



Validation de biomarqueurs prédictifs de la réponse au traitement : extension des courbes de prédictivités à un critère de jugement censuré

Bassirou Mboup

► To cite this version:

Bassirou Mboup. Validation de biomarqueurs prédictifs de la réponse au traitement : extension des courbes de prédictivités à un critère de jugement censuré. Statistiques [math.ST]. Université Paris-Saclay, 2020. Français. NNT : 2020UPASR011 . tel-03176621

HAL Id: tel-03176621

<https://theses.hal.science/tel-03176621>

Submitted on 22 Mar 2021

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

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

Validation de biomarqueurs prédictifs de la réponse au traitement : extension des courbes de prédictivités à un critère de jugement censuré

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 570, Santé publique (EDSP)

Spécialité de doctorat: Santé publique-biostatistiques

Unité de recherche: Institut Curie, PSL Research University, INSERM, U900,

F-92210, Saint Cloud

Référent: : Faculté de Médecine

Thèse présentée et soutenue en visioconférence totale, le 13

Novembre 2020, par

Bassirou Mboup

Composition du jury:

Agathe Guilloux

Professeure, Université d'Évry Val d'Essonne

Présidente

Olivier Bouaziz

Maître de conférence, Université Paris Descartes

Rapporteur et Examineur

Etienne Dantan

Maitre de conférence, Université de Nantes

Rapporteur et Examineur

Aurélien Latouche

Professeur, Conservatoire national des arts et métiers &
Institut Curie

Directeur

À mon défunt ami Fodé Ndiaye

Résumé en Français et en Anglais

Résumé en Français

Il est fréquent en oncologie, de vouloir déterminer si seulement un sous groupe de patients va bénéficier ou pas d'un traitement. C'est un des paradigmes de la médecine personnalisée ou stratifiée. Les biomarqueurs prédictifs sont souvent utilisés pour sélectionner ces patients et la plupart de ces biomarqueurs sont continus. Par exemple les signatures génomiques comme Oncotype-Dx. Il a été proposé dans la littérature une méthodologie d'évaluation d'un biomarqueur avec réponse binaire. L'objectif dans cette thèse est dans un premier travail d'étendre cette méthodologie avec un critère de jugement censuré et de déterminer à différents horizons de prédiction le seuil optimal du biomarqueur au-delà duquel le traitement sera attribué ou évité. Un modèle dont l'estimation de ces paramètres repose sur des pondérations par l'inverse des probabilités de censure est proposé permettant ainsi d'obtenir un estimateur consistant. Une extension des courbes de prédictivité par bras de traitement dépendantes du temps est ainsi proposée. Dans un second travail, un test de l'hypothèse de calibration avec des données censurées à droite a été proposé. Ce test est valide au-delà 60% de taux de censure contrairement à ceux qui existent déjà dans la littérature et nous permettra d'étudier l'influence d'une mauvaise calibration sur la détermination du seuil. Un troisième travail s'intéresse à la détermination du seuil d'un nouveau biomarqueur pronostique pour le cancer de l'ovaire afin de classer les patients en haut ou bas risque de rechute. Enfin, dans un quatrième travail, nous proposons d'employer la régression quantile pour données censurées pour quantifier le bénéfice à long terme de l'immunothérapie.

Mots-clés : Biomarqueurs, pronostique, prédictifs, calibration, cancer de l'ovaire, immunothérapie, régression quantile

Résumé en Anglais

It is common in oncology to want to determine whether or not only a subgroup of patients will benefit from a treatment. This is one of the paradigms of personalized or stratified

medicine. Predictive biomarkers are often used to select these patients and most of these biomarkers are continuous. For example genomic signatures such as Oncotype-Dx. A methodology for evaluating a biomarker with binary response has been proposed in the literature. The objective of this thesis is in first work to extend this methodology with right-censored data and to determine at different prediction horizons the optimal threshold of the biomarker beyond which treatment will be attributed or avoided. A model whose estimates of these parameters are based on inverse censored probability weights is proposed to provide consistent estimators. An extension of the predictiveness curves will be carried out. In a second work, a test of the calibration hypothesis with right-censored data has been proposed. This test will be valid beyond the 60% censoring rate contrary to those already existing in the literature and will allow us to study the influence of a bad calibration on the determination of the threshold. A third work focuses on the determination of the threshold of a new prognostic biomarker for ovarian cancer in order to classify patients at high or low risk of relapse. Finally, a fourth work consists in illustrating the relevance of the censored quantile regression for quantifying the long term benefit of immunotherapy in a reconstructed data set from a single randomized trial. The proposed methodology can be readily employed for individual patients data meta-analysis to summarize evidence of immunotherapy as quantified by the upper quantile of the survival distribution.

Keywords : Biomarkers, pronostic, predictive, calibration, Ovarian cancer, immunotherapy, quantile regression

Remerciements

Je tiens à exprimer mes sincères remerciements dans un premier temps à mon directeur de thèse, **Aurélien Latouche**, pour m’avoir accueilli au sein de son équipe Méthode statistiques pour la médecine de précision et pour m’avoir accordé toute sa confiance. Je le remercie également de m’avoir donné l’opportunité d’assister à plusieurs congrès nationaux et internationaux et de participer à différents projets de recherche et de collaborations nationales. Je lui également reconnaissant pour ses qualités scientifiques et sa rigueur. Pendant ces trois années de formation, j’ai vécu une expérience enrichissante autant au niveau professionnel qu’humain, j’ai beaucoup appris à ses côtés et je lui adresse toute ma gratitude et admiration. **Aurélien**, sans votre aide, ce travail n’aurait jamais vu le jour.

Mes remerciements vont obligatoirement à **Paul Blanche**. Une grande partie de ces travaux a été réalisée en étroite collaboration. Merci pour tous tes conseils, ton aide technique et ta gentillesse. Ta rigueur, ta capacité d’analyse des problèmes et tes connaissances m’ont permis de progresser et ont répondu à plusieurs de mes préoccupations.

Je tiens à exprimer mes remerciements les plus sincères aux rapporteurs de cette thèse, **Olivier Bouaziz** et **Etienne Dantan**, pour avoir accepté de rapporter ce travail de thèse, pour l’intérêt qu’ils y ont apporté et pour leurs conseils avisés.

Je remercie la présidente du jury, **Agathe Guilloux**, pour l’attention qu’elle a accordé à ce travail.

Je souhaite remercier vivement tous les membres de l’équipe STAMPM, pour leur sympathie et leur accueil spécialement **Roman Rouzier**, **Delphine Héquet**, **Pascal**, **Christophe Le Tourneau**, **Xavier Paoletti**. Nos réunions hebdomadaires sont de vrais moments d’échange et de partage de connaissances que j’apprécie particulièrement. Merci pour la bonne ambiance qui règne.

Je remercie particulièrement ma collègue **Alessandra Meddis** pour la bonne ambiance de travail mais également pour les nombreux bons moments passés ensemble.

Un remerciement particulier à **Oumou Goundiam** pour cette collaboration fructueuse. Nos échanges concernant les aspects biologiques ont été très enrichissants.

Je voudrais remercier vivement tous les membres de l'U900 pour leur sympathie et leur accueil spécialement le directeur **Emmanuel Barillot** et la secrétaire **Caroline Dahan Belliere**.

Je souhaite remercier **Babacar Wade** et son épouse Mame Fatou. Quand j'ai besoin d'un avis, tu es toujours là pour me conseiller et m'encourager Babs. Merci pour ton aide. Je remercie également **Ange Sagna** et son épouse Béné. Je n'oublie jamais tes conseils et tes débats passionnants à chaque fois qu'on se rencontre.

Je remercie **Mor absa Loum** pour ton aide et tes encouragements. Un voisin comme toi est très rare.

Je remercie tous les membres du groupe **Dor Kat yi** pour les matchs de football organisés les week-ends.

Je remercie **Lamine Sylla** mon binôme à qui je partageais les amphithéâtres de l'UGB. Je n'oublie jamais les nuits blanches dans les salles de l'UFR. Un grand merci à mes voisins de chambre du 4G2D spécialement à Zakaria Tall, Moustapha Fall, Mouhameth Seck, Alousseynou Ba, Mba sékou Traoré et El hadji Babacar Ndao.

Je souhaite remercier mes amis pour leur soutien. Il s'agit de Mor Mboup, Mayoro Barry, Moustapha Faye, Abou Fofana, Saloum Ndao, Babacar Sine, Adama Faye, Alioune Mboup, Badara Wade, Talla Diop, Tidiane Diallo, Tombong Camara

Toute ma reconnaissance va à mes parents et à mes frères et soeurs, ces années d'études n'auraient pas pu se faire sans votre soutien. Maman **Youma** merci d'être aussi forte. Je t'aime énormément. Papa **Abib**, tu es mon guide et ma source d'inspiration dans cette vie. Je te suis reconnaissant pour tout ce que tu as fais, sans toi, je ne serais jamais où je suis maintenant.

Un grand merci à père **El hadji Malick Mboup** dit Maodo pour tes livres au lycée et pour ton aide.

Je remercie également tous les membres de ma belle famille particulièrement à **Good mère** et **Papa Amady**.

Merci à ma très chère épouse **Mame Dieynaba Ba** pour ton amour, ta douceur, ta patience, ta compréhension, ton soutien infailible et toutes tes attentions. Difficile d'imaginer ce qu'aurait été cette thèse sans toi à mes côtés. Je t'aime. A mon adorable fils **Abib** dont l'amour et les sourires n'ont cessé de me donner de la force là où le doute pouvait s'imposer.

Valorisations Scientifiques

Article en lien avec la thèse

— Article publié :

B. Mboup, P. Blanche, et A. Latouche. On evaluating how well a biomarker can predict treatment response with survival data. Pharmaceutical Statistics, 2020.

— Article accepté dans le Journal Clinical Oncology (JCO) Precision Oncology :

B. Mboup, C. Letourneau, A. Latouche. Insights for quantifying the long-term benefit of immunotherapy using quantile regression.

— Article en cours :

B. Mboup, A. Latouche. Testing the calibration assumption of the risk prediction model in survival data

Article en collaboration

— Morretton, J. P., Herbette, A., Cosson, C., Mboup, B., Latouche, A., Gestraud, P., ... Bataillon, G. (2019). Centrosome amplification favours survival and impairs ovarian cancer progression. BioRxiv, 623983.

