integration. Let A be the (upper) cholesky decomposition of Σ_ν, that is $\Sigma_\nu = AA^\mathsf{T}$. If $X \sim \mathcal{N}(0, I_2)$, then $AX \sim \mathcal{N}(0, \Sigma_\nu)$. Thus,

$$\begin{aligned} I_\nu &= \int_{\mathbb{R}^2} \left\{ \prod_{j=1}^{n_i} \left[\int_{\mathbb{R}^2} \left(p_\phi \left(Y_{ij}, \delta_{ij} \mid M_{ij}, \nu_i, \omega_{ij} \right) \times \prod_{k=1}^{n_{ij}} p_\phi \left(M_{ijk} \mid \nu_i, \omega_{ij} \right) f(\omega_{ij}) \right) \mathrm{d}\omega_{ij} \right] \right\} f(\nu_i) \mathrm{d}\nu_i \\ &\approx \frac{1}{N} \sum_{l=1}^{N} \left\{ \prod_{j=1}^{K_i} \left[\int_{\mathbb{R}^2} \left(p_\phi \left(Y_{ij}, \delta_{ij} \mid M_{ij}, \nu_l, \omega_{ij} \right) \times \prod_{k=1}^{n_{ij}} p_\phi \left(M_{ijk} \mid \nu_l, \omega_{ij} \right) f(\omega_{ij}) \right) \mathrm{d}\omega_{ij} \right] \right\} \end{aligned}$$

where $\nu_1, \ldots, \nu_n$ were generated as :

1. Generate $X_1, \ldots, X_n$ from $\mathcal{N}(0, I_2)$
2. Take ν_i as $\nu_i = AX_i$

Numerical underflow

It may happends that if, for a trial i, the number of subject K_i is large or if there are many observations per individuals (n_{ij}) that the integrand $\prod_{j=1}^{K_i} \left[\int_{\mathbb{R}^2} \left(p_\phi \left(T_{ij}^*, \delta_{ij} \mid \tilde{M}_{ij}, \nu_i, \omega_{ij} \right) \times \prod_{k=1}^{n_{ij}} p_\phi \left(\tilde{M}_{ijk} \mid \nu_i, \omega_{ij} \right) f(\omega_{ij}) \right) \mathrm{d}\omega_{ij} \right]$ is so small it results in an computation underflow: it can no longer be represented by floating point number and is rounded to 0. This is a comon issue and here we cannot work it out using log densities since there are within an integral and $\log(\int f) \neq \int \log(f)$. One possibilty is to weight each individidual contribution by a (large)

constant C so that the product no longer results in an underflow. This constant must not directly depends on the values of the individual densities since it cannot be extrated out the integral (it will depends on the ν_i and ω_{ij} that were're integrating over). Moreover, if this constant is fixed before the maximization process it may happen that this constant works well for the first values of the parameter ϕ, but as the iterations goes and the parameters updates, the scale of the densities may change and C in that case may not be appropriate.

Web Appendix B: Detailed simulation scheme

The core of the simulating process relies on the permutation algorithm developed (Sylvestre and Abrahamowicz, 2008), which allows to easily simulate time-to-event data according to proportional hazard with a time-dependent covariate (in our the longitudinal biomarker). Let be K be the number of trials and n the total number of subjects in the dataset. The simulated model is:

$$\begin{cases} \tilde{M}_{ij}^{z}(t) = \theta_{ij}^{'} f(t) + +(\beta_{Z,M} + \nu_{M,i})z + \beta_2 z \times t + \beta_M' X_{ij}^{M} + \varepsilon_{ij}(t) \\ \lambda_{ij}^{zM^z}(t) = \lambda_0(t) \exp\left((\beta_{Z,T} + \nu_{T,i})\, z + \beta_T' X_{ij}^{T} + \eta' h\left(M^z\right)\right) \end{cases} \tag{2}$$

where $\varepsilon_{ij}(t) \sim \mathcal{N}(0, \sigma^2)$ and

$$f(t) = \begin{pmatrix} 1 \\ t \end{pmatrix}, \quad \theta_{ij} = \begin{pmatrix} \beta_0 + \omega_{ij0} \\ \beta_1 + \omega_{ij1} \end{pmatrix}.$$

Parameters β_0, β_1 and β_2 are respectively the fixed intercept, slope and the effect of the treatment on the slope. The individual-level random effects ω_{ij} were generated from a bivariate gaussian distribution with mean 0 and covariance matrix

$$\begin{pmatrix} \omega_{ij0} \\ \omega_{ij1} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_\omega), \quad \Sigma_\omega = \begin{pmatrix} \sigma_{\omega_1}^2 & \sigma_{\omega_{12}} \\ \sigma_{\omega_{12}} & \sigma_{\omega_2}^2 \end{pmatrix}$$

The trial level random effects were simulated from a centered bivariate gaussian distribution with mean 0 and covariance matrix

$$\begin{pmatrix} \nu_{M,i} \\ \nu_{T,i} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_\nu), \quad \Sigma_\nu = \begin{pmatrix} \sigma_{\nu_M}^2 & \sigma_{\nu_{M,T}} \\ \sigma_{\nu_{M,T}} & \sigma_{\nu_T}^2 \end{pmatrix}$$

The different steps for simulating the datasets according to the parameters of the model are:

1. Generate $\begin{pmatrix} \nu_{M,i} \\ \nu_{T_i} \end{pmatrix}_{1 \le i \le K}$ where $\begin{pmatrix} v_{M,i} \\ v_{T,i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{v_S}^2 & \sigma_{V_{ST}} \\ \sigma_{V_{ST}} & \sigma_{V_T}^2 \end{pmatrix}\right)$
2. Generate $u = (u_i)_{1 \le i \le K}$ where $u_i \sim \mathcal{N}(0, \rho^2)$
3. For $j = 1, \ldots, n$:
   (a) Draw i uniformly from $\{1, \ldots, K\}$
   (b) Draw $\begin{pmatrix} \omega_{ij0} \\ \omega_{ij1} \end{pmatrix} \sim \mathcal{N}(0, \Sigma_\omega)$
   (c) Draw $Z_{ij} \sim \mathcal{B}(p = 0.5)$ and $X_{ij} \sim \mathcal{B}(p = 0.5)$

(d) For t between 0 and 4 with a step 0.001:

- Compute $M_{ij}(t_{ijk}) = \theta_{ij}^{'} f(t) + +(\beta_{Z,M} + \nu_{M,i})Z_{ij} + \beta_2 Z_{ij} \times t + \beta_M X_{ij}$
- Generate $\varepsilon_{ij}(t) \sim \mathcal{N}(0, \sigma^2)$

4. Generate $T_{ij} \sim \exp(\lambda)$ for $j = 1, \ldots, n.$, $\lambda = 0.012$
5. Generate $C_k \sim \mathcal{U}[1, 4001]$
6. Use the permutation algoritm with $[T_{ij}, C_{ij}, h(M_{ij}), Z_{ij}, X_{ij}]$
7. Retrieve the censored time T_{ij}^* ans indicator δ_{ij}
8. Retrieve the observed longitiduinal marker $\tilde{M}_{ij} = M_{ij} + \varepsilon_{ij}$

Web Appendix C: Additional figures

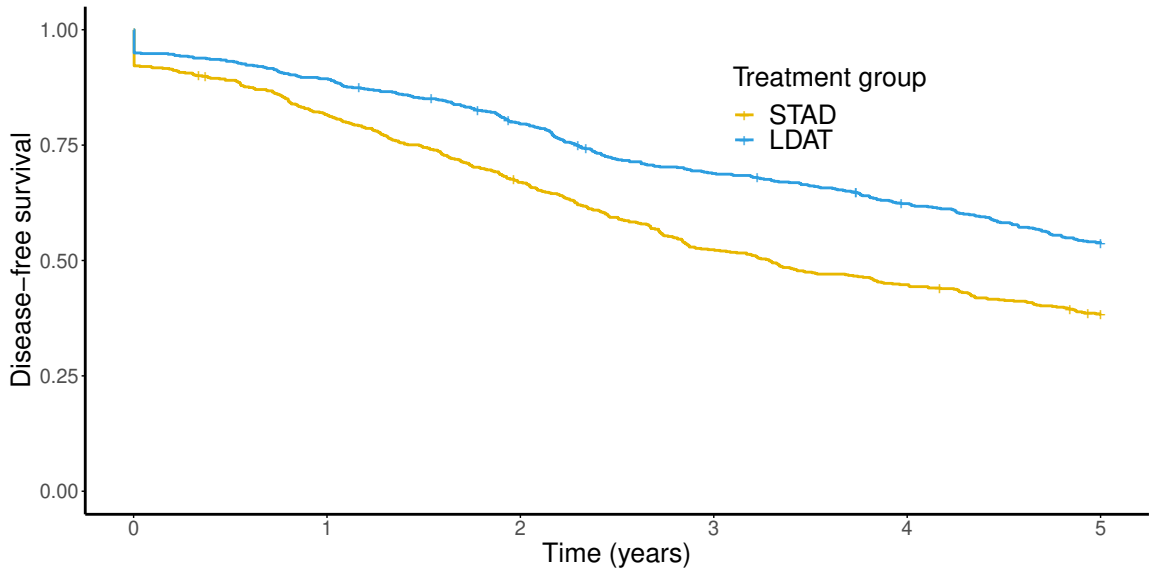


Figure 1: Disease-free survival by treatment arm

References

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics* **11**, 431–441.

Rizopoulos, D. (2012). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis* **56**, 491–501.

Sylvestre, M.-P. and M. Abrahamowicz (2008). Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in medicine* **27**, 2618–2634.

IV.3. | Additional remarks

As of now the current implementation of Model (II.6) is flawed especially regarding the estimation of the standard-errors of the model's parameters. This can be due to several factors and we have yet to identify what causes this issues. The covariance matrix of the standard-errors of the parameters is taken as the inverse of the Hessian matrix of the penalized log-likelihood evaluated at the estimated maximum likelihood. One possible explanations of the issue regarding the standard-errors estimations can come from the evaluation of the derivatives of the penalized log-likelihood. This penalized likelihood is sensitive to several factors. First the penalization term κ is Equation (II.12) can have an important influence on the result. The choice of the appropriate value for κ is therefore crucial. One possibility is to take a value chosen by cross-validation on a reduced survival model for the final endpoint T. However, while this may be an interesting approach to undertake it might have some limitations as the κ obtained on a reduced model can quite differ with the optimal κ in the full joint model. Secondly, the log-likelihood is also very sensitive to the integral approximation used in its computation. The evaluation of log-likelihood of Model (II.6) requires the approximation of three ingrals. A first one over the distribution of the individual-level random effects (which is an integral over $\mathbb{R}^p$ where p is the number of individual-level random effects), a second over the distribution of the trial-level random effects (which is over $\mathbb{R}^2$) and a third one over time (the integral involved in the exponential in Equation (II.10)). These numerous integrals can yields unsatisfactory results with a likelihood not properly approximated. Therefore the choice of approximation method can greatly influence the results. The current implementation of the model estimation uses pseudo-adaptive Gauss-Hermite quadrature rule for the first integral, a Monte-Carlo sampling for the second integral and a Gauss-Kronrod quadrature for the third integral. Investigating other integration methods (for example a full Monte Carlo sampling for the two first integrals, or a full pseudo-adaptive Gauss-Hermite) can help explains to poorly estimated standard-errors.

V

Tutorial for surrogate endpoint validation using joint modeling and mediation analysis

V.1 Introduction 135

V.2 Main article 136

V.1. | Introduction

A large part of this thesis concerned the development and the implementation of the proposed approaches in a user-friendly statistical package. These approaches were implementated in the R package `frailtypack`. They were integrated in two functions of the package, one for the case where both the surrogate the final endpoint are time-to-event and another one for a longitudinal continuous surrogate and a time-to-event final endpoint. The estimation of the developed joint models as well as the resulting surrogate measures (and in particular the proportion of the treatment effect) can require quite a long computation time. In order to reduce these computation times, these methods have been developed in `Fortran` and can be called directly from the two functions in `frailtypack`. Moreover, these methods support parallel computing, and particularly multithreading by being developed with `OpenMP`. In order to facilitate the use of the proposed methods, we have written an article describing them, and illustrated their use through three applications on real data sets from meta-analyses or multicenter trials. The first illustration concerns data from four randomized trials in advanced ovarian cancer and focuses on the study of time to cancer progression as a surrogate for time to death through the meta-analytic approach, i.e. by estimating the association at the individual and trial levels. A second illustration focuses on the application of the first part of this thesis to gastric cancer and investigate the time to cancer recurrence as a surrogate for time to death

through measures related to mediation analysis, i.e. the proportion of treatment effect. A third illustration focuses on a longitudinal endpoint in the context of colorectal cancer. The longitudinal surrogate endpoint is the evolution of the tumor size over time and the final endpoint is time to death. As for the second illustration, the approach presented is based on the estimation of the proportion of treatment effect on the final endpoint through the surrogate endpoint. This article is currently being finalized for submission to a journal dedicated to the software implementation of statistical methods.

V.2. | MAIN ARTICLE

Tutorial for surrogate endpoint validation using joint modeling and mediation analysis

Quentin Le Coent[1], Catherine Legrand[2], and Virginie Rondeau[1]

[1]Department of Biostatistics, Bordeaux Population Health Research Center, INSERM U1219, Bordeaux, France

[2]ISBA/LIDAM, UCLouvain, Louvain-la-Neuve, Belgium

Summary

The use of valid surrogate endpoints is an important stake in clinical research to help reduce both the duration and cost of a clinical trial and speed up the evaluation of interesting treatments. Several methods have been proposed in the statistical literature to validate putative surrogate endpoints. Two main approaches have been proposed: the meta-analytic approach and the mediation analysis approach. The former uses data from meta-analyses to derive associations measures between the surrogate and the final endpoint at the individual and trial levels. The latter rather uses the proportion of the treatment effect on the final endpoint through the surrogate as a measure of surrogacy in a causal inference framework. Both approaches have remained separated as the meta-analytic approach does not estimate the treatment effect on the final endpoint through the surrogate while the mediation analysis approach have been limited to single-trial setting. However, these two approaches are complementary. In this work we propose an approach that combines the meta-analytic and mediation analysis approaches using joint modeling for surrogate validation. We focus on the cases where the final endpoint is a time-to-event endpoint (such as time-to-death) and the surrogate is either a time-to-event or a longitudinal biomarker. Two new joint models were proposed depending on the nature of the surrogate. These model are implemented in the R package `frailtypack`. We illustrate the developed approaches in three applications on real datasets in oncology.

Keywords: Surrogate endpoint, Meta-analyses, Mediation analysis, Joint models.

1 Introduction

Given the continued development of new and more effective therapies in oncology, the use of standard (final) endpoints in clinical trials may affect the feasibility of these trials. For example, a significant increase in overall survival requires long-term follow-up of patients to ensure adequate statistical power and therefore increases the duration and cost of the trial. The use of surrogate endpoints is therefore a possible way to enable such otherwise unfeasible studies. Surrogate endpoints are intermediate endpoints measured earlier and/or more frequently than the final (true) endpoints that help reduce the duration and cost of a trial compared to final endpoints (Buyse et al. (2000)). The benefit of a surrogate endpoint is that is can be used as a primary endpoint in place of the final endpoint in a new clinical trial. However, the final endpoint being the "gold-standard", conclusions of treatment efficacy in a clinical trial based solely on its observed efficacy on the surrogate should be extendable to the final endpoint. Therefore, prior to any use in a new trial, a surrogate endpoint must have been carefully validated. This validation requires that both the final endpoint and the surrogate have been studied in previous studies. Several methods have been proposed in the last decades following the seminal paper of Prentice (Prentice, 1989). Since, two main approaches have been developed, one based on quantifying "associations" between the surrogate and the final endpoint and others based on identification of the proportion of treatment effects on the final endpoint through the surrogate (Joffe and Greene, 2009). The main developments of the "associations" approach is the meta-analytic approach (Daniels and Hughes, 1997; Buyse et al., 2000) while the proportion of treatment effect has been developed in the mediation analysis setting (Vandenberghe et al., 2018; Weir et al., 2022; Parast, Cai, et al., 2017).

The meta-analytic approach uses data from meta-analyses of randomized clinical trials to derive and quantify the association between the surrogate and the final endpoint at two levels. At the individual level the association between the two endpoints is quantified while at the trial-level the association between the treatment effects on both endpoints across the trials is used. For the individual-level association, this approach relies on joint models for the two endpoints using individual (patient) level random effects. These random effects can be shared or correlated between the two endpoints and can be used to estimate an individual-level association between the surrogate and the final endpoint. For the trial-level association, methods based on one or two stages approaches have been developed. In the two-stage approaches, the joint model is estimated with fixed trial-specific treatment effects on the two endpoints. Then at the second stage the estimated treatment effects on the two endpoints for each trial can be regressed in a linear mixed model. The resulting coefficient of determination can then be used as trial-level measure of association between the treatment effects. However, these two stage approaches can be replaced by one stage approaches in which the fixed trial-specific treatment effects in the joint model are replaced by the sum of a fixed effect (not trial-specific) and a trial-level random effect for each endpoint. Then the elements of the covariance matrix of these trial-level random effects are used

to derive the coefficient of determination between the treatment effects. Then, a surrogate endpoint will be claimed validated if the associations at both individual and trial levels are sufficiently large. The meta-analytic approach has been broadly studied and adapted to settings with different types of surrogate and final endpoint: continuous, binary, censored time-to-event (Burzykowski, Buyse, and Molenberghs, 2005).