Communications Orales

— (Invité) Equipe high Dimensional Biostatistics, CESP, Villejuif, novembre 2019
Evaluation de la capacité d'un biomarqueur à prédire la réponse au traitement avec des données de survies

— Séminaire laboratoire Cédric, CNAM, Paris, Octobre 2019
http://maths.cnam.fr/IMG/pdf/seminaire_cnam_bassirou_cle83891e.pdf

— Conférence : Survival Analysis for Junior Researchers, Copenhagen, Danemark, Avril 2019

<https://publicifsv.sund.ku.dk/~safjr2019/ConferenceBookSAfJR2019v2.pdf#page=22>

- 50 ème Journées de statistique, Paris Saclay, Juin 2018
Méthode d'évaluation de la capacité prédictive d'un biomarqueur avec des données censurées à droite
- Journées du G.D.R. Statistique et Santé et de la Société française de biométrie, Bordeaux, Octobre 2017 http://gdr-stat-sante.math.cnrs.fr/spip/IMG/pdf/mboup.biomarqueur_censure.pdf

Table des matières

Table des matières	ix
Liste des figures	xi
Liste des tableaux	xiii
1 Introduction	1
1.1 Contexte	2
1.2 Plan de la thèse	5
2 Rappel de l'évaluation d'un biomarqueur prédictif : cas d'une réponse binaire	9
2.1 Introduction	10
2.2 Notations et Définitions	10
2.3 Modélisation et estimation	11
2.4 Mesures de performances dérivées de $\Delta(Y)$	12
2.5 Simulations	14
2.6 Courbe de prédictivité du biomarqueur par bras de traitement	17
2.7 Conclusion	20
3 Extension aux données censurées à droite	23
3.1 Modélisation et estimation	24
3.2 Article publié dans la revue Pharmaceutical Statistics	25
3.3 Compléments	41
4 Test de l'hypothèse de calibration d'un modèle de prédiction avec des données censurées	45
4.1 Introduction	46
4.2 Statistique de Test	47
4.3 Simulations	48

4.4	Application	50
5	Contribution au développement d'un biomarqueur pronostique pour le cancer de l'ovaire	55
5.1	Introduction	56
5.2	Indexe des centrosomes par noyau (ICN)	57
5.3	Matériels et Méthodes	57
5.4	Résultats	59
5.5	Conclusion	60
5.6	Résumé de l'article en français	60
5.7	Manuscrit publié dans BioRxiv	61
6	Quantification du bénéfice à long terme de l'immunothérapie à l'aide de la régression quantile censurée	105
6.1	Problématique	106
6.2	Modélisation et estimation	106
6.3	Article en révision dans la revue JCO Precision Oncology	108
7	Conclusion et perspectives	121
A	Annexes	127
A.1	Code R pour simulation de données chapitre 2	127
B	Annexes	133
B.1	Code R pour l'extension de données réels	133

Liste des figures

2.1	Le marqueur Y permet à 46% des patients d'éviter la chimiothérapie adjuvante	18
2.2	46% des patients ont un effet absolu traitement négatif donc ils recevront pas le traitement car avant ce seuil le risque des patients traités est supérieur au risque des patients non traités	19
4.1	Distribution de la statistique de test sous l'hypothèse nulle, en Bleu la distribution de la loi théorique du χ^2 à 9 ddl.	51
4.2	Puissance des tests sous le risque mal spécifié sur une taille d'échantillon respective $n = 600$, $n = 1000$ et $n = 5000$	51
5.1	Courbe de prédictivité pour la détermination du seuil de l'indexe des centrosomes par noyaux (ICN)	60
6.1	Scatter plot des résidus de martingale par rapport aux valeurs de quantile ajustées. Les lignes verticales en rouges correspondent aux déciles (de 10% à 60%)	108

Liste des tableaux

2.1	Combinaisons de réponse avec et sans traitement	16
5.1	Quelques paramètres cliniques	58

Chapitre 1

Introduction

Sommaire

1.1	Contexte	2
1.2	Plan de la thèse	5

1.1 Contexte

Les progrès techniques et scientifiques dans des domaines tels que le profilage de l'expression des gènes ont permis de détecter d'éventuels biomarqueurs tumoraux. Un biomarqueur est défini comme une caractéristique mesurée objectivement et évalué comme un indicateur de processus biologique, physiologique, ou pathologique ou de la réponse thérapeutique. Il peut s'agir de protéines, constituants de nos cellules, ou des gènes, fragments d'ADN qui influencent le comportement des cellules. En effet, la médecine personnalisée ou stratifiée correspond à une approche thérapeutique où l'objectif est de sélectionner les patients auxquels administrer un traitement en fonction d'un biomarqueur. Malgré les efforts considérables dans l'évaluation des biomarqueurs, la liste de biomarqueurs utilisés en pratique clinique reste courte notamment à cause d'une mauvaise validation et d'un manque de reproductibilité. Ainsi, le prochain défi consiste à évaluer et valider rigoureusement ces biomarqueurs candidats. Il existe différents types de biomarqueurs qui sont définis en fonction de leur utilité. Nous évoquerons dans cette thèse la valeur pronostique et prédictive d'un biomarqueur.

La valeur pronostique d'un biomarqueur correspond à sa capacité à prévoir l'évolution clinique de la maladie en l'absence de traitement et donc à identifier les patients à fort ou faible risque de survenue d'un événement d'intérêt par exemple la rechute d'un cancer. Un exemple de biomarqueur pronostique et prédictif est la signature génomique, Oncotype-Dx construite à partir de 21 gènes et permet de classer les patients à haut et bas risque de rechute du cancer du sein [Albain et al., 2010].

La valeur prédictive d'un biomarqueur renvoie à sa capacité de prédire la réponse à un traitement donné et donc de distinguer les patients pouvant tirer bénéfice de ce traitement de ceux pour lesquels aucun bénéfice n'est a priori attendu. Démontrer le caractère prédictif d'un marqueur équivaut ainsi à démontrer l'utilité clinique du test l'identifiant, c'est-à-dire son aptitude à améliorer le devenir clinique des patients en aidant à la prise de décision dans le cadre de leur prise en charge thérapeutique. Par exemple, la signature génomique Oncotype-Dx est à la fois pronostique et prédictive et permet la recommandation de la chimiothérapie pour le traitement du cancer du sein.

En oncologie les traitements comme la chimiothérapie sont très efficaces et augmentent la survie globale. Mais certains patients subissent les effets indésirables alors que le traitement est inutile après chirurgie. Par exemple, les femmes souffrant du cancer du sein diagnostiquées avec récepteurs oestrogène positif (ER+) et une atteinte ganglionnaire sont souvent traitées à la fois avec un traitement hormonal, par exemple le tamoxifène, et une chimiothérapie adjuvante après une intervention chirurgicale. Parmi ces femmes, il n'y a qu'un seul groupe qui bénéficie de la chimiothérapie alors que toutes les autres n'en subissent que les effets

indésirables. Il est alors important de trouver des biomarqueurs qui peuvent être utilisés pour prédire quelles femmes sont susceptibles de bénéficier ou pas de la chimiothérapie adjuvante.

La méthodes de validation courante d'un biomarqueur prédictif se déroule en deux étapes : d'abord sur la détermination de sa capacité pronostique puis un test de son interaction avec le traitement.

Cette approche a été invalidée par Huang et al. [2012] dans une simulation réaliste dans laquelle deux biomarqueurs continus, possédant la même interaction avec le traitement mais qui ont une capacité de discrimination très différentes vis à vis du critère de jugement. Par conséquent, une interaction significative est une condition nécessaire mais pas suffisante pour établir la capacité prédictive d'un biomarqueur candidat [Janes et al., 2011].

Un autre aspect fondamental est la détermination de la valeur seuil d'un biomarqueur continu. Les valeurs moyennes ou médianes des biomarqueurs sont largement utilisées comme seuil du biomarqueur dans une majorités d'études et ce sans aucune justification[Chen et al., 2011], mais il est souvent souhaitable de déterminer les seuils en utilisant des méthodes rigoureuses pour obtenir un seuil *optimal*. Nous avons considéré comme point de départ l'approche de Janes et al. [2014] utilisant les courbes de prédictivité par bras de traitement qui est une approche convaincante et compréhensible par les cliniciens pour évaluer la capacité prédictive d'un biomarqueur.

Initialement développée pour un critère de jugement binaire, nous avons étendu cette méthode au cas à une réponse censurée à droite. La motivation pour intégrer la dimension temporelle est que, comme nous le détaillerons au chapitre 3, la valeur optimal du seuil du biomarqueur dépend entre autres, de l'effet du traitement. Nous proposons d'estimer le seuil optimal du biomarqueur à différents horizons de prédiction en tenant compte d'un éventuel effet du traitement dépendant du temps. L'inférence repose sur l'estimation de la différence du risque conditionnel entre les deux bras de traitement. C'est la raison pour laquelle nous avons employé un modèle de risque robuste à savoir le modèle logistique dépendant du temps. La particularité de ce modèle est qu'il permet à la fois de considérer des coefficients constant et dépendant du temps. L'estimation des coefficients de ce modèle repose sur la résolution des équations du score associé à la méthode du maximum de vraisemblance pour données non censurés, auxquelles des pondérations par l'inverse de la probabilité de censure sont ajoutées, pour tenir compte des données censurées.

Une limite de l'estimation du seuil du biomarqueur continu est qu'elle suppose implicitement que le modèle de prédiction du risque est correctement calibré. La calibration est la capacité du modèle à fournir des risques prédits qui sont en cohérence avec les risques observés. Nous avons étudié l'effet d'une mauvaise calibration sur la détermination du seuil

du biomarqueur. En d'autres termes, le seuil serait-il optimal si l'hypothèse de calibration n'est pas vérifiée. Nous proposons en ce sens un test de l'hypothèse de calibration avec des données de survies qui sera valide au-delà de 30% à 50% de taux de censures contrairement à ceux déjà existants [D'agostino and Nam, 2003; Demler et al., 2015].

Nous avons également contribué à l'évaluation d'un nouveau biomarqueur pronostique pour le cancer de l'ovaire. Cette étude est une illustration convaincante de l'utilisation des courbes de prédictivité dans la détermination du seuil du biomarqueur pour classer les patients en haut ou bas risque de rechute.

Une autre situation, où seulement un sous-groupe des patients traités bénéficie du traitement expérimental, émerge actuellement dans le cadre de l'immunothérapie. Le principe de cette nouvelle classe thérapeutique est d'aider le système immunitaire à reconnaître et détruire les cellules tumorales au lieu de les s'attaquer directement. Elle a montré son efficacité sur la survie globale dans plusieurs localisations tumorales [Rittmeyer et al., 2017]. Cependant, son effet n'est pas immédiat et son bénéfice de survie se fait grâce à une durée prolongée du traitement. Cet effet retardé dans le temps se traduit par la séparation tardive des courbes de survies. Une première conséquence, est que l'hypothèse de proportionnalité des risques est mise en défaut. Se pose ensuite la question de la quantification du bénéfice de ce nouveau traitement comparé à une chimiothérapie.

Récemment, Pons-Tostivint et al. [2019] ont défini la proportion de *long-répondeur* comme la proportion de patients avec une survie deux fois plus longue que la médiane du bras de traitement. Avec cette définition, ils ont établi dans une méta-analyse sur des données agrégées que la proportion de patients long-répondeur est significativement plus élevée chez les patients traités par immunothérapie que chez ceux traités par d'autres médicaments anticancéreux.

Nous avons proposé d'améliorer et de généraliser cette définition, car la médiane de survie n'est pas toujours atteinte, en employant la régression quantile pour données censurées pour quantifier le bénéfice éventuel à long terme de l'immunothérapie. Nous avons travaillé sur des données reconstruites à partir de courbes de survie publiées car l'accès aux données individuelles est illusoire au vu de la concurrence actuelle sur ce type de traitement. La régression quantile est un outil très pertinent pour quantifier le bénéfice de survie qu'apporte le traitement au niveau de chaque quantile de la distribution du temps de survie. En effet, elle permet d'avoir une description plus précise de l'effet traitement sur la distribution entière de la variable d'intérêt qu'une simple régression linéaire, qui se focalise sur la moyenne conditionnelle.

1.2 Plan de la thèse

Dans le Chapitre 2, nous présentons la méthodologie d'évaluation d'un biomarqueur prédictif de la réponse au traitement proposée par Janes et al. [2014] qui repose sur une variable réponse binaire et les quantités de mesure de performances d'un biomarqueur prédictif.

Le Chapitre 3 est une extension du Chapitre 2 avec des données censurées. Dans cette partie, nous avons proposé une méthode qui tient compte dans la modélisation un effet du traitement variant avec le temps et estimé le seuil optimal du biomarqueur prédictif. Par conséquent, nous avons montré dans cette partie que ce seuil varie en fonction du temps.

Dans le Chapitre 4, nous développons un test de l'hypothèse de calibration d'un modèle de prédiction de risque avec des données censurées à droite. Ce test est fondé sur la définition de l'hypothèse de calibration au sens faible.

Le Chapitre 5 propose une étude de cas d'un nouveau biomarqueur pronostique dans le cancer de l'ovaire. Cette étude est réalisé dans le cadre d'une collaboration avec une équipe de biologiste de l'institut Curie. Dans cette partie, nous allons voir que l'utilisation de la courbe de prédictivité permet de déterminer le seuil du biomarqueur et de classer les patients en haut et bas risque de rechute.

Le Chapitre 6 présente une perspective pour la quantification du bénéfice à long terme de l'immunothérapie. La pertinence dans cette partie est l'utilisation de la régression quantile censurée avec des données auxquels l'hypothèse de risque proportionnel n'est pas vérifié.

Enfin, nous clôturons ce manuscrit par une conclusion des différents travaux et évoquons quelques perspectives.

Bibliographie

- Albain, K. S., Barlow, W. E., Shak, S., Hortobagyi, G. N., Livingston, R. B., Yeh, I.-T., Ravdin, P., Bugarini, R., Baehner, F. L., Davidson, N. E., et al. (2010). Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy : a retrospective analysis of a randomised trial. *The lancet oncology*, 11(1) :55–65. 2
- Chen, D. T., Hsu, Y. L., Fulp, W. J., Coppola, D., Haura, E. B., Yeatman, T. J., and Cress, W. D. (2011). Prognostic and predictive value of a malignancy-risk gene signature in early-stage non-small cell lung cancer. *J. Natl. Cancer Inst.*, 103(24) :1859–1870. 3
- D’agostino, R. and Nam, B.-H. (2003). Evaluation of the performance of survival analysis models : discrimination and calibration measures. *Handbook of statistics*, 23 :1–25. 4
- Demler, O. V., Paynter, N. P., and Cook, N. R. (2015). Tests of calibration and goodness-of-fit in the survival setting. *Statistics in medicine*, 34(10) :1659–1680. 4
- Huang, Y., Gilbert, P. B., and Janes, H. (2012). Assessing treatment-selection markers using a potential outcomes framework. *Biometrics*, 68(3) :687–696. 3
- Janes, H., Brown, M. D., Huang, Y., and Pepe, M. S. (2014). An approach to evaluating and comparing biomarkers for patient treatment selection. *Int J Biostat*, 10(1) :99–121. 3, 5
- Janes, H., Pepe, M. S., Bossuyt, P. M., and Barlow, W. E. (2011). Measuring the performance of markers for guiding treatment decisions. *Ann. Intern. Med.*, 154(4) :253–259. 3
- Pons-Tostivint, E., Latouche, A., Vaffard, P., Ricci, F., Loirat, D., Hescot, S., Sablin, M.-P., Rouzier, R., Kamal, M., Morel, C., et al. (2019). Comparative analysis of durable responses on immune checkpoint inhibitors versus other systemic therapies : a pooled analysis of phase iii trials. *JCO Precision Oncology*, 3 :1–10. 4
- Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., Von Pawel, J., Gadgeel, S. M., Hida, T., Kowalski, D. M., Dols, M. C., et al. (2017). Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (oak) : a phase 3, open-label, multicentre randomised controlled trial. *The Lancet*, 389(10066) :255–265.

Chapitre 2

Rappel de l'évaluation d'un biomarqueur prédictif : cas d'une réponse binaire

Sommaire

2.1	Introduction	10
2.2	Notations et Définitions	10
2.3	Modélisation et estimation	11
2.4	Mesures de performances dérivées de $\Delta(Y)$	12
2.4.1	Cas dans lequel la stratégie par défaut est de traiter tout le monde	12
2.4.2	Cas dans lequel la stratégie par défaut est de ne traiter personne	13
2.4.3	Variance et gain total	14
2.5	Simulations	14
2.6	Courbe de prédictivité du biomarqueur par bras de traitement	17
2.7	Conclusion	20

2.1 Introduction

En pratique, seul un test d'interaction entre le traitement et le biomarqueur est utilisé pour établir la capacité prédictive d'un biomarqueur pronostique. Il a été montré que deux biomarqueurs possédant la même interaction avec le traitement peuvent avoir des capacités prédictives très différentes (notamment une AUC proche 0.5 ce qui correspond à un marqueur non informatif). La présence d'une interaction entre le traitement et le biomarqueur est une condition nécessaires mais pas suffisante [Janes et al., 2011]. Janes et al. [2014] ont proposé une méthodologie pour valider un biomarqueur prédictif qui va au-delà d'une simple test d'interaction.

Dans cette section, nous présenterons cette méthodologie.

2.2 Notations et Définitions

Nous notons $Treat$ l'indicateur du traitement randomisé qui vaut 1 si le patient reçoit le traitement expérimentale et 0 sinon. D est l'indicateur de survenue de l'événement, par exemple le décès, la récurrence de la maladie ou une toxicité associée au traitement. Notons $\rho_0 = P(D = 1 | Treat = 0)$ et $\rho_1 = P(D = 1 | Treat = 1)$. ρ_0 et ρ_1 vont guider la décision par défaut de recommandation de traitement ou non de tous les patients. $\rho_0 - \rho_1$ est appelé l'effet marginal du traitement et s'interprète comme la diminution, grâce au traitement, de la proportion de patients ayant l'événement.

- Si $\rho_0 - \rho_1 > 0$, alors la décision par défaut est de traiter tous les patients. Dans ce cas bien précis, un biomarqueur Y sera utile pour identifier un sous-groupe de patients pour lequel le traitement pourrait être évité.
- Sinon, la décision par défaut sera de ne traiter personne et un biomarqueur Y sera utile pour identifier un sous-groupe de patient auquel le traitement serait bénéfique.

Une fois que nous avons défini une décision par défaut, sans la connaissance de la valeur du biomarqueur Y , nous pouvons définir une règle de décision qui incorpore l'information disponible sur le marqueur Y . L'intervention du biomarqueur Y va jouer un rôle déterminant dans le processus de sélection de sous-groupe à traiter. L'idée est d'estimer un seuil *optimal* de ce biomarqueur continu Y au delà duquel certains patients seront épargné ou bénéficieront du traitement.

Considérons que la règle de décision adoptée est de traiter tous les patients. Les valeurs du biomarqueur Y étant employées pour déterminer un sous groupe de patients pouvant éviter le traitement. On définira

- Le risque, comme la probabilité d'expérimenter l'événement conditionnellement au traitement et au biomarqueur, *i.e.* $P(D = 1|Treat, Y)$.
- L'effet absolu du traitement conditionnellement au biomarqueur, comme $\Delta(Y) = P(D = 1|Treat = 0, Y) - P(D = 1|Treat = 1, Y)$ et on note F_Δ sa fonction de répartition.

La règle optimale pour affecter le traitement est de :

- recommander le traitement à ceux pour lesquels $\Delta(Y) \geq 0$, les sujets sont appelés marqueurs positifs.
- recommander contre le traitement si $\Delta(Y) < 0$, les sujets sont appelés marqueurs négatifs.

2.3 Modélisation et estimation

On considère le cas d'un essai randomisé comportant n patients. Les observations consistent en un n -uplet $(D_i, Y_i, Treat_i)_{i=1..n}$. Le risque sera modélisé à partir d'un modèle de régression logistique comportant un terme d'interaction entre le traitement $Treat$ et le biomarqueur Y ,

$$g(P(D = 1|Treat, Y)) = \beta_0 + \beta_1 Treat + \beta_2 Y + \beta_3 Treat \times Y$$

avec g la fonction de lien logit, *i.e.* $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$.

Les risques dans chaque bras de traitement sont estimés par substitution comme,

$$\begin{aligned}\widehat{risk}_0(Y) &= \widehat{P}(D = 1|Treat = 0, Y) = g^{-1}(\widehat{\beta}_0 + \widehat{\beta}_2 Y) \\ \widehat{risk}_1(Y) &= \widehat{P}(D = 1|Treat = 1, Y) = g^{-1}(\widehat{\beta}_0 + \widehat{\beta}_1 + \widehat{\beta}_2 Y + \widehat{\beta}_3 Y)\end{aligned}$$

Les coefficients $\widehat{\beta}$ sont des estimations obtenues par la méthode du maximum de vraisemblance. L'effet absolu estimé du traitement est égal à la différence des deux risques

$$\widehat{\Delta}(Y) = \widehat{risk}_0(Y) - \widehat{risk}_1(Y)$$

La solution de l'équation $\widehat{\Delta}(Y) = 0$ est $\widehat{\gamma} = -\widehat{\beta}_1/\widehat{\beta}_3$. $\widehat{\gamma}$ représente le seuil optimal (au sens de Janes) du biomarqueur continu. **Cette valeur dépend donc des effets estimés du traitement et de l'interaction.** Un intervalle de confiance au niveau α peut être obtenu par delta-méthode :

$$\widehat{\gamma} \pm z_{\alpha/2} \sigma_\gamma / \sqrt{n}$$

où $\sigma_\gamma = 1/\beta_3 \sqrt{\text{var}(\beta_1) + 2\gamma \text{Cov}(\beta_1, \beta_3) + \gamma^2 \text{var}(\beta_3)}$ et $z_{\alpha/2}$ est le quantile d'ordre $\alpha/2$ de la distribution normale centrée réduite. L'estimation de cet intervalle de confiance par bootstrap sera détaillé dans le chapitre 3.

2.4 Mesures de performances dérivées de $\Delta(Y)$

La connaissance de la valeur seuil du biomarqueur peut permettre d'épargner ou d'attribuer le traitement. L'estimation des quantités suivantes dépend de l'estimation de l'effet absolu du traitement $\Delta(Y) = P(D = 1|Treat = 0, Y) - P(D = 1|Treat = 1, Y)$.