Another approach, less developed, focuses on identifying the pathways through which the treatment effect on the final endpoint goes. A natural surrogacy measure in that case is the proportion of treatment effect, defined as the ratio of the indirect treatment effect on the final endpoint through the surrogate over the total treatment effect on the final endpoint. Early developments of this approach were based on estimating regression coefficients of the treatment effect in a model for the final endpoint adjusted only on the treatment arm and a model adjusting on both the treatment and the surrogate (Prentice, 1989; Freedman et al., 1992). Then the ratio of the estimated regression coefficient associated with the treatment in the second model over the coefficient estimated in the first model was taken as a measure of the proportion of treatment effect. However, this approach lacks causal interpretability since adjusting on a posttreatment covariate affected by the treatment (the surrogate) may yield biased estimation of the treatment effect by neglecting potential confounding between the surrogate and the final endpoint (Rosenbaum, 1984; Robins, 1986; Frangakis and Rubin, 2002). Moreover, using estimated regression coefficients to define a proportion of treatment effect complicates the interpretation of this proportion on different scales. For example, a high proportion using estimated coefficients in a Cox model might not translates in the same proportion on the survival scale (i.e. by using the difference of survival rates as a measure of the treatment effect). Additionally, estimates of hazard models may also suffer from a lack of causal interpretability in contrary to survival measures (Hernán, 2010). Therefore, methods based on causal mediation analysis have been developed. These methods often rely on the causal framework of potential outcomes (Holland, 1986). The benefit of using causal mediation analysis is two fold: first it does not suffer from a lack of causal interpretability; second, it gives model-free definitions of "treatment effects", which can be defined in a broad variety of situations. For example, if we are interested in the case where the final endpoint is a time-to-event, then it is possible to give a definition of the proportion of treatment effects on the survival scale. However, causal mediation analysis comes at a price: one often needs to rely on strong identifiability assumptions (Imai et al., 2010) and the estimations of the quantities of interest are generally more complex than simply fitting a regression model. Less developments of the mediation analysis for surrogate endpoints validation have been made compared to the meta-analytic approach. One limitation of the mediation analysis is that it is often limited to single-trial data, which may complicate the generalization of its results.

We propose a novel development that combines the meta-analytic and mediation analysis approaches. We focus on the specific case where the final endpoint is a time-to-event. We develop two joint models for the case where the surrogate is also a time-to-event or a continuous longitudinal biomarker. The

first model is an extension of the joint model proposed by Sofeu et al. (2019) that allows for a potential mediated treatment effect on the final endpoint through the surrogate. The second model is a joint model for a time-to-event and a longitudinal surrogate. In both models the use of individual and trial-level random effects allows for i) using meta-analytic data and ii) deriving individual and trial-level associations. From these models we also gave a causal definition of the indirect treatment effect on the final endpoint through the surrogate and the direct treatment effect using mediation analysis. The use of random effects allows us to make less stringent identifiability assumptions than often found in the mediation analysis literature.

This paper is organized as follows. In the first Section we provide a review on the existing R packages available on the CRAN for surrogate endpoints validation with their principal options (type of endpoints, methodology etc.). In a second section we introduce the statistical methodology for surrogate endpoint validation. Two joint models are presented, the first one for the case where the surrogate endpoint is a time-to-event and the second where the surrogate is a longitudinal continuous biomarker. In both models the final endpoint will be a time-to-event. We also present the measures of surrogacy resulting from these models. In the third Section we present two functions implemented in `frailtypack`, `jointSurroPenal` and `longiPenal`, dedicated to each proposed model. We present the main arguments of these functions and their role. In the fourth section we provide three illustrations of the proposed models on real-data examples in oncology. Finally in the fifth section we conclude and provide some further developments.

2 Existing R packages for surrogate validation

In this Section we review some R packages which provide useful measures for surrogate endpoint validation.

The package `Rsurrogate` provides several functions to compute the proportion of treatment effect on a final endpoint through a surrogate. It can be used with a censored time-to-event final endpoint and a surrogate measured at a specific time, with the surrogate being treated as missing if the final endpoint occurs before this time point. It can also be used to compute the proportion of treatment effect on a (time-to-event) final endpoint through several surrogates. It further provides a function to compute this proportion in the case where the final endpoint is a continuous endpoint, where a function to correct for a potential measurement error in the surrogate is also available. In all cases the surrogate (or vector of surrogates) is assumed to be continuous (Parast, 2021a).

The package `SurrogateOutcome` provides a function to estimate the proportion of treatment effect on a time-to-event endpoint through a time-to-event surrogate observed up to a landmark time (Parast, 2021b). These effects are defined as differences in restricted mean survival times. This package also provides a function in which the surrogate is replaced by the final endpoint itself observed up to this

landmark time. The package `longsurr` provide a function to compute the proportion of treatment effect on a final continuous endpoint through a longitudinal surrogate, where it is assumed that each patient has the same number of repeated observations of the surrogate (and that each measurement is taken at the same time) (Parast and Agniel, 2022). The package `surrogate` (Van der Elst et al., 2022) is the most developed package dedicated to surrogate endpoint validation available on the CRAN. It provides numerous functions to assess surrogacy depending on the nature of the surrogate and the final endpoint based on the meta-analytic, information-theoretic, and causal-inference frameworks. For example, this package can be used to assess the trial-level surrogacy for two time-to-event endpoints using a two-stage approach or based on information theory and a two-stage approach. It can also be used to estimate the trial-level surrogacy by providing directly the coefficients associated with the estimated treatment effects on the final and surrogate endpoints. It also provides a function to evaluate surrogacy following Prentice's criteria in a single-trial setting with two continuous endpoints. Although we succinctly presented packages developed in R, implementations for surrogate validation have also been proposed in other softwares such as SAS (Alonso et al., 2016). None of those packages combined the one-step meta-analytic approach with the mediation analysis. We propose an R package to validate a longitudinal biomarker or failure time as a surrogate of a time-to-event using mediation analysis and the meta-analytic approach.

3 Methods: surrogate validation using joint modeling and mediation analysis

In this Section we introduce the two main models: the first one for a time-to-event surrogate and the second one for a longitudinal surrogate. The surrogacy evaluation criteria are presented for both the meta-analytic and the mediation approaches. Estimation method of model parameters is also presented.

3.1 Models and estimation

In a setting of meta-analytic data with K trials, let i denotes the ith trial and ij the jth subject from trial i, $j = 1, \ldots, n_i$. The treatment and covariates are denoted by Z_{ij} and X_{ij}.

3.1.1 Time-to-event surrogate

Let T_{ij} and S_{ij} be the final endpoint and the surrogate endpoint respectively. Let C_{ij} denotes the censoring time. We assume a semi-competing risks setting in that S_{ij} can be censored by T_{ij} but not the opposite. The observed outcomes are therefore $T^*_{ij} = \min(T_{ij}, S_{ij})$, $\delta_{ij} = \mathbb{1}_{T_{ij} \leq C_{ij}}$, $S^*_{ij} = \min(S_{ij}, T^*_{ij})$

and $d_{ij} = \mathbb{1}_{S_{ij} \leq T^*_{ij}}$. The joint model is given by (Le Coënt et al., 2022):

$$\begin{cases} \lambda_S\left(t \mid \omega_{ij}, u_i, \nu_{S,i}, Z_{ij}, X_{ij}\right) = \lambda_{0,S}(t) \exp\Big(\omega_{ij} + u_i + Z_{ij}(\nu_{S,i} + \beta_{Z,S}) + \beta_S' X_{ij}\Big) \\ \lambda_T\left(t \mid \omega_{ij}, u_i, \nu_{T,i}, Z_{ij}, S_{ij}, X_{ij}\right) = \lambda_{0,T}(t) \exp\Big(\zeta\omega_{ij} + \alpha u_i + Z_{ij}(\nu_{T,i} + \beta_{Z,T}) \\ \qquad\qquad + \beta_T' X_{ij} + \gamma(S_{ij}) I\left(S_{ij} \leq t\right)\Big). \end{cases} \tag{1}$$

In this model, ω_{ij} and u_i are individual and trial level random effects respectively used to take into account the heterogeneity between patients at the individual and trial levels. The trial-level random effects $(\nu_{S,i}, \nu_{T,i})$ take into account a potential heterogeneity between treatment effects across trials. All of these random effects are assumed to be Gaussian:

$$\begin{aligned} &\omega_{ij} \sim \mathcal{N}(0, \theta^2) \\ &u_i \sim \mathcal{N}(0, \rho^2) \\ &(\nu_{S,i}, \nu_{T,i})' \sim \mathcal{N}(0, \Sigma_\nu). \end{aligned}$$

The parameters (ζ, α) are power parameters allowing the random effects ω_{ij} and u_i to affect the hazard functions of S and T differently. Parameters $\beta_{Z,S}$ and $\beta_{Z,T}$ are the fixed treatment effects on the surrogate and the final endpoint respectively. The functions $\lambda_{0,T}$ and $\lambda_{0,S}$ are the baseline hazard functions for the two endpoints.

In model (1), the term $\gamma(S_{ij}) I\left(S_{ij} \leq t\right)$ corresponds to a direct link of the surrogate on the final endpoint. This link can be interpreted as a modification of the hazard function of T_{ij} at the time of occurrence of S_{ij} whose magnitude is given by the function $\gamma(\cdot)$. From a mediation perspective this term is important as without it the two endpoints would be independent given the random effects, treatment and covariates, hence no direct effect of the surrogate on the final endpoint would be possible (which corresponds to removing the arrow between S and T in Figure 1(b)).

Outside this mediation perspective, and based only on the meta-analytic approach, a direct link between S and T is not required, and a joint model without the term $\gamma(S_{ij}) I\left(S_{ij} \leq t\right)$ can be used (Sofeu et al., 2019).

3.1.2 Longitudinal surrogate

Let M_{ij} denotes the surrogate which is a continuous longitudinal biomarker (repeated measures over time) and T_{ij} is a time-to-event defined as previously. For subject ij we observe $\tilde{M}_{ij} = \{\tilde{M}_{ij}(t_{ijk}), 1 \leq k \leq n_{ij}\}$ the longitudinal biomarker measured with an error. where $n_i ij$ is the number of repeated measurements of subject ij.

The proposed joint model is given by,

$$\begin{cases} \tilde{M}_{ij}(t) = \theta'_{ij} f(t) + \beta'_M X^M_{ij}(t) + (\beta_{Z,M} + \nu_{M,i}) Z_{ij} + \varepsilon_{ij}(t) \\ \lambda_{ij}(t) = \lambda_0(t) \exp\left((\beta_{Z,T} + \nu_{T,i}) Z_{ij} + \beta'_T X^T_{ij}(t) + \eta' h\left(M_{ij}\right) \right) \end{cases} \tag{2}$$

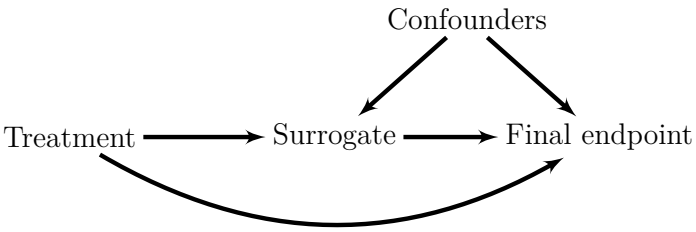
where $M_{ij}(t) = \theta'_{ij} f(t) + \beta'_M X^M_{ij}(t) + (\beta_{Z,M} + \nu_{M,i}) z$ is the true, error-free, marker and $\varepsilon_{ij}(t) \sim \mathcal{N}(0, \sigma^2_\varepsilon)$ (the independent measurement error).

As in Model (1), the random effects $(\nu_{S,i}, \nu_{T,i})$ are trial-level effects taking into account the heterogeneity of the treatment effect across trial and are assumed to be jointly Gaussian: $(\nu_{M,i}, \nu_{T,i})' \sim \mathcal{N}\left(0, \Sigma_\nu\right)$. The function $f(t)$ is a vector that represents the temporal evolution of the biomarker and may be composed of several components. The vector θ is the sum of fixed effects and individual random-effects associated with each component of $f(t)$, $\theta_{ij} = \beta + \omega_{ij}$. For example if $f(t) = (1, t)'$ then $\theta_{ij} = (\beta_0 + \omega_{ij0}, \beta_1 + \omega_{ij1})'$. The vector ω_{ij} is assumed to be Gaussian with mean 0 and unstructured covariance matrix D, $\omega \sim \mathcal{N}(0, D)$.

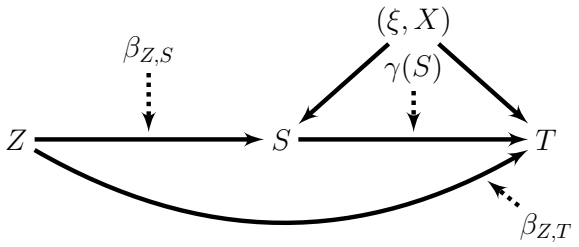
In Model (2), the two submodels are linked together through the "link" function $h\left(M_{ij}\right)$. This function represents the association structure between the longitudinal surrogate and the final endpoint. The strength of this association is taken into account through the "association parameter" η. This parameter quantifies the magnitude of the dependence between both outcomes. The most common link functions found in the literature are the following (Rizopoulos, 2012b):

- The "current level" link: $h\left(M_{ij}\right) = M_{ij}(t)$. This is probably the most natural link between the biomarker and the time-to-event that assumes that the risk of occurrence of T at any time t depends only on the current value of the biomarker at that time.
- The "current slope" link: $h\left(M_{ij}\right) = \frac{\mathrm{d}M_{ij}(t)}{\mathrm{d}t}$. Rather than considering an association between the hazard function of T with the current value of M_{ij} at time t, we consider an association with the increase (or decrease) of M_{ij} at time t.
- $h\left(M_{ij}\right) = \omega_{ij}$, the "shared random-effects" link function. As it is time-independent, it assumes that dependence structure between the two outcomes is constant over time regardless of the biomarker evolution.

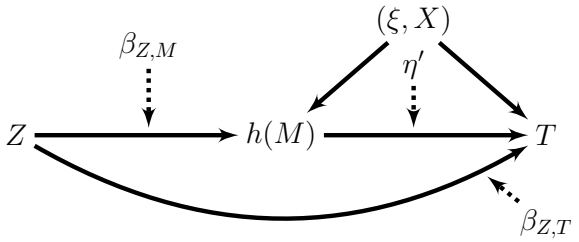
From a mediation viewpoint, in order to enable a potential mediated treatment effect on the final endpoint through the surrogate, the quantity in the link function must depends on the treatment value (which corresponds to the arrow from Z to $h(M)$ in Figure 1(c)). For example in the "current-slope" case the slope must be dependent on the treatment. This can be achieved by adding an interaction between treatment and time in the longitudinal model. Moreover, the shared random-effects link is not relevant in a mediation analysis perspective as the random effects are assumed independent of the covariates.



(a) General case



(b) Model (1)



(c) Model (2)

Figure 1: Directed acyclic graph associated with the joints models: (a) in a general case, (b) for a time-to-event surrogate as represented by Model (1) and (c) for a longitudinal surrogate (Model (2)). Here ξ denotes the vector of all the random effects in models 1 and 2.

3.1.3 Likelihoods and estimations

Both Model (1) and Model (2) can be estimated through maximum likelihood. The baseline hazard functions can be estimated parametrically using M-splines which provides flexibility in their estimations. A penalized likelihood can be maximized instead of the full likelihood to ensure smooth estimate of these functions (Joly et al., 1998; Rondeau, Mathoulin-Pelissier, et al., 2007). Let ϕ_1 and ϕ_2 be the vectors of all the parameters in Model (1) and Model (2) respectively. The penalized likelihood for Model (1) is given by:

$$L_{\text{pen}}(\phi_1) = L(\phi_1) - \kappa_T \int_0^\infty \left(\lambda''_{0,T}(t)\right)^2 \mathrm{d}t - \kappa_S \int_0^\infty \left(\lambda''_{0,S}(t)\right)^2 \mathrm{d}t \tag{3}$$

and for the Model (2):

$$L_{\text{pen}}(\phi_2) = L(\phi_2) - \kappa_T \int_0^\infty \left(\lambda''_{0,T}(t)\right)^2 \mathrm{d}t \tag{4}$$

In these equations, the terms κ_S and κ_T are penalization terms on the second derivatives of the baseline hazard functions that ensure smooth estimations of these functions.

In practice the maximization of the penalized likelihood is carried out using the Levenberg-Marquardt algorithm (Marquardt, 1963).