2.4.1 Cas dans lequel la stratégie par défaut est de traiter tout le monde

- La "Proportion de marqueur négatif" correspond à la proportion des patients qui peuvent renoncer au traitement.

$$P_{neg} = P(\Delta(Y) < 0)$$

- Le "Bénéfice moyen de non traitement parmi les marqueurs négatifs" noté B_{neg} qui est la différence entre la proportion de patients expérimentant l'événement dans chaque bras de traitement, parmi ceux à qui on ne recommande pas le traitement. Cette quantité est définie comme

$$B_{neg} = P(D = 1|Treat = 1, \Delta(Y) < 0) - P(D = 1|Treat = 0, \Delta(Y) < 0)$$

On montre aussi que :

$$\begin{aligned} B_{neg} &= P(D = 1|Treat = 1, \Delta(Y) < 0) - P(D = 1|Treat = 0, \Delta(Y) < 0) \\ &= E\{\mathbb{I}\{D = 1\}|Treat = 1, \Delta(Y) < 0\} - E\{\mathbb{I}\{D = 1\}|Treat = 0, \Delta(Y) < 0\} \\ &= E\{E\{\mathbb{I}\{D = 1\}|Treat = 1, Y\} - E\{\mathbb{I}\{D = 1\}|Treat = 0, Y\}|\Delta(Y) < 0\} \\ &= E\{P\{D = 1|Treat = 1, Y\} - P\{D = 1|Treat = 0, Y\}|\Delta(Y) < 0\} \\ &= E(-\Delta(Y)|\Delta(Y) < 0) \end{aligned}$$

- Le "Bénéfice moyen de traitement parmi les marqueurs positifs" noté B_{pos} est défini comme $B_{pos} = P(D = 1|Treat = 0, \Delta(Y) \geq 0) - P(D = 1|Treat = 1, \Delta(Y) \geq 0)$ et on montre que

$$\begin{aligned} B_{pos} &= P(D = 1|Treat = 0, \Delta(Y) \geq 0) - P(D = 1|Treat = 1, \Delta(Y) \geq 0) \\ &= E\{\mathbb{I}\{D = 1\}|Treat = 0, \Delta(Y) \geq 0\} - E\{\mathbb{I}\{D = 1\}|Treat = 1, \Delta(Y) \geq 0\} \\ &= E\{E\{\mathbb{I}\{D = 1\}|Treat = 0, Y\} - E\{\mathbb{I}\{D = 1\}|Treat = 1, Y\}|\Delta(Y) \geq 0\} \\ &= E\{P\{D = 1|Treat = 0, Y\} - P\{D = 1|Treat = 1, Y\}|\Delta(Y) \geq 0\} \\ &= E(\Delta(Y)|\Delta(Y) > 0) \end{aligned}$$

- La diminution du risque d'événement de la population selon une affectation de traitement fondée sur le marqueur, par rapport à si tous les patients étaient traités, est :

$$\Theta_1 = P(D = 1|Treat = 1) - [P(D = 1, \Delta(Y) \geq 0|Treat = 1) + P(D = 1, \Delta(Y) < 0|Treat = 0)].$$

On établit la relation entre Θ_1 , B_{neg} et P_{neg}

$$\begin{aligned}\Theta_1 &= P(D = 1|Treat = 1) - [P(D = 1, \Delta(Y) \geq 0|Treat = 1) \\ &\quad + P(D = 1, \Delta(Y) < 0|Treat = 0)] \\ &= P(D = 1|Treat = 1) - [P(D = 1|Treat = 1, \Delta(Y) \geq 0)P(\Delta(Y) \geq 0) \\ &\quad + P(D = 1|Treat = 0, \Delta(Y) < 0)P(\Delta(Y) < 0)] \\ &= [P(D = 1|Treat = 1, \Delta(Y) < 0) - P(D = 1|Treat = 0, \Delta(Y) < 0)]P(\Delta(Y) < 0) \\ &= B_{neg} \times P_{neg}\end{aligned}$$

La mesure Θ_1 est recommandée comme une mesure globale de la performance du marqueur de sélection de traitement [Janes et al., 2014]. Θ_1 varie entre 0 pour un marqueur totalement inutile, pour lequel il n'y a pas de différence dans le risque d'événement avec et sans utilisation du marqueur, et $\Theta_1 = \rho_1$ pour un marqueur parfait qui donne un risque d'événement de zéro si le traitement est recommandé en fonction du marqueur Y . Les constituants de Θ_1 , à savoir B_{neg} et P_{neg} , sont informatifs et utiles pour partager l'impact du marqueur. Les mesures B_{neg} et B_{pos} quantifient l'avantage moyen des choix de traitements recommandés respectivement pour les sujets marqueurs négatifs et marqueurs positifs.

2.4.2 Cas dans lequel la stratégie par défaut est de ne traiter personne

Considérons que l'approche adoptée est de ne traiter aucun patient. L'effet marginal du traitement est négatif et la connaissance du marqueur, Y , permet de déterminer un sous groupe de patients qui bénéficiera du traitement. Les mesures de performance sont

- la "Proportion de marqueurs positifs" est la proportion de patients qui peuvent bénéficier du traitement.

$$P_{pos} = P(\Delta(Y) \geq 0)$$

- La différence risque d'événement de la population selon une affectation de traitement fondée sur le marqueur, par rapport à si on ne traite personne, est $\Theta_2 = P(D = 1|Treat = 0) - [P(D = 1, \Delta(Y) \geq 0|Treat = 1) + P(D = 1, \Delta(Y) < 0|Treat = 0)]$.

On montre que :

$$\begin{aligned}
\Theta_2 &= P(D = 1|Treat = 0) - [P(D = 1, \Delta(Y) \geq 0|Treat = 1) \\
&\quad + P(D = 1, \Delta(Y) < 0|Treat = 0)] \\
&= P(D = 1|Treat = 0) - [P(D = 1|Treat = 1, \Delta(Y) \geq 0)P(\Delta(Y) \geq 0) \\
&\quad + P(D = 1|Treat = 0, \Delta(Y) < 0)P(\Delta(Y) < 0)] \\
&= [P(D = 1|Treat = 0, \Delta(Y) \geq 0) - P(D = 1|Treat = 1, \Delta(Y) \geq 0)]P(\Delta(Y) \geq 0) \\
&\quad + [P(D = 1|Treat = 0, \Delta(Y) < 0) - P(D = 1|Treat = 1, \Delta(Y) < 0)]P(\Delta(Y) < 0) \\
&= B_{pos} \times P_{pos}
\end{aligned}$$

La mesure Θ_2 est recommandée comme une mesure globale de la performance du marqueur de sélection de traitement [Janes et al., 2014]. Elle varie entre 0 et ρ_0 . $\Theta_2 = 0$ si $P(\Delta(Y) \geq 0) = 0$ car $\Theta_2 = \rho_0 - [0 \cdot 0 + \rho_0 \cdot 1]$ et $\Theta_2 = \rho_0$ si $P(D = 1|Treat = 1, \Delta(Y) \geq 0) = P(D = 1|Treat = 0, \Delta(Y) < 0) = 0$.

2.4.3 Variance et gain total

— La variance de l'effet absolu du traitement,

$$V_\Delta = Var(\Delta(Y)) = \int (\Delta(Y) - (\rho_0 - \rho_1))^2 dF_\Delta$$

— Le gain total est l'aire entre la courbe de l'effet absolu et la courbe de l'effet marginal du traitement,

$$TG = \int |\Delta(Y) - (\rho_0 - \rho_1)| dF_\Delta$$

Les quantités V_Δ et TG n'ont pas d'interprétation clinique, mais ont l'avantage d'être indépendantes de la règle du traitement et constituent potentiellement la base pour des comparaisons plus efficaces des marqueurs [Janes et al., 2014]. Plus elles sont grandes, plus le marqueur Y est bon. Ces quantités sont considérées comme utiles pour résumer globalement l'utilité de Y . Par contraste, les mesures précédentes sont conditionnelles à une valeur de Y , et donc ce sont des courbes qui résument l'utilité de Y .

2.5 Simulations

Nous détaillons la méthode de simulations proposée par [Janes et al., 2014] qui est utilisée pour reconstruire des données qui imitent celle d'un essai randomisé dans le contexte du cancer du sein.

Nous supposons que, pour chacun des individus i d'un échantillon de taille n , nous observons l'ensemble suivant de variables aléatoires. L'accès au traitement par chimiothérapie est représenté par une variable aléatoire $Treat$, qui prend la valeur 1 si l'individu est traité, 0 sinon, dont la moitié des participants de l'étude sont traitées, i.e. $P(Treat = 1) = P(Treat = 0) = 1/2$ et un biomarqueur Y , qui suit la loi normale centrée réduite. L'efficacité du traitement est mesuré au travers une variable de réponse au traitement, notée D . Le modèle repose sur l'existence de deux variables de réponse au traitement, notées $D(1)$ et $D(0)$, selon que l'individu reçoit le traitement expérimental $Treat = 1$ ou le traitement standard $Treat = 0$. Ces variables correspondent aux réponses contre-factuelles du traitement. Elles ne sont jamais simultanément observées pour une même patiente. Ainsi, pour une patiente traitée, $D(1)$ est observée tandis que $D(0)$ est inconnue. Dans ce cas, la variable $D(0)$ correspond à la réponse qui aurait été réalisée si la patiente n'avait pas été traitée par le traitement expérimental. On dit aussi que la variable $D(0)$ représente la réponse contre-factuelle au traitement. Pour une patiente non traitée, on observe au contraire $D(0)$, tandis que $D(1)$ est inconnue. Les taux de récurrences ou décès ont été générés à 21% et 24% avec et sans traitement par chimiothérapie respectivement, i.e. $P(D(1) = 1) = 21\%$ et $P(D(0) = 1) = 24\%$. Les variables de réponses potentielles sont supposées être chacune engendrée par un modèle de régression logistique de la forme :

$$\begin{aligned}\text{logit}\{P(D(0) = 1|Y)\} &= \beta_0 + \beta_2 Y \\ \text{logit}\{P(D(1) = 1|Y)\} &= \beta_0 + \beta_1 + (\beta_2 + \beta_3) Y \\ P(D(0) = 0, D(1) = 1|Y) &= P(D(0) = 0|Y) \times P(D(1) = 1|Y)\end{aligned}$$

Notons que tous les autres mesures et estimateurs dépendent uniquement des distributions marginales $P(D(0) = 1|Y)$ et $P(D(1) = 1|Y)$ et sont invariantes au choix d'une distribution jointe. La loi jointe des réponses contre factuelles conditionnellement au biomarqueur Y est alors donnée :

$$P(D(0) = 0, D(1) = 1|Y) = P(D(0) = 0|Y) \times P(D(1) = 1|Y)$$

Ce qui implique l'indépendance entre $D(0)$ et $D(1)$ sachant Y , et nous en déduisons les autres combinaisons de probabilités jointes à savoir :

$$\begin{aligned}P(D(0) = 0, D(1) = 0|Y) &= P(D(0) = 0|Y) \times P(D(1) = 0|Y) \\ P(D(0) = 1, D(1) = 0|Y) &= P(D(0) = 1|Y) \times P(D(1) = 0|Y) \\ P(D(0) = 1, D(1) = 1|Y) &= P(D(0) = 1|Y) \times P(D(1) = 1|Y)\end{aligned}$$

Algorithme de simulation

1. Nous simulons les paires de réponses potentielles $(D(0), D(1))$ à partir d'une distribution multinomiale résultante des équations précédentes. Nous représentons toutes les combinaisons de réponses potentielles possibles et leurs probabilités d'apparition dans le tableau suivant :

TABLEAU 2.1 – Combinaisons de réponse avec et sans traitement

D(0)	D(1)	Probabilités
0	0	p_1
0	1	p_2
1	0	p_3
1	1	p_4

Nous déterminons toutes ces probabilités p_1 , p_2 , p_3 et p_4 afin de simuler des paires de réponses potentielles $(D(0), D(1))$ à partir de la distribution multinomiale. Ces probabilités s'obtiennent à partir des équations suivantes :

$$\begin{aligned}
 P(D(0) = 0, D(1) = 0) &= \int P(D(0) = 0, D(1) = 0 | Y = y) f_Y(y) dy \\
 &= p_1 \\
 P(D(0) = 0, D(1) = 1) &= \int P(D(0) = 0, D(1) = 1 | Y = y) f_Y(y) dy \\
 &= p_2 \\
 P(D(0) = 1, D(1) = 0) &= \int P(D(0) = 1, D(1) = 0 | Y = y) f_Y(y) dy \\
 &= p_3 \\
 P(D(0) = 1, D(1) = 1) &= \int P(D(0) = 1, D(1) = 1 | Y = y) f_Y(y) dy \\
 &= p_4
 \end{aligned}$$

où f_Y est la densité normale centrée réduite de Y .

2. L'application du théorème de Bayes permet de calculer la densité conditionnelle $f(Y|D(0), D(1))$:

$$f(y|D(0) = d_0, D(1) = d_1) = \frac{P(D(0) = d_0, D(1) = d_1 | Y = y) f_Y(y)}{P(D(0) = d_0, D(1) = d_1)}$$

où $d_0 \in \{0, 1\}$ et $d_1 \in \{0, 1\}$

Une fois que nous avons estimé la densité conditionnelle $f(Y|D(0), D(1))$, nous estimons la fonction de répartition

$$F_{Y|D(0), D(1)}(y) = \int_{-\infty}^y f(x|D(0), D(1))dx$$

Par conséquent nous simulons Y sachant $D(0)$ et $D(1)$ en utilisant la méthode d'inversion de la fonction de répartition c'est à dire on génère U suivant une loi uniforme sur $[0, 1]$ tel que :

$$Y = F_{Y|D(0), D(1)}^{-1}(U)$$

$F_{Y|D(0), D(1)}^{-1}$ est la fonction inverse généralisée qui coïncide avec $F_{Y|D(0), D(1)}^{-1}$ lorsque $F_{Y|D(0), D(1)}$ est bijective. En prenant pour le couple (d_0, d_1) les valeurs obtenues à l'étape 1 précédent.

3. Nous simulons le traitement $Treat$ indépendamment des réponses potentielles et du marqueur à partir d'une Binomiale de paramètre $1/2$.
4. La variable de réponse au traitement observée peut donc se déduire des réponses potentielles et de la variable de traitement $Treat$ par la relation :

$$D = D(1)Treat + D(0)(1 - Treat).$$

Seul le triplet $(D, Treat, Y)$ est observé pour chaque individu.

2.6 Courbe de prédictivité du biomarqueur par bras de traitement

L'étape pour guider les décisions de traitement repose sur l'estimation de deux courbes de prédictivités, qui représentent les risques conditionnels estimés, $\hat{P}(D = 1|Treat = 0, Y)$ et $\hat{P}(D = 1|Treat = 1, Y)$ en fonction de la proportion de la population ayant une valeur au biomarqueur inférieur à $\hat{\gamma}$, i.e $F(Y)$ (Figure 2.1). Autrement dit, elles représentent les risques d'avoir un événement dans chaque bras de traitement en fonction du percentile du biomarqueur. Cette courbe est un outil graphique permettant de déterminer le seuil du biomarqueur. La première étape dans l'évaluation d'un biomarqueur est la construction de ces courbes.

Une autre façon d'étudier la prédictivité est de représenter l'effet absolu du traitement, $\Delta(Y)$, en fonction du pourcentage de la population ayant une valeur au biomarqueur inférieur à la valeur $\hat{\gamma}$ comme illustré sur la Figure 2.2.

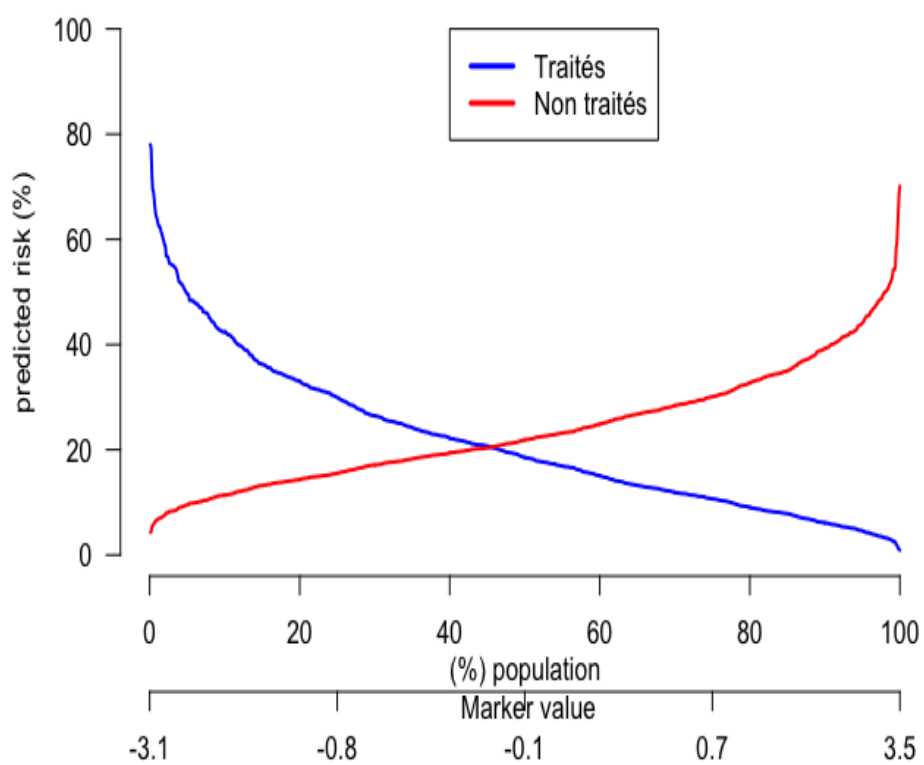


FIGURE 2.1 – Le marqueur Y permet à 46% des patients d’éviter la chimiothérapie adjuvante

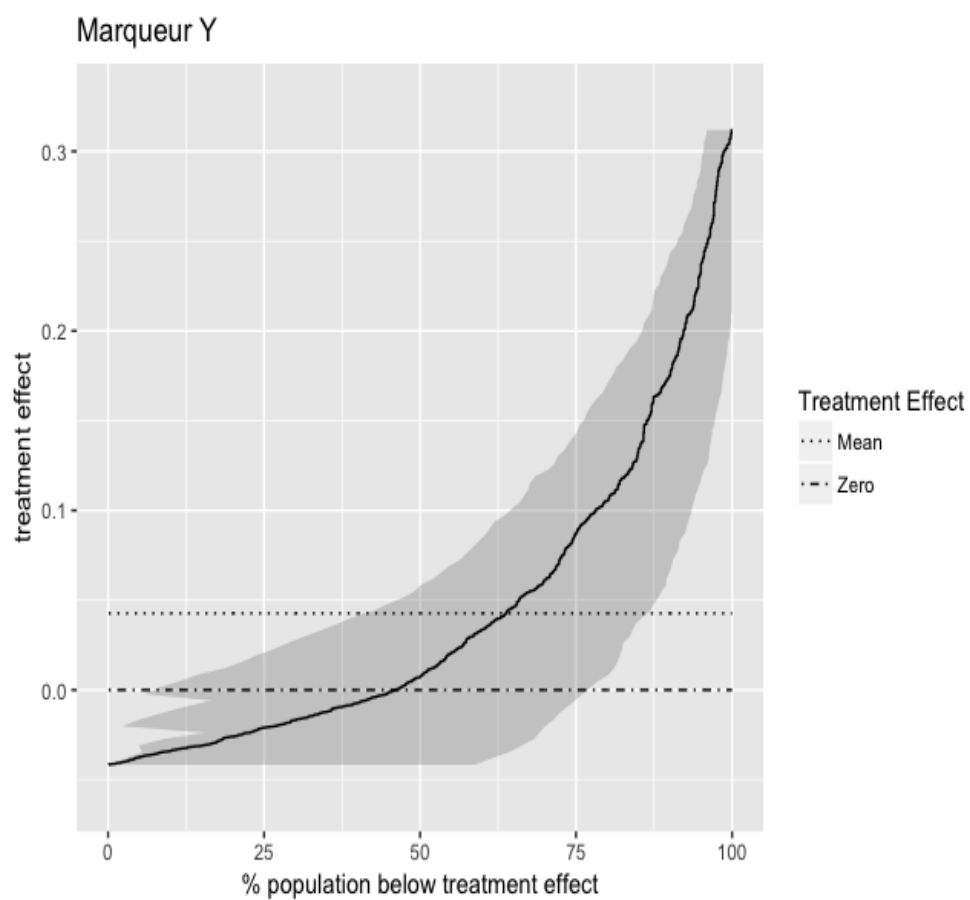


FIGURE 2.2 – 46% des patients ont un effet absolu traitement négatif donc ils recevront pas le traitement car avant ce seuil le risque des patients traités est supérieur au risque des patients non traités

2.7 Conclusion

Dans ce chapitre, nous avons rappelé la méthodologie de Janes et al. [2014] pour valider un biomarqueur prédictif. Elle repose sur une variable réponse binaire et permet d’aller au delà d’un simple test d’interaction entre traitement et le biomarqueur. Cette approche d’évaluation d’un biomarqueur permet aussi de répondre aux questions comme :

- le biomarqueur aide-t-il à prendre des décisions de traitement ?
- quelle proportion de patients auront leurs recommandations de traitement modifiées en utilisant le marqueur ?

Les réponses à ces questions peuvent aider à déterminer si un biomarqueur doit être utilisé pour prendre des décisions de traitement du patient.

Nous étendons ce travail dans le chapitre suivant en introduisant l’aspect temporel avec les données de survie.

Bibliographie

- Janes, H., Brown, M. D., Huang, Y., and Pepe, M. S. (2014). An approach to evaluating and comparing biomarkers for patient treatment selection. *Int J Biostat*, 10(1) :99–121. 10, 13, 14, 20
- Janes, H., Pepe, M. S., Bossuyt, P. M., and Barlow, W. E. (2011). Measuring the performance of markers for guiding treatment decisions. *Ann. Intern. Med.*, 154(4) :253–259. 10

Chapitre 3

Extension aux données censurées à droite

Sommaire

3.1	Modélisation et estimation	24
3.2	Article publié dans la revue <i>Pharmaceutical Statistics</i>	25
3.3	Compléments	41
3.3.1	Une amélioration des intervalles de confiance	41
3.3.2	Data extension	41
3.3.3	Les résultats obtenus	42

L'objectif dans ce chapitre est d'évaluer la capacité prédictive du biomarqueur continu et d'estimer le seuil optimal du biomarqueur et son intervalle de confiance à différents horizon de prédiction car le seuil du biomarqueur est essentiel pour la prise de décision médicale. Nous proposons d'étendre les concepts du chapitre précédent en tenant compte de l'aspect temporel où le temps d'événement est censuré. La motivation pour intégrer cette temporalité vient du fait que le seuil optimal du biomarqueur dans la méthode de Janes dépend de l'effet traitement et de l'effet de l'interaction. Comme, il est fréquent en oncologie que l'effet des traitements varie avec le temps, le seuil peut donc varier avec l'horizon de prédiction (à travers l'effet traitement et/ou l'effet de l'interaction).