3.2 Surrogacy measures

In this section we provide definitions and formulas for the surrogacy measures that can be used with each model. In the meta-analytic approach, several measures have been proposed in the literature: individual and trial-level associations and the surrogate threshold effect (STE). From a mediation perspective the main measure of surrogacy is the proportion of treatment effect (PTE), which quantifies the amount of treatment effect on the final endpoint going through its effect on the surrogate. In the following we present each of these measures. We denote by ξ the vector of all random effects in Model (1) or Model (2).

3.2.1 Individual-level association

Individual-level surrogacy is often represented by a measure of association between the surrogate and the final endpoint, with the Kendall's τ between both endpoints being commonly used. We first introduce the definition of the Kendall's τ and then give a derivation based on the two proposed joint models. Let X and Y be two continuous random variables. For two random subjects (X_1, Y_1) and (X_2, Y_2), Kendall's τ is defined as the probability that the two couples (X_1, Y_1) and (X_2, Y_2) are concordant minus the probability that they are discordant. The couples are concordant if $X_1 > X_2$ and

$T_1 > T_2$ or if $X_1 < X_2$ and $T_1 < T_2$ which is equivalent to $(X_1 - X_2)(Y_1 - Y_2) > 0$. They are discordant otherwise. Therefore the Kendall's τ for (X, Y) can be written as (Wang and Wells, 2000)

$$\tau = 4\mathbb{P}\left(X_1 > X_2, Y_1 > Y_2\right) - 1. \tag{5}$$

In the case where X and Y are independent, $\mathbb{P}\left((X_1 > X_2, Y_1 > Y_2\right) = \mathbb{P}\left(X_1 > X_2\right)\mathbb{P}\left(Y_1 > Y_2\right)$ and by symmetry $\mathbb{P}\left(X_1 > X_2\right) = \mathbb{P}\left(Y_1 > Y_2\right) = \frac{1}{2}$ which gives $\tau = 0$. On the other hand, for perfect correlation $X = Y$, $\mathbb{P}\left(X_1 > X_2, Y_1 > Y_2\right) = \mathbb{P}\left(X_1 > X_2\right) = \frac{1}{2}$ and then $\tau = 1$ while $X = -Y \Rightarrow \mathbb{P}\left(X_1 > X_2, Y_1 > Y_2\right) = 0 \Rightarrow \tau = -1$. Hence, τ takes values in $[-1, 1]$ with values close to 0 meaning small association between X and Y.

(a) Kendall's τ for a time-to-event surrogate

It can be shown that for Model (1) the Kendall's τ is equal to (for the variables S and T),

$$\begin{aligned}\tau = 4\int_{\mathbb{R}^4}\Bigg[\int_0^\infty\int_0^\infty\left(\int_0^\infty \mathbb{P}(T_1 > t \mid S_1 = s+\alpha, \xi_1) f_{T_2}(t \mid S_2 = s, \xi_2)\mathrm{d}t\right)\\ \times f_{S_1}(s+\alpha \mid \xi_1) f_{S_2}(s \mid \xi_2)\mathrm{d}s\mathrm{d}\alpha\Bigg]\mathbb{P}(S_1 > S_2 \mid \xi_1, \xi_2)\mathrm{d}f(\xi_1, \xi_2)\xi_1\mathrm{d}\xi_2 - 1,\end{aligned} \tag{6}$$

where:

$$\begin{aligned}
&\mathbb{P}(T_1 > t \mid S_1 = s+\alpha, \xi_1) = \exp\left(-\int_0^t \lambda_{0,T}(u) e^{\xi_1 + \gamma(s+\alpha) I(s+\alpha \le u)}\mathrm{d}u\right)\\
&f_{T_2}(t \mid S_2 = s, \xi_2) = \lambda_{0,T}(t) e^{\xi_2 + \gamma(t) I(s \le t)} \exp\left(-\int_0^t \lambda_{0,T}(u) e^{\xi_1 + \gamma(s+\alpha) I(s+\alpha \le u)}\mathrm{d}u\right),\\
&f_{S_i}(x \mid \xi_i) = \lambda_{0,S}(x) e^{\xi_i} \exp\left(-\int_0^t \lambda_{0,S}(u) e^{\xi_1}\mathrm{d}u\right),\\
&\mathbb{P}(S_1 > S_2 \mid \xi_1, \xi_2) = \frac{e^{\xi_1}}{e^{\xi_1} + e^{\xi_2}},\\
&f(\xi_1, \xi_2) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\begin{pmatrix}\xi_1 & \xi_2\end{pmatrix}\Sigma^{-1}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}\right).
\end{aligned}$$

In the case where there is no effect of S on T (i.e. independence between S and T given the random

effects or $\gamma(S) = 0$), this expression reduces to

$$\tau = \int_{\mathbb{R}^4} \left[\frac{e^{\omega_2+u_2} + e^{\zeta\omega_2+\alpha u_2}}{\left(e^{\omega_1+u_1} + e^{\omega_2+u_2}\right)\left(e^{\zeta\omega_1+\alpha u_1} + e^{\zeta\omega_2+\alpha u_2}\right)} \right] \frac{1}{4\pi^2\theta\rho} \exp\left(-\frac{1}{2}\left(\frac{\omega_1+\omega_2}{\theta^2} + \frac{u_1+u_2}{\rho^2}\right)\right) \mathrm{d}\omega_1 \mathrm{d}\omega_1 \mathrm{d}u_1 \mathrm{d}u_2. \tag{7}$$

which can be easily approximated using Monte Carlo sampling (Sofeu et al., 2019).

(b) Kendall's τ for longitudinal surrogate

For Model (2), the definition of the Kendall's τ given in Equation 5 can't be used as the event "$M_1 > M_2$" does not make sense for longitudinal processes.
We propose to take a time-dependent Kendall's τ defined as,

$$\tau(t) = 4\mathbb{P}\left(M_1(t) > M_2(t), T_1 > T_2\right) - 1 \tag{8}$$

which now makes sense since $M_i(t)$ is simply a Gaussian variable. Following the definition of Model (2), we have

$$M_i(t) = (\beta + \omega_i)' f(t)$$

where $\omega_i \sim \mathcal{N}(0, D)$. Therefore, $M_i(t)$ is a Gaussian variable whose mean is $\beta' f(t)$ and variance $f(t)' D f(t)$.

$$\begin{aligned}\mathbb{P}\left(M_1(t) > M_2(t), T_1 > T_2\right) &= \mathbb{P}\left(T_1 > T_2 \mid M_1(t) > M_2(t)\right)\mathbb{P}\left(M_1(t) > M_2(t)\right) \\ &= \frac{\mathbb{P}\left(T_1 > T_2 \mid M_1(t) > M_2(t)\right)}{2}\end{aligned}$$

since $\mathbb{P}\left(M_1(t) > M_2(t)\right) = \mathbb{P}(M_1(t) < M_2(t))$. The probability $\mathbb{P}\left(T_1 > T_2 \mid M_1(t) > M_2(t)\right)$ can be computed by conditioning and averaging over ω_1 and ω_2. Since the event $M_1(t) > M_2(t)$ is equivalent to $\omega_1' f(t) > \omega_2' f(t)$, we have:

$$\mathbb{P}\left(T_1 > T_2 \mid M_1(t) > M_2(t)\right) = \int_{\mathbb{R}} \left(\int_{x_2}^{\infty} \mathbb{P}\left(T_1 > T_2 \mid U_1 = u_1, U_2 = u_2\right) f_{U_1}(u_1) \mathrm{d}x_1 \right) f_{U_2}(u_2) \mathrm{d}u_2$$

where for simplicity we note $U_i = \omega_i' f(t) \sim \mathcal{N}(0, f(t)' D f(t))$. The probability $\mathbb{P}\left(T_1 > T_2 \mid U_1 = u_1, U_2 = u_2\right)$

can be computed as,

$$\mathbb{P}\left(T_1 > T_2 \mid U_1 = u_1, U_2 = u_2\right) = \int_0^\infty \mathbb{P}\left(T_1 > t \mid X_1 = x_1\right) f_{T_2|X_2=x_2}(t)\mathrm{d}t$$

where the quantities

$$\mathbb{P}\left(T_1 > t \mid X_1 = x_1\right) = \exp\left(-\int_0^t \lambda_{0,T}(u)\exp(\eta' h(x_1))\mathrm{d}u\right)$$
$$f_{T_2|X_2=x_2}(t) = \lambda_{0,T}(t)e^{\eta' h(x_2)}\exp\left(-\int_0^t \lambda_{0,T}(u)\exp(\eta' h(x_2))\mathrm{d}u\right),$$

can be derived from the model.

The standard error of the estimated Kendall's τ, either $\hat{\tau}$ or $\hat{\tau}(t)$, can be obtained using parametric bootstrap. This can be achieve by sampling parameters of the models based on the estimated distribution of the maximum likelihood estimator and computing $\hat{\tau}$ or $\hat{\tau}(t)$ for each simulated vector of parameters. Then, an estimation of the standard error as well as quantile-based confidence interval can be computed from the resulting sample.

3.2.2 Trial-level association

For both Model (1) and Model (2), the trial-level association is defined as the coefficient of determination of the random effects (ν_S, ν_T) or (ν_M, ν_T). For both models these random effects are assumed to be Gaussian with mean 0 and covariance matrix Σ_ν. If we write,

$$\Sigma_\nu = \begin{pmatrix} \sigma^2_{\nu_1} & \sigma_{\nu_{12}} \\ \sigma_{\nu_{12}} & \sigma^2_{\nu_2} \end{pmatrix}.$$

Then, the trial-level measure of surrogacy, R^2_{trial} is defined as:

$$R^2_{\text{trial}} = \frac{\sigma^2_{\nu_{12}}}{\sigma^2_{\nu_1}\sigma^2_{\nu_2}}.$$

As for the Kendall's τ, standard error for the estimator of $\hat{R}^2_{\text{trial}}$ can be obtained using parametric bootstrap. However, the simple form of R^2_{trial} allows the derivation of an estimator of its standard error directly from the estimated covariance matrix of the vector of parameter using the delta-method and then using the normal approximation to derive a confidence interval. However the later could yield confidence intervals that exceeds the range $[0, 1]$ of R^2_{trial}.

3.2.3 Surrogate threshold effect

The surrogate threshold effect is the smallest treatment effect on the surrogate endpoint required to yield a significant effect on the final endpoint (Burzykowski and Buyse, 2006).

The prediction interval for the treatment effect on the final endpoint in a new trial i_0, $\beta_{Z,T} + \nu_{T,i_0}$, given that the observed effect on the surrogate is $\beta_{Z,S} + \nu_{S,i_0}$ is:

$$\mathbb{E}\left(\beta_{Z,T} + \nu_{T,i_0} \mid \nu_{S,i_0}\right) \pm z_{1-\gamma/2}\sqrt{\mathbb{V}\left(\beta_{Z,T} + \nu_{T,i_0} \mid \nu_{S,i_0}\right)},$$

where $z_{1-\gamma/2}$ is the $(1-\gamma/2)$ percentile of the standard normal distribution. In order to predict a significant effect of the treatment on the final endpoint, the lower bound of this interval should be greater than 0 (or the upper bound lower than 0 depending on the type of effect one is interested in). This interval only depends on ν_{S,i_0}. Let $l(\nu_{S,i_0})$ be the lower bound of this interval, then the STE is the value of ν_{S,i_0} such that

$$l(\nu_{S,i_0}) = 0.$$

3.2.4 Proportion of treatment effect

The PTE is a measure related to mediation analysis. Mediation analysis investigates how the total effect of a treatment on a outcome can be decomposed as a direct effect and an indirect effect through a mediator (intermediate variable). Most development of mediation analysis have been made in the realm of causal inference, mostly based on counterfactual outcomes. In our case, the mediator is the surrogate endpoint of interest and the endpoint is the final endpoint. Let,

$$\mathbb{S}^{zz'}(t) = \mathbb{P}(T(z, S(z') > t),$$

be the survival function of the final endpoint T if the treatment for T is set to z but to z' for the surrogate S (here we do not make a distinction between a time-to-event surrogate or a longitudinal surrogate). Details on the interpretations of the above formula have been published Le Coënt et al., 2022.

For a binary treatment we can derive its natural indirect effect (NIE) on T through S as:

$$\text{NIE}(t) = \mathbb{S}^{11}(t) - \mathbb{S}^{10}(t).$$

The righthand side of this formula is the difference between the survival function of T in the case where the treatment is set to 1 for both S and T and the survival function of T in the case where the treatment is set to 1 for T but to 0 for S. The only difference between $\mathbb{S}^{11}(t)$ and $\mathbb{S}^{10}(t)$ is the treatment value for S, hence their difference quantifies the amount of treatment effect (on the survival scale) on T due to

its effect on S. In the same manner we can define the natural direct effect as:

$$\text{NDE}(t) = \mathbb{S}^{10}(t) - \mathbb{S}^{00}(t).$$

The total treatment effect (TTE) is then the sum of the NIE and NDE, $\text{TTE}(t) = \text{NIE}(t) + \text{NDE}(t)$. Finally, the measure of surrogacy is defined as:

$$\begin{aligned} \text{PTE}(t) &= \frac{\text{NIE}(t)}{\text{TTE}(t)} \\ &= \frac{\mathbb{S}^{11}(t) - \mathbb{S}^{10}(t)}{\mathbb{S}^{11}(t) - \mathbb{S}^{00}(t)}. \end{aligned} \tag{9}$$

which is the proportion of treatment effect going through the surrogate. To be able to compute the function $\text{PTE}(t)$ we need to compute the functions $\mathbb{S}^{zz'}(t)$, which can be done from Model (1) and Model (2) using the formula (Le Coënt et al., 2022; Zheng and Liu, 2021),

$$\begin{aligned} \mathbb{S}^{zz'}(t) &= \int_{\mathcal{X}} \int_0^{\infty} \int_{\xi} \mathcal{S}_T(t \mid s, z, \xi, X = x) f_S(s \mid z', \xi, x) f_\xi(\xi) f_X(x) \mathrm{d}s \mathrm{d}\xi \mathrm{d}x \quad \text{for Model (1)} \\ \mathbb{S}^{zz'}(t) &= \int_{\mathcal{X}} \int_{\xi} \mathcal{S}_T\left(t \mid h(M^{z'}), \xi, z, x\right) f_\xi(\xi) f_X(x) \mathrm{d}\xi \mathrm{d}x \quad \text{for Model (2),} \end{aligned} \tag{10}$$

where $f_\xi(\cdot)$ and $f_X(\cdot)$ are the density functions of the random-effects and covariates in the models respectively. In the first equation, $S_T(t \mid s, z', \xi, x)$ is the survival function of T evaluated at t, given the treatment is set to z, the surrogate $S = s$, the random effects ξ and covariates x. The function $f_S(s \mid z, \xi, x)$ is the density of S given the treatment is set to $Z = z'$, the random effect and covariates. In the second equation, $S_T\left(t \mid h(M^{z'}), \xi, Z = z, X = x\right)$ is the survival function of T given the treatment $Z = z$, the random effects x, the covariates x and the link function $h(M^{z'})$ given the treatment for M is set to z'. Note that in the second case, in contrary to the first case, we do not need to integrate over the distribution of $M^{z'}$ since it only depends on z', the random effects ξ, and the covariates x. From Equation 10, $\text{PTE}(t)$ can be obtained by setting the appropriate values for z and z' following Equation 9. All the functions involved in the righthand side of Equation 10 can be derived from the estimated parameters of the model, except $f_X(\cdot)$ for which averaging over the observed distribution of the covariates in the dataset can be used to approximate the integral over $\mathcal{X}$. As for the Kendall's τ and R^2_{trial}, the standard error and confidence interval of $\text{PTE}(t)$ can be obtained through parametric bootstrap.

3.3 Implementation in frailtypack

In this Section we present two main functions of the `frailtypack` package for investigating surrogacy and their main arguments. A detailed explanation of these functions can be found in the reference manual of the package available in R or on the CRAN (Rondeau and Gonzalez, 2005).

3.3.1 Function jointSurroPenal

The function `jointSurroPenal` can be used to investigate surrogacy when both the surrogate S and the final endpoint T are time-to-event (model (1)). The call to this function is as follows (the values given for each parameters are the default values):

```
model<- jointSurroPenal(data, maxit = 50,
          indicator.zeta = 1,
          indicator.alpha = 1,frail.base = 1,
          n.knots = 6, LIMparam = 0.001,
          LIMlogl = 0.001, LIMderiv = 0.001,
          nb.mc = 300,nb.gh = 32, nb.gh2 = 20,
          adaptatif = 0, int.method = 2,
          nb.iterPGH = 5, nb.MC.kendall = 10000,
          nboot.kendall = 1000, true.init.val = 0,
          theta.init = 1, sigma.ss.init = 0.5,
          sigma.tt.init = 0.5, sigma.st.init = 0.48,
          gamma.init = 0.5,alpha.init = 1, zeta.init = 1,
          betas.init = 0.5, betat.init = 0.5,
          scale = 1, random.generator = 1, kappa.use = 4,
          random = 0,random.nb.sim = 0, seed = 0,
          init.kappa = NULL, ckappa = c(0, 0), nb.decimal = 4,
          print.times = TRUE, print.iter = FALSE,
          mediation = FALSE, g.nknots = 1, pte.times = NULL,
          pte.ntimes = NULL, pte.nmc = 500, pte.boot = FALSE,
          pte.nboot = 2000, pte.boot.nmc = 500,
          pte.integ.type = 2)
```

In order to use this function, the user has to provide a dataset (argument `data`) of the following structure:

```
head(data)
patientID    timeT    timeS statusT statusS trt trialID
```

```
1 9.057946 2.217739       1       1   0       1
2 2.986813 1.389263       1       1   0       1
3 8.874237 8.874237       1       0   1       1
4 3.245388 1.809671       1       1   1       1
5 4.448964 2.603604       1       1   0       1
```

The dataset must contain one line per subject and seven columns: one for the subject's identification number (column `patientID`), for the trial number (`trialID`), treatment indicator (`trt`) as well as one for the follow-up time for the surrogate (`timeS`) and censoring indicator (`statusS`) and for the follow-up time for the final endpoint (`timeT`) and censoring indicator (`statusT`).