Plus précisément nous considérons l'indicateur de survenue de l'événement avant l'instant t , $D(t) = \mathbb{I}(T \leq t)$ pour un délai d'événement T . Par conséquent, le risque conditionnel est une fonction dépendante du temps de la forme $P(D(t) = 1|Treat, Y)$. Un effet absolu du traitement dépendant de l'horizon de prédiction t est alors défini comme $\Delta(t, Y) = P(D(t) = 1|Treat = 0, Y) - P(D(t) = 1|Treat = 1, Y)$. En outre, nous généralisons la définition des courbes de prédictivité du marqueur par bras de traitement [Janes et al., 2014] à des courbes dépendantes du temps.

3.1 Modélisation et estimation

Soient T le délai d'évènement, C le délai de censure, $\tilde{T} = \min(T, C)$ le délai de suivi observé, $\delta = \mathbb{I}_{(T \leq C)}$ l'indicatrice de survenue de l'évènement et $D(t) = \mathbb{I}_{(T \leq t)}$ l'indicatrice de l'évènement à l'horizon de prédiction t . Nous observons $(\tilde{T}_i, \delta_i, Y_i, Treat_i)$ pour $i = 1 \dots n$, des vecteurs de variables indépendantes et identiquement distribuées.

Le modèle de régression logistique dépendant du temps avec les effets du marqueurs et du traitement variant dans le temps est défini avec la forme suivante

$$g(P(D(t) = 1|Treat, Y)) = \alpha(t) + \beta_1(t)Treat + \beta_2(t)Y + \beta_3(t)Treat \times Y \quad (3.1)$$

où g est la fonction logit, ie $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$.

Autrement dit,

$$\text{logit}(P(T \leq t|Treat, Y)) = \alpha(t) + \beta_1(t)Treat + \beta_2(t)Y + \beta_3(t)Treat \times Y.$$

Nous supposons dans la suite que l'effet du biomarqueur et de l'interaction sont constants. L'intercepte du modèle ainsi que l'effet du traitement dépendent du temps. Ce modèle a été utilisé par Zheng et al. [2006] pour évaluer la valeur pronostique de plusieurs biomarqueurs. Par conséquent, le modèle 3.1 peut être réécrit sous la forme suivante

$$\text{logit}(P(T \leq t|Treat, Y)) = \beta_0(t) + \beta_1(t)Treat + \beta_2Y + \beta_3Treat \times Y. \quad (3.2)$$

Le risque s'exprime donc comme :

$$P(T \leq t | Treat, Y) = \text{logit}^{-1} \{ \beta_0(t) + \beta_1(t)Treat + \beta_2 Y + \beta_3 Treat \times Y \} \quad (3.3)$$

L'estimation des coefficients (ou effets) du modèle repose sur la résolution d'équations du score associées à la méthode du maximum de vraisemblance pour données non censurées, auxquelles des pondérations par l'inverse de la probabilités de censure (IPCW) sont ajoutées, pour tenir compte des données censurées. Suivant Scheike et al. [2008], nous définissons le vecteur des estimations $\hat{\beta}(t)$ comme les solutions de l'équation estimée

$$\sum_{i=1}^n D_{\beta}^{\top}(t; \mathbf{W}_i) \left[\frac{\delta_i \mathbb{I}(\tilde{T}_i \leq t)}{\hat{G}_i(\tilde{T}_i | \mathbf{W}_i)} - P_{\beta}(t; \mathbf{W}_i) \right] = 0, \quad (3.4)$$

où $P_{\beta}(t; \mathbf{W}_i)$ est le terme à droite de l'équation 3.3, $\mathbf{W}_i = (Treat_i, Y_i)$ et $D_{\beta}^{\top}(t; \mathbf{W}_i) = \partial P_{\beta}(t; \mathbf{W}_i) / \partial \beta(t)$. Pour tenir compte de la censure, le poids $\hat{G}_i(\tilde{T}_i | \mathbf{W}_i)$ doit être un estimateur consistant de la probabilité de non censure au temps $\tilde{T}_i$, étant donné $\mathbf{W}_i$, i.e. $G(\tilde{T}_i | \mathbf{W}_i) = P(C_i > \tilde{T}_i | \mathbf{W}_i, \tilde{T}_i)$. Plus précisément, nous supposons que la censure C est indépendante de la variable d'intérêt T conditionnellement à $\mathbf{W} = (Treat, Y)$. L'utilisation des pondérations conduit à des équations d'estimation sans biais car

$$\begin{aligned} \mathbb{E} \left\{ \frac{\delta_i \mathbb{I}(\tilde{T}_i \leq t)}{G(\tilde{T}_i | \mathbf{W}_i)} \middle| \mathbf{W}_i \right\} &= \mathbb{E} \left\{ \mathbb{E} \left\{ \mathbb{I}(T_i \leq t) \frac{\mathbb{I}(T_i \leq C_i)}{G(T_i | \mathbf{W}_i)} \middle| \mathbf{W}_i, T_i \right\} \middle| \mathbf{W}_i \right\} \\ &= \mathbb{E} \left\{ \frac{\mathbb{I}(T_i \leq t)}{G(T_i | \mathbf{W}_i)} \mathbb{E} \left\{ \mathbb{I}(T_i \leq C_i) \middle| \mathbf{W}_i, T_i \right\} \middle| \mathbf{W}_i \right\} \\ &= \mathbb{E} \left\{ \frac{\mathbb{I}(T_i \leq t)}{G(T_i | \mathbf{W}_i)} G(T_i | \mathbf{W}_i) \middle| \mathbf{W}_i \right\} \\ &= P(D_i(t) = 1 | \mathbf{W}_i), \end{aligned}$$

L'estimation des coefficients et la résolution de l'équation $\hat{\Delta}(t, Y) = 0$, nous permettent d'estimer le seuil optimal du biomarqueur $\forall t$, $\hat{\gamma}(t) = -\hat{\beta}_1(t) / \hat{\beta}_3$ et son intervalle de confiance par **bootstrap**. Nous avons montré que si l'effet du traitement varie avec le temps alors le seuil du biomarqueur varie en fonction du temps. La méthode d'estimation des coefficients est implémentée dans le **package R timereg** et **comp.risk** est la fonction utilisée.

Un biomarqueur continu peut être considéré comme prédictif si l'hypothèse nulle $H_0 : \beta_3 = 0$ est rejetée et le seuil du biomarqueur $\hat{\gamma}(t)$ est dans le support de Y i.e si la valeur seuil du biomarqueur est dans l'ensemble des valeurs que peut prendre le biomarqueur.

3.2 Article publié dans la revue Pharmaceutical Statistics

Cet article est publié dans la revue Pharmaceutical Statistics et présente l'extension de la méthodologie de Janes et al. [2014] en tenant compte de l'aspect temporel, considère un

effet du traitement dépendant du temps, propose un modèle adéquat permettant d'estimer le seuil optimal du biomarqueur et en évalue les performances.

Les méthodes développées sont implémentées et accessibles à l'adresse https://github.com/BMBOUP/Optimal_threshold.

Ci-après est présenté l'article.



MAIN PAPER

WILEY

On evaluating how well a biomarker can predict treatment response with survival data

Bassirou Mboup¹ | Paul Blanche^{2,3,4} | Aurélien Latouche^{1,5} ¹INSERM, Institut Curie, PSL Research University, Paris, France²Department of Cardiology, Copenhagen University Hospital Herlev and Gentofte, Hellerup, Denmark³Department of Public Health, Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark⁴Department of Cardiology, The Heart Centre, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark⁵Department of Mathematics and Statistics, Conservatoire National des Arts et Métiers, Paris, France**Correspondence**Bassirou Mboup, INSERM, U900, Institut Curie, PSL Research University, F-92210 Saint Cloud, Paris, France.
Email: bassirou.mboup@curie.fr**Summary**

One of the objectives of personalized medicine is to take treatment decisions based on a biomarker measurement. Therefore, it is often interesting to evaluate how well a biomarker can predict the response to a treatment. To do so, a popular methodology consists of using a regression model and testing for an interaction between treatment assignment and biomarker. However, the existence of an interaction is not sufficient for a biomarker to be predictive. It is only necessary. Hence, the use of the marker-by-treatment predictiveness curve has been recommended. In addition to evaluate how well a single continuous biomarker predicts treatment response, it can further help to define an optimal threshold. This curve displays the risk of a binary outcome as a function of the quantiles of the biomarker, for each treatment group. Methods that assume a binary outcome or rely on a proportional hazard model for a time-to-event outcome have been proposed to estimate this curve. In this work, we propose some extensions for censored data. They rely on a time-dependent logistic model, and we propose to estimate this model via inverse probability of censoring weighting. We present simulations results and three applications to prostate cancer, liver cirrhosis, and lung cancer data. They suggest that a large number of events need to be observed to define a threshold with sufficient accuracy for clinical usefulness. They also illustrate that when the treatment effect varies with the time horizon which defines the outcome, then the optimal threshold also depends on this time horizon.

KEYWORDS

personalized medicine, predictive biomarker, time-to-event data

1 | INTRODUCTION

Personalized medicine currently relies on the use of biomarkers capable of discriminating patients into subgroups in order to give the most appropriate treatment. The term biomarker refers to an objectively measured characteristic (i.e. with sufficient precision and reproducible) and is evaluated as an indicator of a physiological or pathological process. It can be a measurement, such as prostate-specific antigen level, or a genetic signature. Often, biomarker measurements are numerical values, and cutoffs are defined to map a biomarker value into a binary value. This defines a biomarker test, which can be either positive or negative. However, defining appropriate cutoffs is a challenging task.

Ideally both biomarkers and cutoffs need to be carefully evaluated before being used to guide treatment decisions in clinical practice.

Establishing the predictive capacity of a biomarker to predict treatment response is often done by testing for an interaction between treatment assignment and biomarker value. The interaction quantifies how the treatment effect depends on the biomarker value. A large sample size is often needed to test for this interaction with sufficient power. Besides, Janes et al¹ have pointed out that two biomarkers having the same interaction with the treatment may actually have very different predictive capabilities. Thus, a strong interaction is a necessary but not sufficient condition for a biomarker to be predictive. Janes et al² have proposed a formal and comprehensive approach for the evaluation of biomarker performance.

So far this framework has been mainly discussed for a binary outcome, in settings in which censoring does not occur. Exceptions include Janes et al³ and Barlow,⁴ who present interesting ideas to handle survival outcome which can be right censored. The modeling assumptions imply constant effects in the risk model, which might be seen as an unnecessary limitation. Specifically, in this manuscript, we illustrate this limitation and its consequences when the treatment effects on the short- or long-term outcomes differ. In this case, the optimal cutoff to define a biomarker test depends on the time horizon which defines the outcome and so should the estimated cutoff be. Therefore, the objective of this manuscript is to present a method that takes that into account and which is a natural extension of the method of Janes et al² to survival data.

Our methodology could typically be used in the context of a retrospective evaluation of a predictive biomarker, with the aim to refine the value of a threshold that could guide a decision to treat or not to treat. This updated threshold could then be used in a (prospective) confirmatory controlled randomized trial.

The rest of the paper is organized as follows. Section 2 presents the methods for uncensored data on which we build on. Section 3 describes our proposed extensions to survival data and summarizes the main ideas of Janes et al.³ The finite sample performances of the inference procedures are evaluated by simulations in Section 4. Section 5 presents three applications of the method to publicly available randomized clinical trial data sets. We close in Section 6 with conclusions and perspectives. The supplementary material, which contains R code, is available at https://github.com/BMBOUP/Optimal_threshold.