The arguments `indicator.zeta` and `indicator.alpha` are used to indicate if one wants to estimate the parameters ζ and α in Model (1). If they are set to 0 then these parameters will not be estimated and assumed to be equal to 1, which means that the random effects will have the same effect on S and T. The parameter `frail.base` indicates if the user wants to include the trial-level random effects u_i in the model. If set to 0 then these random effects are not included, and neither will be the parameter α associated with u_i. The number of knots used in the M-splines basis for estimating the functions $\lambda_{0,S}(\cdot)$ and $\lambda_{0,T}(\cdot)$ can be set through the argument `n.knots`, with allowed values between 4 and 20 knots.

For the parametric bootstrap used to derive the standard error and confidence interval for the estimated Kendall's τ, the argument `nboot.kendall` can be used to fix the number of bootstrap samples to be generated.

The option to include the function $\gamma(S)$ in Model (1) is given by setting the argument `mediation` to TRUE. In that case the function $\gamma(S)$ is estimated using a basis of B-splines whose number of knots is given by the argument `g.nknots`, which can take any value between 1 and 5 and the function PTE(t) will also be estimated. The timepoints at which this function has to be evaluated can be specified through the argument `pte.times`. Note that these times should be between 0 and the maximum observed final endpoint times. If one does not want to specify any timepoints, the argument `pte.ntimes` can be used instead to specify the number of timepoints at which PTE(t) should be evaluated. These points will then be selected evenly on the range of the observed event times. The argument `pte.boot` is used if we want to compute quantile-based confidence bands of $\hat{\text{PTE}}(t)$ using parametric bootstrap. If set to TRUE, then the number of bootstrap samples to be used can be set with `pte.nboot`.

Other parameters are mainly pertain to computational details. A complete description of each parameter can be found in the documentation of the package `frailtypack` available on the CRAN.

The function `jointSurroPenal` returns an R object of class `jointSurroPenal` if the argument `mediation` is set to FALSE and of class `jointSurroMed` otherwise. In both cases, common R functions such as `summary`, `print` and `plot` can be used as will be illustrated in Section 4.

3.3.2 Function longiPenal

The function `longiPenal` can be used to investigate surrogacy when the surrogate outcome is a longitudinal biomarker and the final endpoint is a time-to-event (Model (2)). The call to this function is as follows (the values given to each parameters are the default values):

```
model<- longiPenal(formula, formula.LongitudinalData, data,
          data.Longi,formula.Binary=FALSE,
          random,random.Binary=FALSE,
          fixed.Binary=FALSE, GLMlog=FALSE,
          MTP=FALSE, id, intercept = TRUE,
          link="Random-effects",timevar=FALSE,
          left.censoring=FALSE,n.knots, kappa,
          maxit=350, hazard="Splines", init.B,
          mediation=FALSE,centers=NULL,treatment=NULL,
          pte.times=NULL,pte.ntimes=NULL,pte.boot=FALSE,
          pte.nboot=2000, pte.nmc=500,init.Random, init.Eta,
          method.GH = "Standard",seed.MC=1, n.nodes,
          LIMparam=1e-3,LIMlogl=1e-3,
          LIMderiv=1e3,print.times=TRUE,method_int=1,
          nquad=9,nmcmed=100L,Cmult=0L)
```

This function requires the specification of two datasets. The first one, specified through the argument `data`, contains the data regarding the final endpoint T such as the follow-up time for each subject, the censoring indicator, and potential baseline covariates. Note that this dataset requires one line per subject and therefore does not allow for time-dependent covariates to be included. Associated with this dataset is the `formula` wich is a formula object, with the response on the left of a $\sim$ operator, and the covariates on the right. The response must be a survival object as returned by the `Surv` function of the R `survival` package. The variables used in `formula` should be the ones contained in `data`. For the longitudinal part, the repeated measurements data are specified in a separate dataset through the argument `data.Longi`. The specification of the longitudinal submodel is made through `formula.LongitudinalData` which is a R formula with the observed biomarker on the left and the different covariates on the right. Both the names for the biomarker and the covariates specified in this formula should correspond to columns in the dataset `data.Longi`.

Both `data` and `data.Longi` should have a column labelled "`id`" that corresponds to the identificator of each subject in order to link the two datasets, i.e., `id=1` in `data` should corresponds to the same individual with `id=1` in `data.Longi`. Note that for simpliciy the variable `id` should takes

values between 1 and n_i where n_i is the total number of subjects.

The argument `intercept` can be used to indicate if one wants to include the fixed intercept β_0 in the longitudinal submodel, with default value being TRUE.

The argument `timevar` is used to specify the timepoints of the repeated measurements in `data.Longi` for each individual. This variable can be included in `formula.LongitudinalData` but is not required, a transformation of `timevar` can be used instead to take into account non linear time evolution. However, in both case a variable `timevar` must be included in the dataset which contains the measurement times.

The individual-level random effects to be included in the longitudinal submodel are specified through the argument `random` which must be a vector of characters corresponding to covariates in `data.Longi`. The character `"1"` is used to indicate a random intercept. For example if one wants to include a random intercept and a random slope then one should specify `random=c("1","timevar")` (given that `timevar` has been included in `formula.LongitudinalData`).

The baseline hazard function estimation type is chosen via the argument `hazard`. Possible values are `"Splines"` and `"Splines-per"` for flexible hazard functions estimated with M-Splines with respectively equidistant or percentile knots, or `"Weibull"` to specify a parametric Weibull hazard function. For `"Splines"` and `"Splines-per"` the number of knots used is set through the argument `n.knots`.

The mediation analysis is enabled by setting the argument `mediation` to TRUE. In that case one should also specify the name of the variable in `data` that corresponds to the treatment through the argument `treatment`. Moreover, one should also specify the centers (or trials) to which each subject belongs through the argument `centers`. This argument takes a character string which must correspond to the name of the variable in `data` indicating the center/trial of each subject. Note that as the variable `id`, the variable indicating the center for each subject should take values between 1 and K where K is the total number of centers/trials. If `mediation` is set to TRUE then the function PTE(t) will be estimated. As for the function `jointSurroPenal`, one can specify the timepoints at which $\hat{\text{PTE}}(t)$ should be evaluated or the number of timepoints at which it should be evaluated. The argument `pte.boot` takes values TRUE/FALSE to indicate if the parametric bootstrap estimation of the standard-error of $\hat{\text{PTE}}(t)$ and its confidence bands should be computed. If set to TRUE then the number of bootstrap samples is specified by `pte.nboot` with a default value of 2000. A complete description of each parameter can be found in the documentation of the package `frailtypack` available on the CRAN.

The function `longiPenal` returns a R object of class `longiPenal` on which the usual R functions `summary`, `print` and `plot` can be applied as will be illustrated in Section 4.

4 Illustrations

We illustrate the proposed methods in three applications on cancer data from meta-analyses or multicentric randomized clinical trial. The first application is based on a dataset on ovarian cancer, the second on gastric cancer and the third on colorectal cancer.

4.1 Progression-free survival as a surrogate of overall survival in advanced ovarian cancer: meta-analytic approach

In this first application we are interested in evaluating progression-free survival as a surrogate endpoint for overall survival in advanced ovarian cancer using the classical meta-analytic approach and its two-level validation. The dataset is the `dataOvarian` which can be loaded from `frailtypack` using the following command:

```
data("dataOvarian",package = "frailtypack")
```

This dataset combines the data of four double-blind randomized clinical trials. These trials investigated the efficacy of cyclophosphamide plus cisplatin versus cyclophosphamide plus adriamycin plus cisplatin in advanced colorectal cancer (Ovarian Cancer Meta-Analysis Project, 1991). The structure of the dataset is as follows:

```
head(dataOvarian)
```

```
  patientID trialID trt      timeS statusS     timeT statusT
1         1       2   0 0.10515873       1 0.1857143       1
2         2       2   0 0.89523809       1 1.4087302       1
3         3       2   0 0.07896825       1 0.1261905       1
4         4       2   1 1.73928571       0 1.7392857       0
5         5       2   0 0.09126984       1 0.1273810       1
6         6       2   1 0.16984127       1 0.2253968       1
```

The columns `dataOvarian` and `trialID` indicate the identifiers of the subject and the trial respectively. The variable `trt` takes value 0 for cyclophosphamide plus cisplatin and 1 for cyclophosphamide plus adriamycin plus cisplatin. The variables `timeS` and `statusS` are the follow-up and censoring indicator of the time-to-progression (the surrogate) while `timeT` and `statusT` are the follow-up and censoring indicator of the time-to-death (from any cause).

4.1.1 Model fitting and surrogacy evaluation

Once the dataset has been loaded, we can call the `jointSurroPenal` function as follows:

```
require(frailtypack)
mod.ovar <- jointSurroPenal(data = dataOvarian, n.knots = 8,
                            indicator.alpha = 0,
                            nb.mc = 500, scale = 1/365)
```

Here we do not specify the option `mediation`=TRUE (the default value being FALSE), therefore the approach used corresponds to a "classical" meta-analytic approach rather than a mediation approach and is based on Model (1) without inclusion of the function $\gamma(S)$.

We specify that the number of knots (inner + boundary) in the M-splines basis for estimating the two baseline hazard functions $\lambda_{0,S}$ and $\lambda_{0,T}$ is 8 (`n.knots`=8). The argument `indicator.alpha = 0` indicates that we don't estimate the parameter α which will be fixed to 1 throughout the estimation procedure (see Model (1)). The argument `nb.mc = 500` specifies that the number of Monte Carlo points used in the approximation of the integral on the trial-level random effects is set to 500. Finally `scale = 1/365` is used to rescale the follow-up times in days (originally expressed in years) .
The object `mod.ovar` belongs to the class `jointSurroPenal` on which we can apply the R `summary` function.

The results of the estimation can be simply outputed using the command `mod.ovar`. The command `summary` only displays estimators of the estimated fixed treatment effects and surrogacy measures.

```
mod.ovar
```

```
Estimates for variances parameters of the random effects

Individual-level random effect
                    Estimate Std Error      z      P
 Variance (theta^2)    6.852    0.3769 18.180 < e-10 ***
 Association (zeta)    1.785    0.0621 28.739 < e-10 ***

Trial-level random effect
                  Estimate Std Error      z      P
 Variance (rho^2)    1.785    0.0621 28.739 < e-10 ***

Trial-level treatment heterogeneity
         Estimate Std Error     z      P
sigma2_S    0.460    0.3712 1.239 0.2154
sigma2_T    1.501    1.0543 1.423 0.1547
sigma_ST    0.831    0.6189 1.342 0.1795


Estimates for the fixed treatment effects
       Estimate Std Error      z        P
beta_S   -0.699    0.2210 -3.164 0.001555 **
beta_T   -1.037    0.3869 -2.680 0.007362 **
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

hazard ratios (HR) and confidence intervals for the fixed treatment effects
       exp(coef) Inf.95.CI Sup.95.CI
beta_S     0.497     0.322     0.766
beta_T     0.355     0.166     0.757

Surrogacy evaluation criterion
             Level Estimate Std Error Inf.95.CI Sup.95.CI Strength
Ktau    Individual    0.675        --     0.662     0.693
R2trial      Trial    1.000     0.002     0.995     1.005     High
R2.boot      Trial    0.968        --     0.738     1.000     High
---
Correlation strength: <= 0.49 'Low'; ]0.49 - 0.72[ 'Medium'; >= 0.72 'High'
---

The treatment effects on the surrogate endpoint (beta_S) that can predict a nonzero
treatment effect on the true endpoint (beta_T) belongs to the interval: ]-Inf ; -0.315[ : HR= ]0 ; 0.73[

Surrogate threshold effect (STE) : -0.315 (HR = 0.73 )

Convergence parameters
Penalized marginal log-likelihood =  -10892.287
Number of iterations =  31
Smoothing parameters =  60093516.4969293 0.000167782176816666
Number of spline nodes =  8
LCV = the approximate likelihood cross-validation criterion
      in the semi parametrical case     =  9.161
Convergence criteria:
  Parameters =  0.0009353 Likelihood =  3.562e-06 Gradient =  7.698e-10
Estimation based on a combination of both Gaussian-Hermite quadrature and Monte Carlo integration
```

The first part of the results shows the estimated variance of the individual and trial-level random effects and the power parameters associated with ω_{ij} and u_i in the hazard function of T (respectively α and ζ). Note that here we only have an estimation for ζ since we specified `indicator.alpha=0`. These results show a significant variance of the individual-level random effect ω_{ij} (variable `theta^2`). Since the parameter ζ is also significantly different from 0, both S and T are positively correlated at the individual level. On the other hand, the trial-level random effects variances are not significantly different from 0. The variable `rho^2` is the variance of the random effect u_i while `sigma2_S`, `sigma2_T` and `sigma_ST` are respectively the estimations of the variance of $\nu_{S,i}$, the variance of $\nu_{T,i}$ and their covariance.

The second part of the results show the estimated fixed treatment effects, their estimated standard errors and the associated Wald statistics and p-values. There are significant treatment effects on both endpoints, therefore the addition of adriamycin to cyclophosphamide plus cisplatin reduces the risk of the occurrences of progression and death.

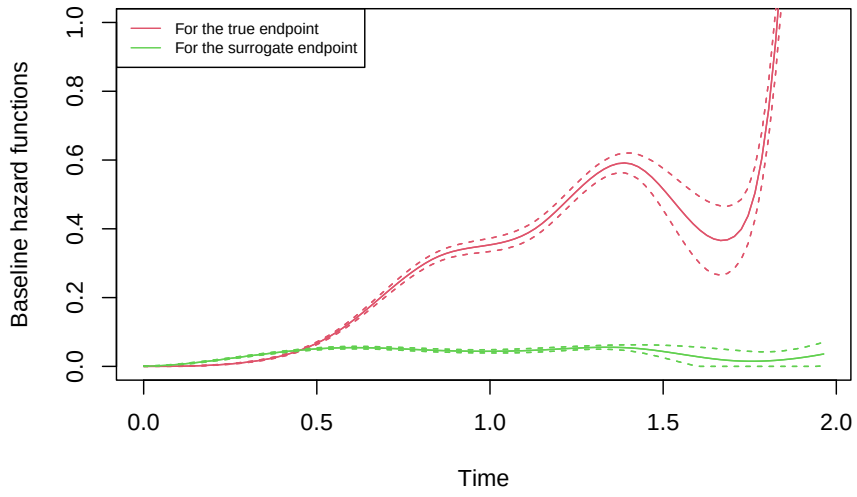
The third part of the results shows the hazard-ratios associated with the estimated fixed treatment together with their 95% confidence intervals, with a hazard ratio of 0.497 ($[0.322, 0.766]$) for the surrogate and 0.355 ($[0.166, 0.757]$) for the final endpoint.

The fourth part of the results shows the estimated surrogacy measures at the individual and trial level. The variable `Ktau` is the estimated Kendall's τ between the surrogate and the final endpoint with its associated 95% confidence interval. The estimated value of 0.675 suggests a mild association between the two endpoints. The variable `R2trial` is the estimated R^2_{trial} with its estimated standard error (obtained using the delta-method) and its 95% confidence interval while `R2.boot` obtained using parametric bootstrap. Their values close to 1 suggest a high association between the treatment effects on the surrogate and the final endpoint at the trial-level. Finally these results suggest that the time-to-progression can be considered as a good surrogate of the overall survival in this setting of advanced ovarian cancer.

Additionally, the results give the surrogate threshold effect estimated at -0.315 (associated hazard-ratio 0.73). This means that in a new trial, the minimal treatment effect on the surrogate to be observed in order to predict a significant treatment effect on the final endpoint is -0.315.

Moreover, the estimated baseline hazard functions $\lambda_{0,S}$ and $\lambda_{0,T}$ can be plotted directly from the `mod.ovar` object using the R function `plot`:

```
plot(mod.ovar, pos.legend = "topleft")
```



It should be noted that the results in this application have to be interpreted with caution. First, the few number of trials (6) in the dataset make the estimation of the R^2_{trial} complicated. Moreover, its

broad confidence interval and a point estimate of 1.0 may indicates convergence issues, even if the three convergence criteria (based on the likelihood, its first order derivative and its second order derivative) show that the model converged.