2 | THE FRAMEWORK OF JANES ET AL

We consider the setting of a randomized clinical trial with two arms, with an untreated group ($Treat = 0$) and a treated group ($Treat = 1$). Let Y denote the continuous candidate biomarker measured at baseline. We define the risk as the probability of experiencing a binary outcome D (event, i.e., death) conditional on the treatment and the biomarker, that is, $P(D = 1 | Treat, Y)$, where $D = 1$ indicates the occurrence of the event and $D = 0$ that the event does not occur. Let $\rho_0 = P(D = 1 | Treat = 0)$, and $\rho_1 = P(D = 1 | Treat = 1)$ denote the marginal probability of experiencing an event in the untreated and treated group, respectively. The difference $\rho_0 - \rho_1$ defines the marginal treatment effect.

- If $\rho_0 - \rho_1 > 0$, the “standard” strategy is to treat all the patients. Thus, the aim of using the biomarker will be to identify a subgroup of patients for which the treatment might not be beneficial.
- If $\rho_0 - \rho_1 < 0$, the “standard” strategy is to treat no one.

The absolute effect of the treatment conditionally on the biomarker Y between untreated and treated is defined as $\Delta(Y) = P(D = 1 | Treat = 0, Y) - P(D = 1 | Treat = 1, Y)$. Without using the biomarker, the treatment will be given either to all subjects or to no one. Janes et al² proposed the following logistic regression with an interaction between the treatment indicator and the biomarker to model the risk: $\text{logit}\{P(D = 1 | Treat, Y)\} = \beta_0 + \beta_1 \text{Treat} + \beta_2 Y + \beta_3 \text{Treat} \times Y$, where $\text{logit}(p) = \log\{p/(1 - p)\}$. Estimates $\hat{=} (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ of $(\beta_0, \beta_1, \beta_2, \beta_3)$ can be obtained by usual maximum-likelihood method and plug-in estimates can be defined as

$$\hat{P}(D = 1 | Treat = 0, Y) = \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_2 Y)$$

$$\hat{P}(D = 1 | Treat = 1, Y) = \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1 + (\hat{\beta}_2 + \hat{\beta}_3) Y)$$

$$\hat{\Delta}(Y) = \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_2 Y) - \text{logit}^{-1}(\hat{\beta}_0 + \hat{\beta}_1 + (\hat{\beta}_2 + \hat{\beta}_3) Y).$$

Solving, $\hat{\Delta}(Y) = 0$ with respect to Y leads to an interesting estimated cutoff value, $\hat{\gamma} = -\hat{\beta}_1/\hat{\beta}_3$. This estimates the “true” biomarker value γ , which leads to the same risk among treated and untreated patients. If the estimated value $\hat{\gamma}$ lies within the support of Y , that is, within the range of values that we can observe for the biomarker Y , then the estimated value $\hat{\gamma}$ is of interest. For instance, this suggests that patients with a biomarker $Y > \hat{\gamma}$ should be treated, and patients with a biomarker $Y \leq \hat{\gamma}$ should avoid the treatment, when $\hat{\Delta}(\cdot)$ is increasing.

To account for the uncertainty of $\hat{\gamma}$, the upper ($\hat{\gamma}_{up}$) or lower limit ($\hat{\gamma}_{low}$) of the confidence intervals (CIs) of γ can be used instead of $\hat{\gamma}$.⁴ For instance, assume that $\hat{\Delta}(\cdot)$ is increasing. If the “standard” strategy is to treat everybody, the aim of using the biomarker is to avoid unnecessary treatment, and it might be sensible to use the cutoff $\hat{\gamma}_{low}$ instead of $\hat{\gamma}$. This leads to a conservative approach that recommends not only to treat when there is sufficient evidence that the treatment benefit does not exist. Similarly, if the “standard” strategy is to not treat anybody, a conservative approach could be to use $\hat{\gamma}_{up}$, as a conservative approach which recommends to deviate from the “standard” strategy only when there is sufficient evidence to justify it.

As pointed out by Barlow,⁴ computing a CI for γ might make more sense if the null hypothesis $\mathcal{H}_0: \beta_3 = 0$ has been rejected first, as otherwise it might be that $\gamma = \infty$ and that the distribution of $\hat{\gamma}$ is degenerated. However, it is interesting to note that after rejecting $\mathcal{H}_0: \beta_3 = 0$ there is no need to account for multiple-testing to compute the CI for γ and test the corresponding null hypothesis $\mathcal{H}_0$: “ γ is not within the support of Y ,” to test whether Y is a predictive biomarker. This is because performing the second hypothesis test only when the first is significant naturally defines a gatekeeping procedure which controls the type-I error.⁵ To sum-up, an interesting hypothesis test to test whether Y is a predictive biomarker with type-I error α is defined as follows. First, test $\mathcal{H}_0: \beta_3 = 0$ with type-I error α . If the null hypothesis is not rejected, stop and conclude that you have not shown that Y is a predictive biomarker. If this is significant, compute a two-sided $100(1 - \alpha)\%$ CI for γ . If the interval is included within the range of values that the biomarker Y can take, then conclude that Y is a predictive biomarker, otherwise conclude that you have not shown that Y is a predictive biomarker.

The CI for γ can be computed from the asymptotic normality of $\hat{\gamma}$ and the estimate of the variance-covariance matrix of $\hat{\gamma}$. This leads to $\hat{\gamma} \pm z_{1-\alpha/2} \hat{\sigma}_{\hat{\gamma}}$ where

$$\hat{\sigma}_{\hat{\gamma}} = (1/\hat{\beta}_3) \sqrt{\widehat{\text{var}}(\hat{\beta}_1) + 2 \cdot \hat{\gamma} \cdot \widehat{\text{cov}}(\hat{\beta}_1, \hat{\beta}_3) + \hat{\gamma}^2 \cdot \widehat{\text{var}}(\hat{\beta}_3)}$$

follows from the delta-method, and $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -quantile of the univariate standard normal distribution.

An interesting graphical tool for descriptive analysis is the marker by treatment predictiveness curves.¹ It displays the distribution of the risk as a function of the quantile of the biomarker for each treatment group. That is, it displays the values of $P(D = 1 \mid \text{Treat} = 0, Y = y_l)$ and $P(D = 1 \mid \text{Treat} = 1, Y = y_l)$ against the biomarker percentiles $q_l = l/100$, for $l = 1, \dots, 100$, where $y_l = F_Y^{-1}(q_l)$ and $F_Y(y) = P(Y \leq y)$ denote the cumulative distribution function of Y . See Figure 1 for an example. This plot nicely illustrates the treatment effect heterogeneity, in the same way that any predictiveness curve nicely displays risk heterogeneity across patients.⁶

3 | EXTENSIONS TO SURVIVAL DATA

Suppose that the interest lies on predicting the occurrence of a 5-year mortality posttreatment. Some patients will be observed to have experienced an event within the 5 years, whereas other will be observed event free after 5 years. However, often some others might be lost of follow-up within the 5 years, which results in right censored observations. Therefore, we extend the approach of Janes et al² to time to event data with right censored data. That is, the observed data consist of $\{(\tilde{T}_i, \delta_i, \text{Treat}_i, Y_i), i = 1, \dots, n\}$, a sample of n independent copies of $(\tilde{T}, \delta, \text{Treat}, Y)$. Let T denote the failure time, C the right censored time, $\tilde{T} = \min(T, C)$ the observed failure time and $\delta = 1_{(T \leq C)}$ the failure indicator. Incorporating the time dimension into the building bricks of Janes's leads to

$$\rho_0(t) - \rho_1(t) = P(D(t) = 1 \mid \text{Treat} = 0) - P(D(t) = 1 \mid \text{Treat} = 1)$$

$$\Delta(t, Y) = P(D(t) = 1 \mid \text{Treat} = 0, Y) - P(D(t) = 1 \mid \text{Treat} = 1, Y)$$

where $D(t) = 1_{(T \leq t)}$ indicates the occurrence of event within t years. For a given time t , the value of the biomarker Y which solves $\Delta(t, Y) = 0$ defines the biomarker time-dependent threshold $\gamma(t)$. The above notations emphasize the important role of the follow-up time t which defines the time-dependent outcome $D(t)$.

Treatment recommendations to optimize the risk of outcome $D(t)$, marker by treatment predictiveness curves and other related concepts are defined as in the previous Section 2, but they are now explicitly defined as time-dependent.

3.1 | Inference using a time-dependent logistic model

We suggest to make inference using a time-dependent logistic model and inverse probability of censoring weighting (IPCW). This means that we propose to use the same logistic model as in Section 2. It leads to a direct and natural extension of the nice framework of Janes et al² to survival data. For any time $t > 0$, the model is

$$P(D(t) = 1 \mid \text{Treat}, Y) = \text{logit}^{-1}\{\beta_0(t) + \beta_1(t) \cdot \text{Treat} + \beta_2(t) \cdot Y + \beta_3(t) \cdot \text{Treat} \times Y\}. \quad (1)$$

In this Equation (1), the effect of some of the covariates may be allowed to change with time while some others could be assumed to be constant. When the coefficients $\beta_j(t)$, $j = 1, 2, 3$, are assumed to be constant with time, the model corresponds to the proportional odds model of Bennett.⁷

Several authors have previously suggested to use an IPCW approach to make inference with for this model with right censored data (Zheng et al,⁸ Scheike et al,⁹ Scheike and Zhang,¹⁰ Gron and Gerds¹¹). Following Scheike et al⁹ (eq. 3), we define the vector of estimates $\hat{\tau}(t)$ as the solutions of the estimating equations

$$\sum_{i=1}^n D^\top(t; \mathbf{w}_i) \left[\frac{\delta_i 1_{(\tilde{T}_i \leq t)}}{\hat{G}_i(\tilde{T}_i | \mathbf{w}_i)} - P(t; \mathbf{w}_i) \right] = 0, \quad (2)$$

where $P(t; \mathbf{w}_i)$ denotes the term in the right part of Equation (1), $\mathbf{w}_i = (\text{Treat}_i, Y_i)$ and $D^\top(t; \mathbf{w}_i) = \partial P(t; \mathbf{w}_i) / \partial(t)$. To account for censoring, the weight $\hat{G}_i(\tilde{T}_i | \mathbf{w}_i)$ needs to be a consistent estimate of the probability of not being censored at time $\tilde{T}_i$, given $\mathbf{w}_i$, that is, of $G(\tilde{T}_i | \mathbf{w}_i) = P(C_i > \tilde{T}_i | \mathbf{w}_i, \tilde{T}_i)$. Using such weights leads to unbiased estimating equations essentially because

$$\begin{aligned} \mathbb{E} \left\{ \frac{1_{(\tilde{T}_i \leq t)} \delta_i}{G(\tilde{T}_i | \mathbf{w}_i)} \middle| \mathbf{w}_i \right\} &= \mathbb{E} \left\{ \mathbb{E} \left\{ 1_{(T_i \leq t)} \frac{1_{(T_i \leq C_i)}}{G(T_i | \mathbf{w}_i)} \middle| d\mathbf{w}_i, T_i \right\} \middle| d\mathbf{w}_i \right\} \\ &= \mathbb{E} \left\{ \frac{1_{(T_i \leq t)}}{G(T_i | \mathbf{w}_i)} \mathbb{E} \{ 1_{(T_i \leq C_i)} \middle| d\mathbf{w}_i, T_i \} \middle| d\mathbf{w}_i \right\} \\ &= \mathbb{E} \left\{ \frac{1_{(T_i \leq t)}}{G(T_i | \mathbf{w}_i)} G(T_i | \mathbf{w}_i) \middle| d\mathbf{w}_i \right\} \\ &= P(D_i(t) = 1 | \mathbf{w}_i), \end{aligned} \quad (3)$$

where the first equation follows from the law of iterated expectation and the fact that $\delta_i = 1$ implies $\tilde{T}_i = T_i$.