4.2 Time-to-relapse as a surrogate of overall survival using proportion of treatment effect in gastric cancer: a mediation approach

Our second application is on a meta-analysis on resectable gastric cancer patients investigating the addition of adjuvant chemotherapy after surgery versus surgery alone (Paoletti et al., 2010). In this illustration, the final endpoint is the time between randomization and death from any cause while the surrogate is the time-to-relapse, defined as the time between randomization and disease recurrence or occurrence of a second cancer, whichever occurred first. Therefore both endpoints might be right censored due to loss to follow-up and moreover the surrogate endpoint might be censored by the final endpoint. We are interested in estimating the proportion of treatment effect (adjuvant chemotherapy or not) on overall survival that goes through its effect on time-to-relapse.

4.2.1 Dataset

As for the first illustration, the dataset `gastadj` can be loaded directly from `frailtypack` using the command,

```
data("gastadj",package="frailtypack")
```

This dataset contains the data of 3288 patients from 14 randomized clinical trials. Out of these 3288 patients, 1654 were assigned to the control group of no adjuvant chemotherapy, and the remaining 1634 patients were assigned to receive adjuvant chemotherapy. The dataset has the following structure, using the command `head(gastadj)`,

```
head(gastadj)
```

```
  trialID patientID trt timeT statusT timeS statusS
1       1         1   1  4636       0  4636       0
2       1         2   1  4536       0  4536       0
3       1         3   0  3151       1  3151       1
4       1         4   1   485       1   432       1
5       1         5   0   435       1   300       1
6       1         6   0   187       1   137       1
```

The columns `trialID, patientID, trt` are the trial, patient and treatment indicator respectively. The variables `timeT, timeS` are the follow-up times for the final endpoint and surrogate endpoints respectively and `statusT`, `statusS` their associated censoring indicator.

In this dataset, the variable `timeS` corresponds to a time-to-progression defined as the earliest between cancer recurrence, occurrence of a second cancer or death. Therefore this endpoint includes death as a composite endpoint which raises questions from a mediation analysis viewpoint since the final endpoint always triggers the surrogate. To circumvent this, we instead analyzed the time-to-relapse (cancer recurrence or second cancer) by censoring them at the time of death. In the dataset this change can be made using the following command,

```
gastadj[gastadj$timeS==gastadj$timeT &
        gastadj$statusS==1,c("statusS")]<-0
```

For practical purposes and to reduce the computation time, we restrain this illustration on a subset of the original dataset, by selecting 20% of the patients at random.

Moreover, to circumvent some computing issues, we divide the time variable (originally represented in a daily scale) by 365 in the yearly scale. Therefore, the full call for data preparation is

```
require(frailtypack)
data(gastadj)
gastadj$timeS<-gastadj$timeS/365
gastadj$timeT<-gastadj$timeT/365
#"statusS" corresponds now
#to a time-to-relapse event
gastadj[gastadj$timeS==gastadj$timeT &
        gastadj$statusS==1,c("statusS")]<-0

# select 20% of the original dataset
set.seed(1)
n<-nrow(gastadj)
subset<-gastadj[sort(sample(1:nrow(gastadj),
                round(n*0.2),replace = F)),]
```

4.2.2 Model fitting and surrogacy evaluation

The call to the function `jointSurroPenal` is the following:

```
mod.gast<-jointSurroPenal(subset,n.knots = 4,
                          indicator.zeta = 0,
                          indicator.alpha = 0,
                          mediation=TRUE,g.nknots=1,
                          pte.ntimes=30,pte.nmc=10000,
                          pte.boot=TRUE,rt.nboot=1000,
                          rt.boot.nmc=1000)
```

In this call we set both `indicator.zeta` and `indicator.alpha` to 0, therefore the parameters ζ and α in Model (1) are not estimated and assumed to be equal to 1.

Here we specify that `mediation`=TRUE, therefore the function $\gamma(S)$ in Model (1) will be estimated using B-Splines. The number of inner knots used in the spline basis is fixed to 1 via the command `g.nknots`=1. Since we are interested in the mediation analysis setting, we specify that we want the function PTE(t) to be evaluated at 30 timepoints by setting `pte.ntimes` to 30. The number of Monte-Carlo points used in the approximation of the integral over the random effects in Equation 10 is set to 10000. The use of parametric bootstrap to derive estimated standard-errors and confidence bands for PTE(t) is given by `pte.boot`=TRUE where we also specify that we want this bootstrap to be based on 1000 sampling via `rt.nboot`=1000. Finally, for illustration purposes and to reduce computation time, we also set the number of Monte Carlo points used for each bootstrap sample to 1000. However, in practice this number should be the same as for the estimation of PTE(t). The object `mod.gast` has R class `jointSurroMed`, and we can apply the `summary` function to display the results.

```
summary(mod.gast)
```

```
Results of a joint surrogate mediation model using a penalized likelihood
on the baseline hazard functions.

Estimates for variances parameters of the random effects

Individual-level random effect
                    Estimate Std Error     z         P
 Variance (theta^2)    0.398    0.0743 5.359 8.352e-08 ***

Trial-level random effect
                  Estimate Std Error     z         P
 Variance (rho^2)    0.398    0.0743 5.359 8.352e-08 ***

Trial-level treatment heterogeneity
         Estimate Std Error      z      P
sigma2_S    0.032    0.0272  1.173 0.2409
sigma2_T    0.009    0.0118  0.720 0.4713
sigma_ST   -0.016    0.0133 -1.203  0.229
```

```
Estimates for the fixed treatment effects
       Estimate Std Error      z        P
beta_S   -0.248    0.0771 -3.215 0.001303 **
beta_T    0.095    0.0675  1.404   0.1602
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hazard ratios (HR) and confidence intervals for the fixed treatment effects
       exp(coef) Inf.95.CI Sup.95.CI
beta_S     0.780     0.671     0.908
beta_T     1.099     0.963     1.255

Individual and trial level associations
             Level Estimate Std Error Inf.95.CI Sup.95.CI Strength
Ktau    Individual    0.252        --     0.210     0.293
R2trial      Trial    0.945     0.376     0.209     1.681      Low
R2.boot      Trial    0.636        --     0.006     0.999      Low
---
Correlation strength: <= 0.49 'Low'; ]0.49 - 0.72[ 'Medium'; >= 0.72 'High'
---

Estimated function g at 3 times of occurence of the surrogate
    Time     g          CI.95
1  4.683 3.563  [3.321;3.805]
2  9.388 3.559   [2.999;4.12]
3 14.094 0.801 [-2.345;3.947]

Estimated function PTE(t), natural direct, indirect and total effect at 3 time points
    Time   PTE      CI.95.PTE    TE       CI.95.TE    NDE      CI.95.NDE   NIE
1  6.939 1.334  [0.581;3.842] 0.030       [0;0.06] -0.010 [-0.025;0.007] 0.040
2 14.644 1.178  [0.729;4.216] 0.035  [0.001;0.057] -0.006 [-0.024;0.006] 0.041
3 22.348 1.669 [-3.315;6.775] 0.026 [-0.007;0.048] -0.017 [-0.025;0.006] 0.044
      CI.95.NIE
1 [0.013;0.067]
2 [0.011;0.063]
3 [0.007;0.054]

Convergence parameters
Penalized marginal log-likelihood =  -6747.349
Number of iterations =  15
LCV = the approximate likelihood cross-validation criterion
      in the semi parametrical case     =  2.059
Convergence criteria:
  parameters =  4.57e-05 likelihood =  8.116e-06 gradient =  1.522e-09
```

As in the first illustration, we first have the estimation of the random effects variances. The parameters `theta^2` and `rho^2` (the variances of ω_{ij} and u_i in Model (1) respectively) are significantly different from 0; there is an association between S and T at the individual-level (through ω_{ij}) and trial-level (u_i). However, the variances of ν_S and ν_T associated with the heterogeneity of the treatment effects between trials are not significantly different from 0.

The estimated fixed treatment effect on the surrogate and the final endpoint are respectively -0.248 and 0.095, while the latter is not significant. However this does not suggest no treatment effect on the final endpoint, because the value 0.095 is associated with the parameter $\beta_{Z,T}$ in Figure 1(b) which corresponds to the *direct* effect of the treatment on the final endpoint. A null direct effect does not mean no effect at all since the indirect effect may be different from 0.

Regarding the association measures of surrogacy we have an estimated Kendall's τ of 0.252 with a 95% confidence interval $[0.210, 0.293]$ which suggests a poor individual-level surrogacy. At the trial-level the estimated R^2_{trial} is 0.945 but its 95% confidence interval is wide ([0.209,1.681]). The bootstrapped estimation and confidence interval are 0.636 and $[0.006, 0.999]$, which suggest not enough evidence to support a good trial-level surrogacy.

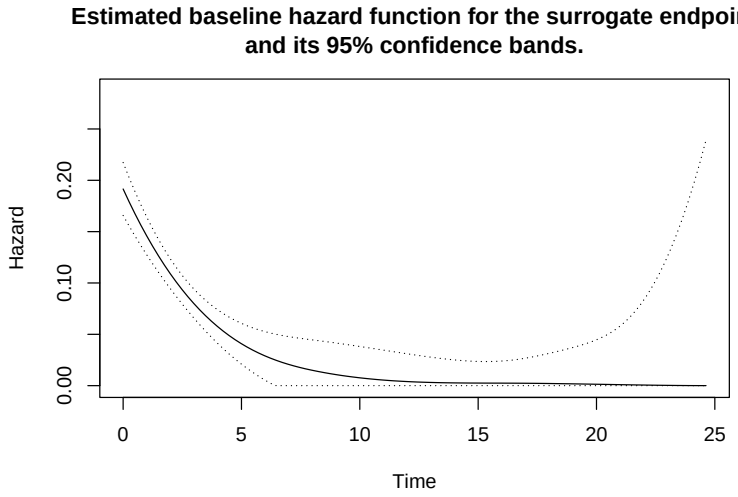
The next part of the results shows the value of the estimated function $\gamma(\cdot)$ at three timepoints. We can see that for the two first times the estimated values are significantly different from 0, which suggest that there is a direct link between the occurrence of the surrogate S and the risk of occurrence of T. Moreover the estimated function $\gamma(S)$ being positive translates as an increased risk of occurrence of T once the surrogate occurs. However for the third timepoint the estimated value of $\gamma(\cdot)$ is not significant anymore. This can be explained by a lack of power at latter timepoints as less patients are still under observation.

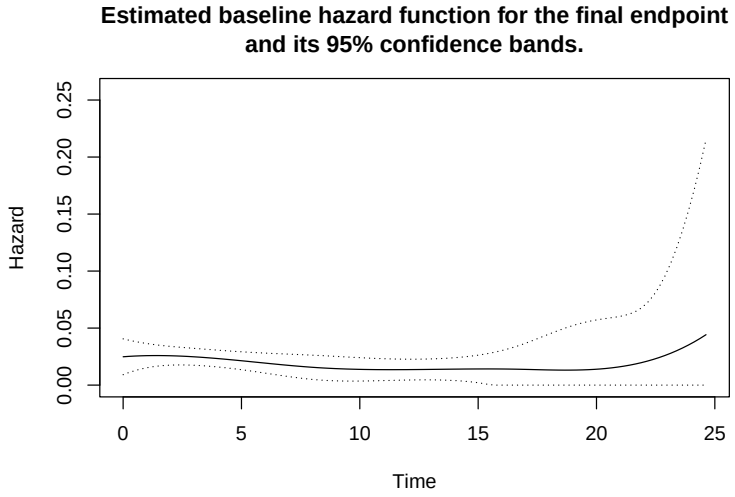
The next part of the results displays the estimated values of the function PTE(t) at the same three timepoints, as well as the estimated total treatment effect, TTE(t), natural indirect and direct effects, NIE(t) and NDE(t).

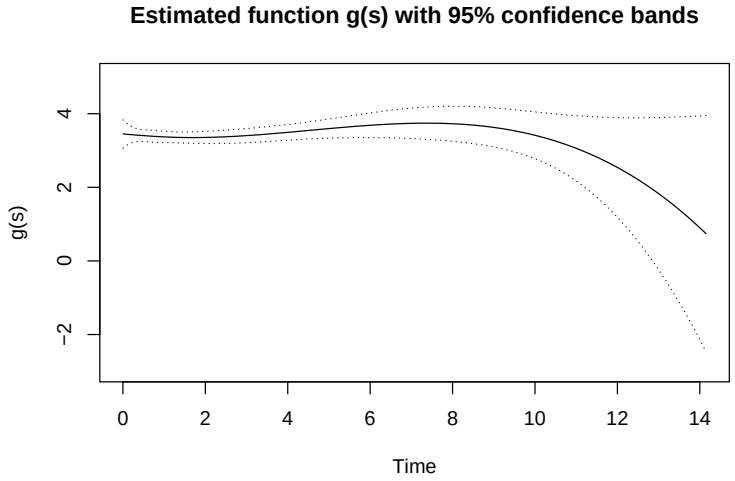
First, for PTE(t), we can see that for the two first timepoints, the estimated values of PTE(t) is high with a large confidence interval excluding the value 0, while for the third timepoint the conficence interval is even larger. This is partly due to a total treatment effect, TTE(t), being not significant at the third time point (confidence interval $[-0.007, 0.048]$), therefore PTE(t) defined as the ratio of NIE(t) over TTE(t) may be extremely instable if TTE$(t) \approx 0$, resulting in the wide confidence interval. For the direct effect, NDE(t) is not significant at the three timepoints, which is in agreement with the non significant estimated value of $\beta_{Z,T}$. For the indirect effect, the estimated values are significant and around 0.04. Recall that these effects are given in the survival scale. The interpretation of a NIE(t) of 0.04 is therefore that survival gain by setting $Z = 1$ for both T and S versus $Z = 1$ for T but 0 for S is about 4% at a time t.

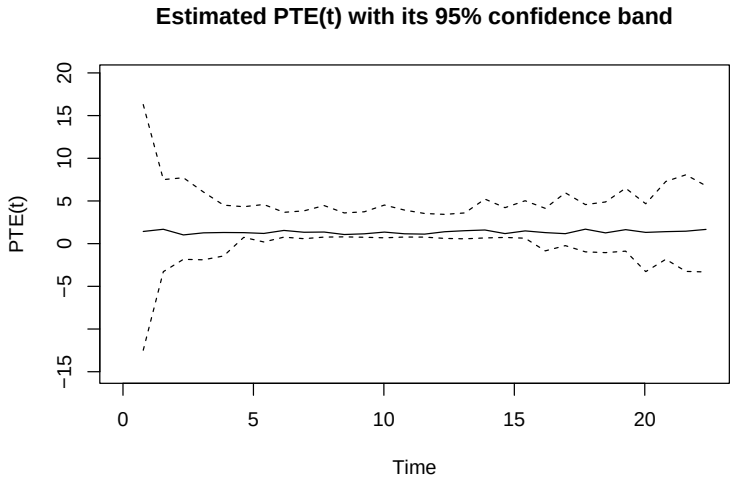
The mediation measures of surrogacy, essentially PTE(t) and associated NIE(t) and NDE(t), are inherently time-dependent. Therefore a graphical representation might be of interest when investigating surrogacy through theses measures. Especially the `plot` function can be applied to the object `mod.gast` to display the estimated results.

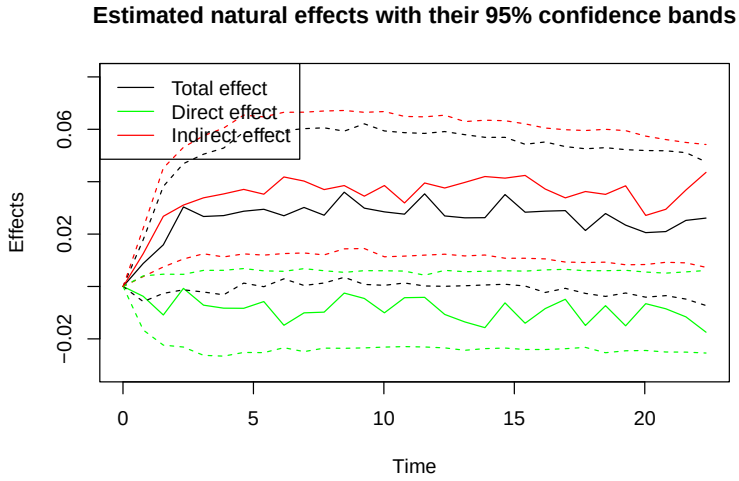
```
plot(mod.gast,type.plot="Survival",plot.mediation="All")
```






Estimated function g(s) with 95% confidence bands
g(s)
−2
0
2
4
0
2
4
6
8
10
12
14
Time


Estimated PTE(t) with its 95% confidence band
PTE(t)
−15
−5
0
5
10
15
20
0
5
10
15
20
Time



We first have the estimated baseline survival functions associated with the baseline hazard function $\hat{\lambda}_{0,S}(t)$ and $\hat{\lambda}_{0,T}(t)$ and their 95% confidence bands through the command `type.plot="Survival"`. Another option is to set `type.plot="Hazard"` to plot the $\hat{\lambda}_{0,S}(t)$ and $\hat{\lambda}_{0,T}(t)$. The argument `plot.mediation` can takes the values `"g"`, `"effects"`, `"pte"` or `"All"` depending if one want to plot only $\gamma(\cdot)$, the natural effects (NIE(t),NDE(t) and TTE(t)), the PTE(t) or all at the same time.

From these figures we see that the direct effect is indeed not significant, while the indirect effect is. However, even if the direct effect is not significantly different from 0, most of its confidence bands contains negative values therefore the NIE and NDE are in opposite direction. Therefore we have $0 \leq$ TTE $\leq$ NIE and thus PTE ≥ 1. Moreover, the large confidence interval of PTE can be explained by the confidence bands of TTE(t) containing values very close to 0, hence a large instability of PTE(t).

Finally, even if from an "association" perspective the time-to-relapse is a poor surrogate for overall survival, the mediation measure suggest that most of the treatment effect on overall survival goes through its effect on the time-to-relapse. However a limitation is the small total treatment effect which makes difficult the interpretation of the PTE(t) since it can be very unstable in that case.

4.3 Tumor size as a surrogate biomarker of overall survival in colorectal cancer: a mediation approach

In this third application we are interested in evaluating the tumor size evolution over time as a surrogate of the overall surival in colorectal cancer. Since the tumor size evolution is a longitudinal biomarker we will base the analysis on the function `longiPenal`.

We will use a dataset containing 150 patients randomly selected from the FFCD 2000-05 multicenter

phase III clinical trial (Ducreux et al., 2011). This trial originally included 410 patients with metastatic colorectal cancer randomized into two treatment strategies: combination and sequential chemotherapy. The dataset contains times of observed appearances of new lesions censored by a terminal event (death) with some baseline characteristics.

Unfortunately, the available dataset does not contain the identificator of the center of the patients (on which we could have based the "meta-analytic" approach). We therefore reduce to generate fictive center for each subject, with the number of center equals to 5. The data are actually composed of two datasets, one for the survival part and another containing the repeated measurements of tumor sizes. As in the two previous illustrations these two datasets can be loaded from `frailtypack`.

```
data(colorectal,package = "frailypack")
data(colorectalLongi,package="frailtypack")
```

The dataset `colorectal` contains several observations per subject, one for each new lesions in addition to a follow-up time and a censoring indicator for death.

Therefore we only want to retrieve the last observation for a subject. In this dataset the variable `new.lesions` takes the value 1 if a new lesion is record and 0 otherwise. Therefore if a subject has n_i observations, the observations $1, \ldots, n_{i-1}$ all have the status `new.lesions` equals to 1 (since the repeated follow-up are based on the appearance of new lesions). Hence, the last observation for each subject can be taken as the only one for which `new.lesions` equals 0:

```
colorectalSurv <- subset(colorectal, new.lesions == 0)
```

We also generate random centers for each subject,

```
set.seed(1)
colorectalSurv$centers<-sample(1:5,nrow(colorectalSurv),T)
```

In the dataset the variable `treatment` takes the value `"S"` for "sequential" and `"C"` for "combined", for interpretability we simply make this variable binary 0/1,

```
colorectalSurv$treatment<-sapply(colorectalSurv$treatment,
                                 function(t) ifelse(t=="S",1,0))
```

To keep the illustration simple we only adjust on the variable `age` as a categorical variable: <60 years, 60-69 years or >69 years. The call to the function is:

```
mod.col=longiPenal(Surv(time1, state) ~ age+treatment,
            tumor.size ~ age+year*treatment,
            data=colorectalSurv, data.Longi = colorectalLongi,
            random = c("1", "year"),id = "id",
            link = "Current-level",timevar="year",
            method.GH = "Pseudo-adaptive",
            mediation = TRUE,
            treatment = "treatment",
            pte.times=c(1,2,3),pte.boot=TRUE,
            pte.nboot=2000,pte.nmc=1000,
            centers = "centers",
            nmcmed = 500,n.knots = 7,
            kappa = 2)
```

In this call we fit a model using a "current-level" link function between the longitudinal biomarker and the final endpoint. We specify a random slope and intercept in the longitudinal submodel.

The argument `mediation=TRUE` indicates that we want to compute the natural direct and indirect effects as well as the proportion of treatment effect, PTE(t). We require that this function to be evaluated at three specific timepoints, $1, 2$ and 3, through the argument `pte.times`. Moreover, we also require that the bootstrap standard error and confidence interval for PTE(t) computed using 2000 samples.

The result can be displayed by applying the R function `print` to the object `mod.col`. This function may also takes an argument `ptetimes` if one wants to display the results regarding the mediation analysis only at specific times (note that the times specified in `ptetimes` must be contained in those specified `pte.times` in the call of the function).

```
print(mod.col,ptetimes=c(1,2,3))
```

```
Call:
longiPenal(formula = Surv(time1, state) ~ age, formula.LongitudinalData = tumor.size ~
    age + year, data = colorectalSurv, data.Longi = colorectalLongi,
    random = c("1", "year"), id = "id", link = "Current-level",
    timevar = "year", n.knots = 7, kappa = 2, mediation = TRUE,
    centers = colorectalSurv$centers, treatment = colorectalSurv$treatment,
    method.GH = "Pseudo-adaptive", method_int = 1, nmcmed = 500,
    type.ma = 3, timetype = 0, Cmult = 6L)


  Joint Model for Longitudinal Data and a Terminal Event
  Using meta-analytic data and mediation analysis
```

```
  Parameter estimates using a Penalized Likelihood on the hazard function

Longitudinal outcome:
----------------
                      coef SE coef (H) SE coef (HIH)         z          p
Intercept         2.322874    0.194668      0.194668 11.932497     <1e-16
age60-69 years    0.739922    0.196995      0.196995  3.756051 1.7262e-04
age>69 years      0.428919    0.211281      0.211281  2.030088 4.2348e-02
year             -0.529501    0.099566      0.099566 -5.318068 1.0488e-07
treatment         0.871515    0.205881      0.205881  4.233104 2.3049e-05
treatment x year -0.609927    0.263546      0.263546 -2.314308 2.0651e-02

     chisq df global p
age 14.431  2 0.000735

Terminal event:
-------------
                     coef exp(coef) SE coef (H) SE coef (HIH)         z
age60-69 years  0.055660  1.057238    0.254667      0.254667  0.218559
age>69 years   -0.076333  0.926508    0.224731      0.224731 -0.339663
treatment      -0.096558  0.907957    0.202280      0.202280 -0.477348
                        p
age60-69 years 8.2699e-01
age>69 years   7.3411e-01
treatment      6.3311e-01

    chisq df global p
age 0.285  2    0.867


Components of individual random-effects covariance matrix B1:

Intercept  2.300209 -0.531238
year      -0.531238  0.777910

Components of Trial-level random effects covariance matrix:

nu_M 5.05048e-12 2.77344e-12
nu_T 2.77344e-12 1.75415e-12

Association parameters:
                   coef        SE       z          p
Current level 0.329153 0.0802414 4.10204 4.0952e-05

Residual standard error:  0.932979  (SE (H):  0.026479 )


Estimated total treatment effect on the final endpoint

 Time     TTE           95% CI
    1 -0.0036 [-0.1016,0.0386]
    2 -0.0079   [-0.0977,0.06]
    3 -0.0056 [-0.0799,0.0817]
```

```
Estimated natural indirect treatment effect

 Time     NIE           95% CI
    1 -0.0211 [-0.1067,0.0071]
    2  -0.035 [-0.0949,0.0128]
    3 -0.0375 [-0.0764,0.0245]

Estimated natural direct treatment effect

 Time    NDE           95% CI
    1 0.0175 [-0.0355,0.0822]
    2 0.0271   [-0.04,0.0964]
    3 0.0319 [-0.0422,0.0977]

Estimated proportion of treatment effect

 Time    PTE             95% CI
    1 5.8065 [-10.6727,13.6746]
    2 4.4188 [-11.3467,12.0111]
    3 6.6781 [-11.0315,14.2366]

penalized marginal log-likelihood = -1678.41
Convergence criteria:
parameters = 0.00017 likelihood = 0.000358 gradient = 1.47e-06

LCV = the approximate likelihood cross-validation criterion
in the semi parametrical case     = 1.61403

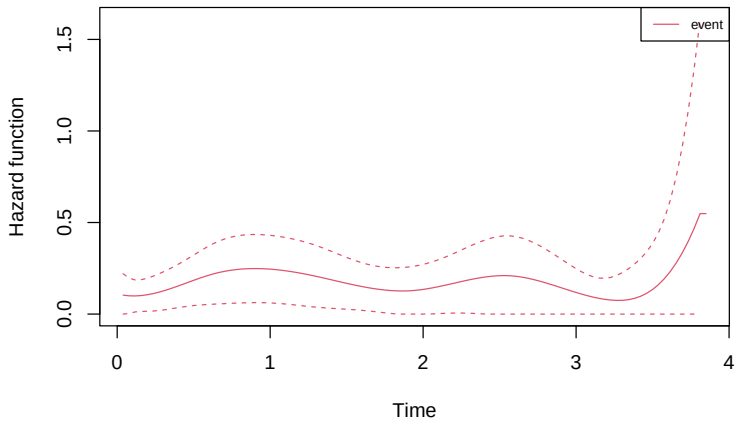
      n= 150
      n repeated measurements= 906
      n events= 121
      number of iterations:  20

      Exact number of knots used:  7
      Value of the smoothing parameter:  2
      Gaussian quadrature method:  Pseudo-adaptive with 9 nodes
```

From these results we see a significant effect of the treatment on the surrogate (at both the baseline level and through the slope), while the estimated $\beta_{Z,T}$ is not significantly different from 0. The individual random effects suggest a heterogeneity at baseline of the tumor size value. Moreover the estimated covariance matrix suggest no treatment heterogeneity across the trial, which is expected due to randomization (of the trials).

The estimated baseline hazard function $\hat{\lambda}_{0,T}(t)$ can be plotted using the R function `plot` with the additional argument `type="Hazard"`.

```
plot(mod.col,type="Hazard")
```

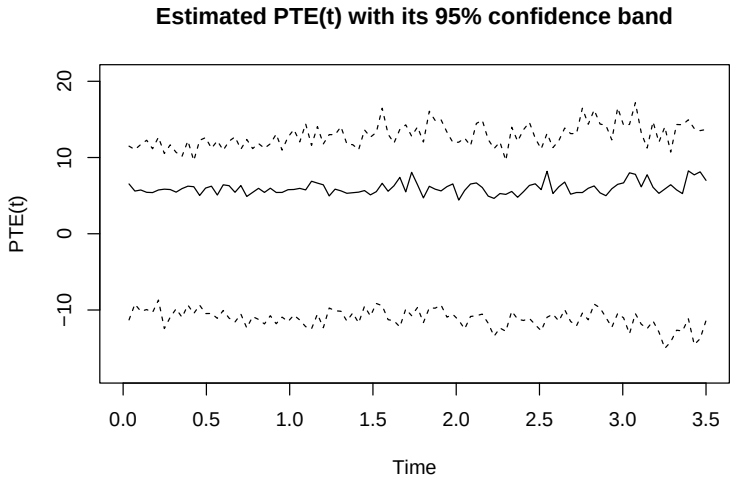


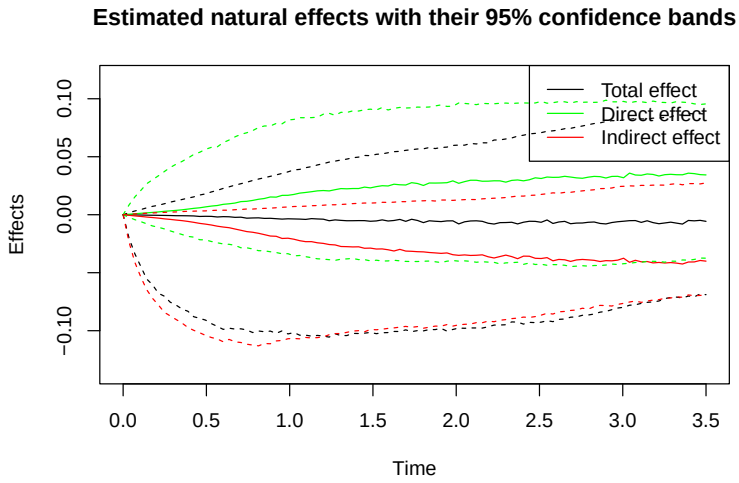
Note that we could have used `type="Survival"` to plot the estimated baseline survival function associated with $\hat{\lambda}_{0,T}(t)$.

Regarding the mediation measures, we can see that neither the direct effect, NDE(t), indirect effect NIE(t) nor the total effect, TTE(t) are significant, which results in a very large confidence interval for the PTE(t).

This can be better seen by plotting the different estimated functions by instead using the argument `type="PTE"`.

```
plot(mod.col,type="PTE")
```





As we can see, the confidence bands for the estimated PTE(t) are very wide (from -10 to 10). This is explained by the absence of treatment effect on the final endpoint; therefore TTE$(t) \approx 0$ which results in very high uncertainty regarding PTE(t). As such, it is impossible to conclude anything, since the absence of treatment effect on the final endpoint renders the search for surrogate endpoints pointless. The validation of surrogates endpoint should only be made for treatments that have proved their efficacy regarding the final endpoint. In this application the low number of patients (150) is a clear limitation.

5 Discussion

In this paper we presented new implementations of surrogate endpoint validation based on joint models that combines meta-analytic data and mediation analysis. These joint models have been implemented in the R package `frailtypack`. These models combine the two main approaches that have been developed in the statistical literature for surrogate endpoint validation: the meta-analytic approach and the causal mediation analysis approach.

To achieve a shorter computation time, the core of these functions were developed in `Fortran` and support multithreading through `OpenMP`. This parallelization takes place at the numerical approximation of the integral over the random-effects where the iterations (either in a Monte Carlo or quadrature approach) are dispatched between the available threads.

Numerical issues can happen with the estimation of these complex models, especially when computing the likelihood which requires computing integrals over the random effects (both individual and trial level). These integrals do not have closed-form solutions and therefore need to be numerically approximated. Two main approaches for carrying out these approximations can be used, either by using Monte Carlo sampling or Gauss-Hermite quadrature rule. Based on simulations results, it appears that

better numerical approximation and convergence of the maximization algorithm is achieved when a Monte-Carlo approach is used to integrate over the trial-level random effects and a Gauss-Hermite rule is used for the integral over the individual-level random effects. The latter can be further improved by using pseudo-adaptive quadrature in which the nodes in the quadrature are chosen to properly recover the scale and shape of the integrand (Rizopoulos, 2012a).

Another computational issue concerns specifically Model (2). The likelihood of this model requires computing the product of the individual-likelihoods $\prod_{j=1}^{n_i} L_{ij}$ for all the patients in trial i. Numerical instabilities can occur if n_i is large and/or if the number of observations per subject is large. If all individual likelihoods are small, say $L_{ij} \approx 10^{-3}$, and if n_i is large, $n_i \approx 200$ as can be the case for clinical trials, then the product will result in a numerical underflow and rounded to 0. One possible way around is to use a large constant, M, and multiply each individual likelihood by M so that there is no longer a numerical underflow. Since M is a constant it can then be easily taken out of the integral and substracted (when taking the log of this integral). However, in an iterative maximization algorithm, M may be suited for the first iterations but as the vector of parameter is updated, so will be the "scale" of the individual-likelihoods L_{ij} which depend on the parameters. Therefore ML_{ij} may be too large at a given iteration which can also result in numerical issues (numerical overflow this time). One possibility is to also update to value of M at each iteration so that it keeps track of the scale of the individual likelihoods. For now, only a fixed value M can be used (through the argument `Cmult` in the `longiPenal` function). The constant M used in practice in the likelihood is 10^{Cmult}.

Regarding the individual surrogacy measure (the Kendall's τ), the introduction of a direct effect of the surrogate on the final endpoint creates a difficulty regarding both its definition and computation since this link influences the association between the two endpoints and therefore must be taken into account. For example, in Model (1), the Kendall's without a direct link between S and T (i.e. without the function $\gamma(S)$), is given by Equation (7) which only requires integrating over the distribution of the random effects while the more general τ in the presence of $\gamma(S)$ is given by Equation (6) and requires more complex integration over both the distribution of S and the distribution of the random effects. For the trial-level association measure, R^2_{trial}, this issue does not occur since the trial-level random effects take into account all the association between the treatment effects on the surrogate and the final endpoint.

For now the `jointSurroPenal` does not allow adjustment on covariates beside the treatment itself, and the surrogacy measures (especially the Kendall's τ and the PTE(t)) are defined and computed in a somewhat "covariate-free" manner. Further developments will allow the adjustment on potential covariates in the model and the computation of these surrogacy measures conditionally to a given set of covariates as well as at a marginal (populational) level (Emura et al., 2021). Adjustment on covariates is already possible in the `longiPenal` function. However it should be noted that computing a marginal measure averaged over the distribution of the covariates can significantly increase the computation time

in the presence of continuous variables. Indeed, marginalization in the case of categorical covariates only requires to compute a τ or PTE(t) for each possible group and then averaging them (and weighting by the proportion of each group in the dataset). On the other hand, marginalization in the presence of a continuous covariates would require computing these measures for each observed values of the covariate in the dataset (i.e. for each individual) which can significantly increase the computation time.

Contrary to the meta-analytic approach, the mediation analysis does not require, *per se*, the use of meta-analytic data. The combination of meta-analytic data and mediation analysis allows broader generalisation of the results by introducing heterogeneity in the data. If one only has access to single-trial data, a possibility is to use different clustering units instead of "trials", such as "institution" in multicentric studies. Even if our proposed approaches were developed within a meta-analytic setting which is more appropriate for surrogate validation, a further development would be to run a mediation analysis without the requirement of meta-analytic or clustered data (which would yield a simpler model). However, in that case only the individual-level association in addition to the proportion of treatment effect would be available since the trial-level association would not be identifiable.

We want to point out that in a perspective of surrogate validation, there is no acknowledged threshold yet for the surrogacy measures, especially the proportion of treatment effect, above which a surrogate would be claimed validated. Such threshold warrants more investigation in order to give a clear decision rule for the clinicians and statisticians.

Finally, further developments of the `frailtypack` package will concern the extension of the proposed functions to validate surrogate endpoints. In order to improve the flexibility of the proposed approaches, other options for the type of surrogate or final endpoint will be proposed, for example in the case of a binary final endpoint.

References

Alonso, A. et al. (2016). *Applied surrogate endpoint evaluation methods with SAS and R*. CRC Press New York.

Burzykowski, T. and M. Buyse (2006). Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* **5**, 173–186.

Burzykowski, T., M. Buyse, and G. Molenberghs (2005). *The evaluation of surrogate endpoints*. Springer.

Buyse, M. et al. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* **1**, 49–67.

Daniels, M. J. and M. D. Hughes (1997). Meta-analysis for the evaluation of potential surrogate markers. *Statistics in medicine* **16**, 1965–1982.

Ducreux, M. et al. (2011). Sequential versus combination chemotherapy for the treatment of advanced colorectal cancer (FFCD 2000–05): an open-label, randomised, phase 3 trial. *The lancet oncology* **12**, 1032–1044.

Emura, T., C. L. Sofeu, and V. Rondeau (2021). Conditional copula models for correlated survival endpoints: Individual patient data meta-analysis of randomized controlled trials. *Statistical Methods in Medical Research* **30**, 2634–2650.

Frangakis, C. E. and D. B. Rubin (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

Freedman, L. S., B. I. Graubard, and A. Schatzkin (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in medicine* **11**, 167–178.

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology (Cambridge, Mass.)* **21**, 13.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association* **81**, 945–960.

Imai, K., L. Keele, and T. Yamamoto (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science* **25**, 51–71.

Joffe, M. M. and T. Greene (2009). Related causal frameworks for surrogate outcomes. *Biometrics* **65**, 530–538.

Joly, P., D. Commenges, and L. Letenneur (1998). A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics*, 185–194.

Le Coënt, Q., C. Legrand, and V. Rondeau (Nov. 2022). Time-to-event surrogate endpoint validation using mediation analysis and meta-analytic data. *Biostatistics*. kxac044.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics* **11**, 431–441.

Ovarian Cancer Meta-Analysis Project (1991). Cyclophosphamide plus cisplatin versus cyclophosphamide, doxorubicin, and cisplatin chemotherapy of ovarian carcinoma: a meta-analysis. *J Clin Oncol* **9**, 1668–74.

Paoletti, X. et al. (2010). Benefit of adjuvant chemotherapy for resectable gastric cancer: a meta-analysis. *Jama* **303**, 1729–1737.

Parast, L. (2021a). *Rsurrogate: Robust Estimation of the Proportion of Treatment Effect Explained by Surrogate Marker Information*. R package version 3.1.

VI

Discussion

VI.1 Joint modeling and mediation analysis 177

VI.2 Perspectives 179

VI.3 Conclusion 180

This thesis proposed developments for surrogate endpoint validation in clinical research based on the combinaition of mediation analysis, joint modeling and meta-analytic data. The rationale behind this work stands in the complementary nature of the meta-analytic approach and the mediation analysis approach for surrogate validation, yet both approaches are often developed separately.

In this thesis we proposed developments for the statistical validation of surrogate endpoints, relying on causal mediation analysis to determine the proportion of treatment effect on the final endpoint operating through the surrogate endpoint and also on the use of data from meta-analysis. Two joint models were developed depending on the nature (event time or longitudinal biomarker) of the surrogate endpoint from which measures related to the mediation analysis were defined. In this context, contrary to previous developments of mediation analysis which were limited to analyses on data from a single clinical trial, the contribution of data from meta-analyses allows a greater generalizability of the results obtained.

VI.1. | Joint modeling and mediation analysis

The starting point of this thesis was the joint model proposed by (Sofeu et al., 2019), which aimed at validating a time-to-event endpoint as a surrogate of another time-to-event endpoint through a meta-analytic approach. This model was proposed for a semi-competing risk setting as the final endpoint occurrence precludes the occurrence of the surrogate. From a mediation

analysis however this model does not allow a potential mediated effect since it assumes a conditional independence of the two endpoints given the random effects. In order to enable a mediation, we had to introduce a "direct" link between the two endpoints. We chose to define this link as a modification of the risk of occurrence of the final endpoint T at the occurrence of the surrogate endpoint S. This modification is taken into account through a function $\gamma(S)$. In practice this function is unknown and is estimated using cubic B-splines. The motivation to the use of a function that depends on the surrogate occurrence time is that this time may have an influence on the risk of occurrence of T. Another (simpler) possibility would have been to use a simple fixed effect, $\gamma(S) = \gamma$, in which case it is only the occurrence of S that influences the risk of T: the time at which S occurs has therefore no influence. It is more restrictive as in some settings, it may happens that an early occurrence of S may have a greater impact than a late occurrence. For example, a cancer progression at 6 months might have a greater (or lesser) impact on the risk of death than a progression at 5 years. Therefore we chose to use a combination of both the occurrence of S (through the indicator $I_{S\leq t}$ in Model II.19) and the time of occurrence of S, $\gamma(S)$. While we introduced a direct effect of S on T this way, other possibilites in this setting of semi-competing risks are possible. WEIR et al., 2022 used the framework of multi-state modeling to give a definition of the indirect effect of the treatment on the final endpoint through the surrogate while GAO et al., 2020 defined an indirect effect using principal stratification, i.e. within subgroups of patients having the same surrogate outcomes under both treatment arm.

The analog of the function $\gamma(S)$ in Model (II.6) is the link function $h(\cdot)$ which represents the direct link between the longitudinal surrogate and the final endpoint. In this thesis we focused mainly on two link functions, the "current-level" and "current-slope". The choice of link function can greatly impact the results in terms of mediation and surrogacy, as illustration in the application on the PSA in Chapter IV. Further work might is needed to investigate other link functions and study how they may influence the mediation analysis and the proportion of treatment effect.

It should be noted that the total treatment effect on the final endpoint can be estimated in two ways. Since in the developments we proposed the final endpoint is a time-to-event, the total treatment effect on the survival scale can be obtained by simply comparing Kaplan-Meier estimators between the two arms. On the other hand, it can be computed from the joint

models, in the same way as the indirect and direct effect. Since the proportion of treatment effect relies on both the indirect and total treatment effect, in order to remain consistent we chose to also compute the total treatment effect through the joint models.

VI.2. | Perspectives

One limitation of the current implementation of Model (II.19) is that it does not allow for adjustment on covariates. While we could argue that the primary interest lies in surrogacy validation rather than estimating the effects of covariates, allowing such adjustment has at least two benefits. First, from a mediation viewpoint we require that there is no unmeasured confounder. Even if the random effects may help capture potential confounding, it is not perfect as the random effects are assumed to have a very specific Gaussian distribution. For confounders that have a distribution far from being Gaussian (or just far from being symmetric), the random effects may fail to take properly into account this confounding. Moreover, these random are time-fixed therefore may not be appropritate for time-dependent confounding. Secondly, it may be interesting to compute the different surrogacy measures conditional on some covariates to investigate if the surrogacy behaves differently accross groups of patients (Emura et al., 2021). Therefore, extension of the current implementation of Model (II.19) to allow adjustment on covariates is attractive.

In this thesis we focused on the case where only a single surrogate was studied but developments using the combination of several surrogates to derive the proportion of treatment effect were recently proposed (Wang, Parast, Han, et al., 2022). The combination of several surrogates is attractive to increase the degree of surrogacy compared to the use of a single surrogate. In this regards, one possible extention of Model (II.19) would be to study repeated surrogates event times instead of a single event (Rondeau et al., 2007). For Model II.6 this would be to use several longitudinal biomarkers combined. While this may seem burdensome from a computational point of view, methods have been proposed in the joint modeling framework to incorporate several biomarkers (Hickey et al., 2018). Moreover, in this work we studied either a time-to-event or a longitudinal biomarker as a surrogate of a final time-to-event endpoint. A possible extension would be to use a combination of both a time-to-event and a longitudinal biomarker as surrogate endpoints. While this may strengthens the surrogacy, from a mediation perspective this may complicates the causal definition and interpretation of the resulting

indirect and direct effects.

The proposed approaches could require significant computation time. Most of this computation time comes from the iterative process of the maximization of the joint model. Long computation times are not unusual in joint modeling and several methods have been proposed to overcome this. For example, in the bayesian framework R-INLA can be used to estimate joint models more quickly than most frequentits counterparts (RUSTAND et al., 2020).

One limitation of the proportion of treatment effect as a surrogacy measure, is the difficulty in estimation if the total treatment effect on the final endpoint is small. In that case the proprotion of treatment effect can be highly variable and its confidence interval very broad. It may be interesting to investigate other potential measures of surrogacy derived from the mediation analysis that do not relate directly to the proportion of the treatment effect.

Other perspectives emerge from this thesis. First, the extension of the proposed methods to cases where the final endpoint does not necessarily correspond to a time-to-event time is envisaged. For example, binary endpoints (success/failure) are also widely used in clinical reasearch, as well as continuous endpoints. Extentions of the proposed methods to these different types of endpoints will require the definition of appropriate joint models, but the overall spirit of the validation would remain essentially the same; the proportion of the treatment effect on the endpoint will be determined from these new models.

VI.3. | CONCLUSION

The work of this thesis has allowed the development of new joint models and new approaches for the validation of substitution criteria in the case where the final endpoint is an event time, based on joint modeling, mediation analysis and meta-analyses. These approaches have been implemented in a statistical package to facilitate their diffusion and use. The illustrations of the proposed methods on real datasets were all in oncology. However, these methods can be applied to other areas of clinical research in which measures of treatment efficacy correspond to time to occurrence of a clinical event of interest.

Bibliography

Avin, C., I. Shpitser, and J. Pearl (2005). Identifiability of Path Specific Effects. *In Proceedings of International Joint Conference on Artificial Intelligence*, 357–363.

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology & therapeutics* **69**, 89–95.

Burzykowski, T. and M. Buyse (2006). Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* **5**, 173–186.

Burzykowski, T., M. Buyse, and G. Molenberghs (2005). *The evaluation of surrogate endpoints*. Springer.

Burzykowski, T., G. Molenberghs, et al. (2001). Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **50**, 405–422.

Buyse, M., T. Burzykowski, K. Carroll, et al. (2007). Progression-free survival is a surrogate for survival in advanced colorectal cancer. *Journal of clinical oncology* **25**, 5218–5224.

Buyse, M., T. Burzykowski, S. Michiels, et al. (2008). Individual-and trial-level surrogacy in colorectal cancer. *Statistical methods in medical research* **17**, 467–475.

Buyse, M. and G. Molenberghs (1998). Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* **54**, 1014–1029.

Buyse, M., G. Molenberghs, T. Burzykowski, et al. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* **1**, 49–67.

Buyse, M., G. Molenberghs, X. Paoletti, et al. (2016). Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. *Biometrical Journal* **58**, 104–132.

CANCER RESEARCH UK (2020). *Surrogate Endpoint Resources for Drug and Biologic Development.* https://www.cancerresearchuk.org/about-cancer/breast-cancer/survival.

CARPENTER, D., A. S. KESSELHEIM, and S. JOFFE (2011). Reputation and Precedent in the Bevacizumab Decision. *New England Journal of Medicine* **365**, e3.

DANIELS, M. J. and M. D. HUGHES (1997). Meta-analysis for the evaluation of potential surrogate markers. *Statistics in medicine* **16**, 1965–1982.

DELGADO, A. and A. K. GUDDATI (2021). Clinical endpoints in oncology-a primer. *American journal of cancer research* **11**, 1121.

DEMPSTER, A. P., N. M. LAIRD, and D. B. RUBIN (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**, 1–22.

EMURA, T., C. L. SOFEU, and V. RONDEAU (2021). Conditional copula models for correlated survival endpoints: Individual patient data meta-analysis of randomized controlled trials. *Statistical Methods in Medical Research* **30**, 2634–2650.

FRANGAKIS, C. E. and D. B. RUBIN (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

FREEDMAN, L. S., B. I. GRAUBARD, and A. SCHATZKIN (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in medicine* **11**, 167–178.

GAO, F., F. XIA, and K. C. G. CHAN (2020). Defining and Estimating Subgroup Mediation Effects with Semi-Competing Risks Data. *arXiv preprint arXiv:2010.00061*.

GHOSH, D., J. M. TAYLOR, and D. J. SARGENT (2012). Meta-analysis for surrogacy: Accelerated failure time models and semicompeting risks modeling. *Biometrics* **68**, 226–232.

GIESSEN, C. et al. (2013). Progression-Free Survival as a Surrogate Endpoint for Median Overall Survival in Metastatic Colorectal Cancer: Literature-Based Analysis from 50 Randomized First-Line TrialsPFS as a Surrogate Endpoint in mCRC. *Clinical Cancer Research* **19**, 225–235.

Global Cancer Observatory (2022). *All cancers fact sheet*. https://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf.

Hernan, M. and J. Robins (2020). *Causal Inference: What if*. Chapman & Hill/CRC.

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology (Cambridge, Mass.)* **21**, 13.

Hickey, G. L. et al. (2018). joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC medical research methodology* **18**, 1–14.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association* **81**, 945–960.

Hughes, M. D. (1993). Regression dilution in the proportional hazards model. *Biometrics* **49**, 1056–1066.

Imai, K., L. Keele, and D. Tingley (2010). A general approach to causal mediation analysis. *Psychological methods* **15**, 309–334.

Imai, K., L. Keele, and T. Yamamoto (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science* **25**, 51–71.

Joffe, M. M. and T. Greene (2009). Related causal frameworks for surrogate outcomes. *Biometrics* **65**, 530–538.

Joly, P., D. Commenges, and L. Letenneur (1998). A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics* **54**, 185–194.

Kalbfleisch, J. D. and R. L. Prentice (2011). *The statistical analysis of failure time data*. John Wiley & Sons.

Kemp, R. and V. Prasad (2017). Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC medicine* **15**, 1–7.

Kim, C. and V. Prasad (2015). Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: an analysis of 5 years of US Food and Drug Administration approvals. *JAMA Internal Medicine* **175**, 1992–1994.

Lange, T. et al. (2017). Applied mediation analyses: a review and tutorial. *Epidemiology and health* **39**.

LAURIE, D. (1997). Calculation of Gauss-Kronrod quadrature rules. *Mathematics of Computation* **66**, 1133–1145.

LE COËNT, Q., C. LEGRAND, and V. RONDEAU (Nov. 2022). Time-to-event surrogate endpoint validation using mediation analysis and meta-analytic data. *Biostatistics*. kxac044.

MARQUARDT, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics* **11**, 431–441.

MAUGUEN, A. et al. (2013). Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: a re-analysis of meta-analyses of individual patients' data. *The lancet oncology* **14**, 619–626.

MICHIELS, S. et al. (2009). Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data. *The lancet oncology* **10**, 341–350.

MILES, D. W. et al. (2010). Phase III study of bevacizumab plus docetaxel compared with placebo plus docetaxel for the first-line treatment of human epidermal growth factor receptor 2–negative metastatic breast cancer. *Journal of clinical oncology* **28**, 3239–3247.

MILLER, K. et al. (2007). Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *New England journal of medicine* **357**, 2666–2676.

PARAST, L., T. CAI, and L. TIAN (2017). Evaluating surrogate marker information using censored data. *Statistics in medicine* **36**, 1767–1782.

PARAST, L., T. CAI, and L. TIAN (2022). Using a Surrogate with Heterogeneous Utility to Test for a Treatment Effect. *arXiv preprint arXiv:2209.08315*.

PEARL, J. (2022). Direct and Indirect Effects. *Probabilistic and Causal Inference: The Works of Judea Pearl*. Association for Computing Machinery.

PETERSEN, M. L., S. E. SINISI, and M. J. van der LAAN (2006). Estimation of direct causal effects. *Epidemiology* **17**, 276–284.

PRENTICE, R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine* **8**, 431–440.

RENARD, D. et al. (2003). Validation of a longitudinally measured surrogate marker for a time-to-event endpoint. *Journal of Applied Statistics* **30**, 235–247.

Renfro, L. A. et al. (2012). Bayesian adjusted R2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials. *Statistics in Medicine* **31**, 743–761.

Rizopoulos, D. (2012a). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis* **56**, 491–501.

Rizopoulos, D. (2012b). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.

Robert, N. J. et al. (2011). RIBBON-1: randomized, double-blind, placebo-controlled, phase III trial of chemotherapy with or without bevacizumab for first-line treatment of human epidermal growth factor receptor 2–negative, locally recurrent or metastatic breast cancer. *Journal of clinical oncology* **29**, 1252–1260.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling* **7**, 1393–1512.

Robins, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. *Oxford Statistical Science Series*, 70–82.

Robins, J. M. and S. Greenland (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143–155.

Rondeau, V. et al. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* **8**, 708–721.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society: Series A (General)* **147**, 656–666.

Rotolo, F. et al. (2019). A Poisson approach to the validation of failure time surrogate endpoints in individual patient data meta-analyses. *Statistical Methods in Medical Research* **28**, 170–183.

Royston, P. and M. K. Parmar (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC medical research methodology* **13**, 1–15.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* **66**, 688.

RUBIN, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American statistical association* **75**, 591–593.

RUSTAND, D. et al. (2020). *Bayesian Estimation of Two-Part Joint Models for a Longitudinal Semi-continuous Biomarker and a Terminal Event with R-INLA: Interests for Cancer Clinical Trial Evaluation*.

SARGENT, D. J. et al. (2005). Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *Journal of Clinical Oncology* **23**, 8664–8670.

SCHOENFELD, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics* **39**, 499–503.

SELF, S. and Y. PAWITAN (1992). Modeling a marker of disease progression and onset of disease. *AIDS epidemiology*. Springer.

SHI, Q. et al. (2011). Comparative assessment of trial-level surrogacy measures for candidate time-to-event surrogate endpoints in clinical trials. *Computational Statistics & Data Analysis* **55**, 2748–2757.

SOFEU, C. L., T. EMURA, and V. RONDEAU (2019). One-step validation method for surrogate endpoints using data from multiple randomized cancer clinical trials with failure-time endpoints. *Statistics in medicine* **38**, 2928–2942.

SRIDHARA, R. et al. (2010). Review of oncology and hematology drug product approvals at the US Food and Drug Administration between July 2005 and December 2007. *Journal of the National Cancer Institute* **102**, 230–243.

SWEETING, M. J. and S. G. THOMPSON (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal* **53**, 750–763.

TSIATIS, A. A. and M. DAVIDIAN (2004). Joint modeling of longitudinal and time-dependent data: an overview. *Statistica Sinica* **14**, 809–834.

U.S. Food and Drug Administration (2018). *Surrogate Endpoint Resources for Drug and Biologic Development*. https://www.fda.gov/drugs/development-resources/surrogate-endpoint-resources-drug-and-biologic-development.

U.S. National Library of Medicine (2022). *Trends, Charts, and Maps*. https://www.clinicaltrials.gov/ct2/resources/trends.

Vandenberghe, S. et al. (2018). Surrogate marker analysis in cancer clinical trials through time-to-event mediation techniques. *Statistical methods in medical research* **27**, 3367–3385.

VanderWeele, T. J. (2011a). Causal mediation analysis with survival data. *Epidemiology (Cambridge, Mass.)* **22**, 582.

VanderWeele, T. J. (2011b). Controlled direct and mediated effects: definition, identification and bounds. *Scandinavian Journal of Statistics* **38**, 551–563.

VanderWeele, T. J. and M. A. Hernan (2013). Causal inference under multiple versions of treatment. *Journal of causal inference* **1**, 1–20.

Wang, X., L. Parast, L. Han, et al. (2022). Robust approach to combining multiple markers to improve surrogacy. *Biometrics*.

Wang, X., L. Parast, L. Tian, et al. (2022). Towards Optimal Use of Surrogate Markers to Improve Power. *arXiv preprint arXiv:2209.08414*.

Weir, I. R., J. R. Rider, and L. Trinquart (2022). Counterfactual mediation analysis in the multistate model framework for surrogate and clinical time-to-event outcomes in randomized controlled trials. *Pharmaceutical Statistics* **21**, 163–175.

Worl Health Organization (2022). *Cancer*. https://www.who.int/news-room/fact-sheets/detail/cancer.

Wulfsohn, M. S. and A. A. Tsiatis (1997). A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics* **53**, 330–339.

Zheng, C. and L. Liu (2022). Quantifying direct and indirect effect for longitudinal mediator and survival outcome using joint modeling approach. *Biometrics* **78**, 1233–1243.

Appendices

A

Proofs

A.1. | Mediation analysis formula

$$
\begin{aligned}
\mathbb{E}\left[T(z, S(z')) \mid X\right] &= \int_{\mathcal{S}} \mathbb{E}\left[T(z, S(z')) \mid S(z') = s, X\right] f_{S(z')}(s)\mathrm{d}s \\
&= \int_{\mathcal{S}} \mathbb{E}\left[T(z, S(z')) \mid Z = z', S(z') = s, X\right] f_{S(z')|X}(s)\mathrm{d}s \\
&= \int_{\mathcal{S}} \mathbb{E}\left[T(z, S(z')) \mid Z = z', X\right] f_{S(z')|X}(s)\mathrm{d}s \\
&= \int_{\mathcal{S}} \mathbb{E}\left[T(z, S(z')) \mid Z = z, X\right] f_{S(z')|Z=z',X}(s)\mathrm{d}s \\
&= \int_{\mathcal{S}} \mathbb{E}\left[T(z, S(z')) \mid S(z) = s, Z = z, X\right] f_{S(z')|Z=z',X}(s)\mathrm{d}s \\
&= \int_{\mathcal{S}} \mathbb{E}\left[T \mid S = s, Z = z, X\right] f_{S(z')|Z=z',X}(s)\mathrm{d}s \\
&= \int_{\mathcal{S}} \mathbb{E}\left[T \mid S = s, Z = z, X\right] f_{S|Z=z',X}(s)\mathrm{d}s.
\end{aligned}
\tag{A.1}
$$

The first equality simply comes from averaging $\mathbb{E}\left[T(z, S(z')) \mid S(z') = s, X\right]$ over the distribution of $S(z')$. The second equality is an implication of the first part of the SI: if

$$\{T(z, s), S(z')\} \perp\!\!\!\perp Z \mid X,$$

then,

$$T(z, S(z')) \perp\!\!\!\perp Z \mid S(z') = s, X.$$

The third equality comes from the second part of the SI, the fourth from the first part of the SI, the fifth from the second part of the SI. Finally the sixth and seventh equalities come from the consitency assumption.

A.2. | Kendall's τ

Our purpose is to compute the Kendall's τ associated with the following joint frailty model:

$$
\begin{cases}
\lambda_S\left(t \mid \omega, u\right) = \lambda_{0,S}(t) \exp\left(\omega + u\right) \\
\lambda_T\left(t \mid \omega, u\right) = \lambda_{0,T}(t) \exp\left(\zeta\omega + \alpha u\right).
\end{cases}
\tag{A.2}
$$

In this model (ω, u) is a bivariate gaussian vector

$$\begin{pmatrix} \omega \\ u \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \theta^2 & 0 \\ 0 & \rho^2 \end{pmatrix} \right).$$

The Kendall's τ is defined as:

$$\tau = \mathbb{P}\left((S_1 - S_2)(T_1 - T_2) > 0\right) - \mathbb{P}\left((S_1 - S_2)(T_1 - T_2) < 0\right).$$

for two individuals 1 and 2 draw at random. We have,

$$\begin{aligned} \mathbb{P}\left((S_1 - S_2)(T_1 - T_2) > 0\right) - \mathbb{P}\left((S_1 - S_2)(T_1 - T_2) < 0\right) &= 2\mathbb{P}\left((S_1 - S_2)(T_1 - T_2) > 0\right) - 1 \\ &= 2\Big\{\mathbb{P}\left(S_1 > S_2, T_1 > T_2\right) \\ &\quad + \mathbb{P}\left(S_1 < S_2, T_1 < T_2\right)\Big\} - 1 \\ &= 4\mathbb{P}\left(S_1 > S_2, T_1 > T_2\right) - 1. \end{aligned}$$

We are interested in the computation of $\mathbb{P}\left(S_1 > S_2, T_1 > T_2\right)$. Averaging over the random effects gives,

$$\mathbb{P}\left(S_1 > S_2, T_1 > T_2\right) = \mathbb{E}_{(\omega_1, u_1, \omega_2, u_2)}\left[\mathbb{P}\left(S_1 > S_2, T_1 > T_2 \mid \omega_1, u_1, \omega_2, u_2\right)\right],$$

with $(\omega_1, u_1) \perp\!\!\!\perp (\omega_2, u_2)$. By assumption of the model (A.2), T and S are independent given (ω, u) therefore,

$$\mathbb{P}\left(S_1 > S_2, T_1 > T_2 \mid \omega_1, u_1, \omega_2, u_2\right) = \mathbb{P}\left(S_1 > S_2 \mid \omega_1, u_1, \omega_2, u_2\right) \mathbb{P}\left(T_1 > T_2 \mid \omega_1, u_1, \omega_2, u_2\right).$$

First we have,

$$\begin{aligned} \mathbb{P}\left(S_1 > S_2 \mid \omega_1, u_1, \omega_2, u_2\right) &= \int_0^\infty \mathbb{P}\left(S_1 > s \mid \omega_1, u_1, \omega_2, u_2\right) f_{S_2}\left(s \mid \omega_1, u_1, \omega_2, u_2\right) \mathrm{d}s \\ &= \int_0^\infty \exp\left(-H_{0,S}(s) e^{\omega_1 + u_1}\right) \lambda_{0,S}(s) e^{\omega_2 + u_2} \exp\left(-H_{0,S}(s) e^{\omega_2 + u_2}\right) \mathrm{d}s. \end{aligned}$$

Let $W_i = \exp(\omega_i + u_i)$ we have:

$$\begin{aligned}
\mathbb{P}\left(S_1 > S_2 \mid \omega_1, u_1, \omega_2, u_2\right) &= \int_0^\infty \exp\left(-H_{0,S}(s)W_1\right) \lambda_{0,S}(s) W_2 \exp\left(-H_{0,S}(s)W_2\right) \mathrm{d}s \\
&= W_2 \int_0^\infty \lambda_{0,S} \exp\left(-H_{0,S}(s)\left(W_1 + W_2\right)\right) \mathrm{d}s \\
&= \frac{W_2}{W_1 + W_2} \underbrace{\int_0^\infty \left(W_1 + W_2\right) \lambda_{0,S} \exp\left(-H_{0,S}(s)\left(W_1 + W_2\right)\right) \mathrm{d}s}_{=1} \\
&= \frac{W_2}{W_1 + W_2} \\
&= \frac{e^{\omega_2 + u_2}}{e^{\omega_1 + u_1} + e^{\omega_2 + u_2}}.
\end{aligned}$$

The above integral is equal to 1 since it is of the form $\int \lambda(t) \exp\left(-\int_0^t \lambda(u)\mathrm{d}u\right) \mathrm{d}t$ with $\lambda(t)$ a hazard function which result in $\lambda(t) \exp\left(-\int_0^t \lambda(u)\mathrm{d}u\right)$ being a density. The same process applies to compute $\mathbb{P}\left(T_1 > T_2 \mid \omega_1, u_1, \omega_2, u_2\right)$ which yields

$$\mathbb{P}\left(T_1 > T_2 \mid \omega_1, u_1, \omega_2, u_2\right) = \frac{e^{\zeta\omega_2 + \alpha u_2}}{e^{\zeta\omega_1 + \alpha u_1} + e^{\zeta\omega_2 + \alpha u_2}}.$$

And finally,

$$\begin{aligned}
\mathbb{P}\left(S_1 > S_2, T_1 > T_2\right) &= \mathbb{E}_{(\omega_1, u_1, \omega_2, u_2)} \left[\frac{e^{\omega_2 + u_2}}{e^{\omega_1 + u_1} + e^{\omega_2 + u_2}} \frac{e^{\zeta\omega_2 + \alpha u_2}}{e^{\zeta\omega_1 + \alpha u_1} + e^{\zeta\omega_2 + \alpha u_2}} \right] \\
&= \mathbb{E}_{(\omega_1, u_1, \omega_2, u_2)} \left[\frac{e^{\omega_2 + u_2} + e^{\zeta\omega_2 + \alpha u_2}}{\left(e^{\omega_1 + u_1} + e^{\omega_2 + u_2}\right)\left(e^{\zeta\omega_1 + \alpha u_1} + e^{\zeta\omega_2 + \alpha u_2}\right)} \right] \\
&= \int_{\mathbb{R}^4} \left[\frac{e^{\omega_2 + u_2} + e^{\zeta\omega_2 + \alpha u_2}}{\left(e^{\omega_1 + u_1} + e^{\omega_2 + u_2}\right)\left(e^{\zeta\omega_1 + \alpha u_1} + e^{\zeta\omega_2 + \alpha u_2}\right)} \right] \\
&\qquad \frac{1}{4\pi^2\theta\rho} \exp\left(-\frac{1}{2}\left(\frac{\omega_1 + \omega_2}{\theta^2} + \frac{u_1 + u_2}{\rho^2}\right)\right) \mathrm{d}\omega_1 \mathrm{d}\omega_1 \mathrm{d}u_1 \mathrm{d}u_2.
\end{aligned}$$

Développement de méthodes pour la validation de critères de substitution : comparaison des approches causales et non causales pour des données de méta-analyses

Résumé : En recherche clinique l'utilisation de critères de substitution permet d'accélérer le processus d'évaluation d'un traitement par rapport à l'utilisation d'un critère de jugement classique. Les critères de substitution doivent au préalable avoir été validés statistiquement avant toute utilisation dans un nouvel essai clinique. Dans cette thèse nous proposons la validation d'un critère de substitution basée sur une approche causale utilisant l'analyse de médiation qui vise à décomposer l'effet du traitement sur le critère de jugement final en un effet indirect via le critère de substitution et un effet direct indépendant de celui-ci. L'effet total du traitement est donc la somme de l'effet indirect et de l'effet direct. Un critère de substitution sera validé si l'essentiel de l'effet du traitement correspond à l'effet indirect (le traitement opère donc principalement via le critère de substitution). Nous nous intéressons particulièrement au cas où le critère de jugement est un temps d'évènement comme le temps avant décès et à un critère de substitution étant soit également un temps d'évènement ou un biomarqueur longitudinal. Des modèles conjoints ont été développés desquels les effets direct et indirect du traitement peuvent être obtenus. Ces modèles permettent également de prendre en compte des données hétérogènes issues de méta-analyses ou d'études multi-centriques affin de renforcer le processus de validation. Des études de simulations ont été conduites pour evaluer la performance de ces approches qui ont ensuite été appliquées à des jeux de données réelles en cancérologie : une méta-analyse en cancer gastrique pour évaluer la survie sans maladie comme critère de substitution de la survie globale et une étude multi-centrique en cancer de la prostate pour évaluer l'évolution au cours du temps du taux d'antigène spécifique de la prostate comme critère de substitution de la survie sans maladie. Ces approches ont été implémentées dans un package statistique visant à faciliter leur utilisation.
Mots-clés : Critère de substitution, Modélisation conjointe, Analyse de médiation, Méta-analyses

---

Development of methods for surrogate endpoints validation: comparison of causal and non causal approaches for meta-analytic data.

Abstract: In clinical research, the use of surrogate endpoints speeds up the process of evaluating a treatment compared to the use of a final endpoint. Surrogate endpoints must first be statistically validated before being used in a new clinical trial. In this thesis we propose the validation of a surrogate endpoint based on a causal approach using mediation analysis which aims at decomposing the effect of the treatment on the final endpoint into an indirect effect through the surrogate endpoint and a direct effect independent of it. The total effect of the treatment is therefore the sum of the indirect effect and the direct effect. A surrogate endpoint will be validated if most of the treatment effect corresponds to the indirect effect (the treatment therefore operates mainly through the surrogate criterion). We are particularly interested in the case where the endpoint is a time-to-event such as time to death and a surrogate endpoint is either also a time-to-event or a longitudinal biomarker. Joint models have been developed from which direct and indirect treatment effects can be derived. These models also allow heterogeneous data from meta-analyses or multi-center studies to be taken into account in order to strengthen the validation process. Simulation studies were conducted to evaluate the performance of these approaches which were then applied to real cancer datasets: a meta-analysis in gastric cancer to evaluate disease-free survival as a surrogate for overall survival and a multi-center study in prostate cancer to evaluate the evolution over time of the level of prostate-specific antigen as a surrogate for disease-free survival. These approaches have been implemented in a statistical package to facilitate their use.
Keywords: Surrogate endpoint, Joint modeling, Mediation analysis, Meta-analyses

---

Bordeaux Population Health Research Center
INSERM U1219, Université de Bordeaux, 33000 Bordeaux, France.