If the censoring time C can be assumed independent of the treatment Treat and the biomarker Y then $P(C_i > u | \mathbf{w}_i) = P(C_i > u)$ can be estimated using the Kaplan-Meier estimator for any $u > 0$. Alternatively, a regression model for censored data (e.g. a Cox model) can be used instead.⁹ As this independent censoring assumption is often realistic and simplifies the computation, we make this assumption and rely on the consistency of the Kaplan-Meier estimator for all numerical applications presented in this manuscript.

Under weak regularity conditions, Scheike et al⁹ showed that $\hat{\tau}(t)$ is asymptotically normally distributed and provided a consistent estimator of its variance-covariance matrix. Inference procedures have been implemented in the function `comp.risk()` of the R package `timereg`.¹² Although the `comp.risk()` function and the methods of

Scheike et al⁹ were originally introduced in the setting of competing risks data, they cover our simpler setting as a special case.

From the estimate $\hat{\gamma}(t)$ and the corresponding estimate of the variance–covariance matrix, we propose to compute $\hat{\gamma}(t) = -\hat{\beta}_1(t)/\hat{\beta}_3(t)$ and the corresponding CI, as in Section 2. We also suggest to define plug-in estimates of the marker by treatment predictiveness curve and key quantities, such as $\rho_1(t)$ or $\Delta(t, Y)$, as in Section 2, by using the estimate $\hat{\gamma}(t)$.

3.2 | Inference using a proportional hazards model

We now summarize the main idea for making inference with a proportional hazards model. Most of these ideas are presented in Barlow⁴ and Janes et al³ and a software implementation is available from the **R** package **TreatmentSelection**.¹³

Let $\lambda(t; \text{Treat}, Y) = \lim_{\Delta t \rightarrow 0} P(t \leq T \leq t + \Delta t | T \geq t, \text{Treat}, Y) / \Delta t$ denote the hazard function at time t given treatment and biomarker value. Let us consider the following Cox proportional hazards model $\lambda(t; \text{Treat}, Y) = \lambda_0(t) \exp(c_1 \text{Treat} + c_2 Y + c_3 \text{Treat} \times Y)$ where $\lambda_0(t)$ is the unknown baseline hazard function and $c = (c_1, c_2, c_3)$ is the vector of regression coefficients. This model leads to $P(D(t) = 1 | \text{Treat}, Y) = 1 - \exp\{-\Lambda_0(t) \exp(c_1 \text{Treat} + c_2 Y + c_3 \text{Treat} \times Y)\}$ where $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ is the cumulative baseline hazard function. One can compute $\hat{c}$ to estimate c by maximizing the Cox partial likelihood and compute the Breslow estimate $\hat{\Lambda}_0(s)$ to estimate $\Lambda_0(s)$, for any $s > 0$. See, for example, Martinussen and Scheike.¹⁴

One can then compute the plug-in estimate $\hat{P}(D(t) = 1 | \text{Treat}, Y) = 1 - \exp\{-\hat{\Lambda}_0(t) \exp(\hat{c}_1 \text{Treat} + \hat{c}_2 Y + \hat{c}_3 \text{Treat} \times Y)\}$ and define others plug-in estimates as in Section 2. Interestingly, the solution of equation $\hat{\Delta}(t, Y) = 0$ now becomes $\hat{\gamma}(t) = -\hat{c}_1/\hat{c}_3$, which is time-independent. The proportional hazards model assumption also implies that the true threshold $\gamma(t) = -c_1/c_3$ is constant over time.

4 | SIMULATION STUDY

In this section, we conducted a simulation study to assess the performance of our method. We generated right censored data to estimate the threshold of the biomarker with the time-dependent logistic model and the proportional hazards model. Both time-varying and constant effects were simulated. Our main objectives were (a) to illustrate how the time horizon t that defines the outcome $D(t)$ can impact the predictive capability of a biomarker and (b) to evaluate the finite sample size performances of the proposed inference procedures. To estimate the coefficients of the time-dependent logistic model, we use the function **comp.risk** implemented in the **timereg** package of **R**.¹²

4.1 | Data generation

For all the simulation scenarios, the biomarker Y follows the standard normal distribution $\mathcal{N}(0, 1)$, the treatment Treat follows a Bernoulli variable with $P(\text{Treat} = 1) = 1/2$, and C is exponentially distributed with mean 20. The survival time T is generated with a conditional cumulative distribution function, given Y and Treat , which corresponds to the time-dependent logistic model of Equation (1). We considered three sample sizes (500, 1000, and 5000), and three scenarios were defined for the treatment effect $\beta_1(t)$:

1. A constant treatment effect: $\beta_1(t) = -0.29$.
2. An increasing treatment effect with time: $\beta_1(t) = -1.24$ if $t \leq 3$ and -0.31 if $t > 3$.
3. A decreasing treatment effect with time: $\beta_1(t) = 1/(1 + t)$.

In the first and second scenarios, $\beta_0(t) = \log(t/10)$, $\beta_2(t) = 0.6$, and $\beta_3(t) = -1.5$, whereas in scenario 3, $\beta_3(t) = -2$.

In this setting, an increasing (resp decreasing) treatment effect $\beta_1(t)$ will lead to an increasing threshold $\gamma(t) = -\beta_1(t)/\beta_3(t)$ (resp decreasing) because $\beta_3(t)$ was assumed constant over time and negative.

The resulting proportions of events and censored observations observed before times $t = t_1, t_2$, and t_3 that were generated are presented in Table 1. The three time horizons t_1, t_2 , and t_3 have been chosen, respectively, as the 25%, 50%, and 75% percentile of the observed failure time, $\tilde{T}$.

4.2 | Results

Table 2 summarizes the simulation results across 1000 simulated datasets, each of size $n = 500$, $n = 1000$, or $n = 5000$. We report the bias, mean squared error, coverage probabilities (CP), and average CI lengths for the estimation of $\hat{\gamma}(t)$. The Figures 2–4 display the time-dependent marker by treatment predictiveness curves for the three scenarios. For each scenario, the proposed methodology performs well, with estimates of $\hat{\gamma}(t)$ close to the true value. As expected with a constant treatment effect, for example, scenario 1 from Table 2, the estimated threshold $\hat{\gamma}(t)$ is the same for each of the three time points t . It should be pointed out that even with a large sample size of 500 and 1000, with right censoring up to 50%, the CI of the estimated threshold is large as seen in Figure 2. With time-dependent treatment effect, the proposed method performs well and accommodates the changes either in the scenario 2 or the scenario 3. Similarly with the sample size of 500 and 1000, the CI is large for all scenarios and will encompass a large spectrum of values. The CP are close to the nominal value of 95% with the largest sample size $n = 5000$, whereas they can be too conservative for the smaller sample size. Table 3 presents the estimation of the threshold when a proportional hazard model is used. For the constant treatment-effect scenario, the results are in agreement with those of our proposed approach. As expected, the misspecified Cox model provides markedly biased estimation of the time-dependent threshold with a time-varying effect. The estimate $\hat{\gamma}(t)$ is constant over time and seems to be a weighted average of the true values of $\gamma(t)$ over time.

These simulations emphasize that even with a large number of events observed, the CIs of the threshold can be large. For instance, Table 2 illustrates that with $n = 5000$, the average length of 95% CI ranges between 24% and 52% of the SD of the biomarker.

5 | APPLICATIONS

We illustrate the proposed approach using data from three randomized clinical trials Byar and Green,¹⁵ Zhu et al¹⁶ and Schlichting et al.¹⁷

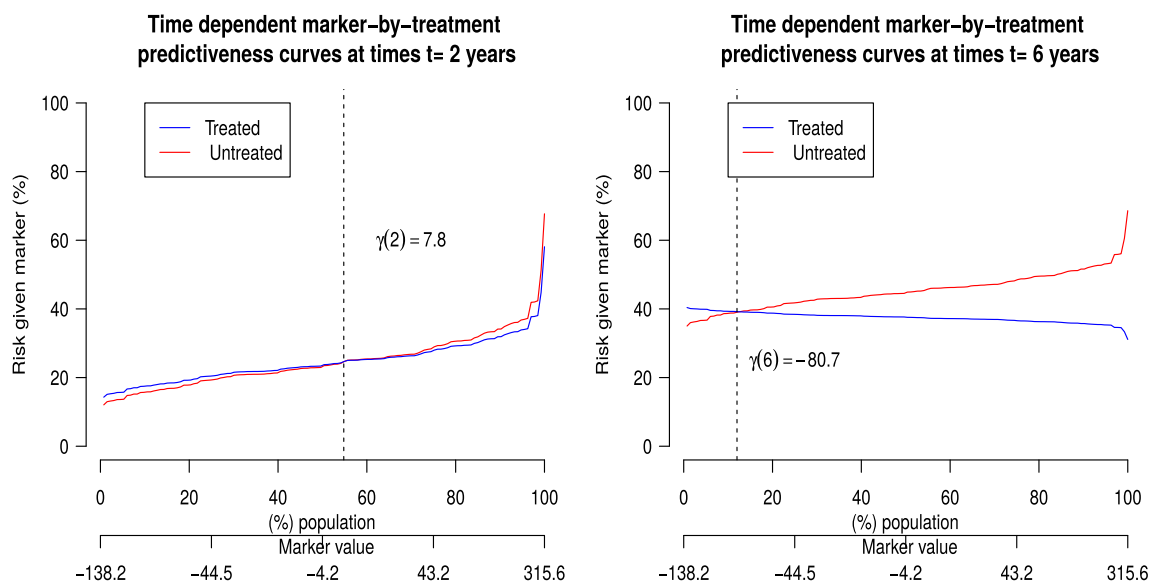


FIGURE 1 Application 2: JBR.10 data

TABLE 1 Event and censoring rates for each time horizon, for each scenario

Scenario	$P(\tilde{T} \leq t, \delta = 1)(\%)$			$P(\tilde{T} \leq t, \delta = 0)(\%)$		
	t_1	t_2	t_3	t_1	t_2	t_3
1	8.9	16.1	22.4	17.2	34.6	53.5
2	8.9	14.1	17.9	17.1	36.3	57.8
3	6.5	10.7	20.7	17.5	30.8	54.3

TABLE 2 Mean bias, mean square error (MSE), coverage probability (CP), and average confidence interval length (ACL) of the threshold $\hat{\gamma}(t)$ when $\gamma(t)$ is estimated from a time-dependant logistic regression model at several time t

Times	(t)	Bias	MSE	CP	ACL
Scenario 1 with n = 500					
$t_1 = 1.00$	-0.193	-0.015	1.178	0.993	4.253
$t_2 = 2.43$	-0.193	-0.022	0.097	0.981	1.219
$t_3 = 5.13$	-0.193	-0.008	0.106	0.978	1.274
Scenario 1 with n = 1000					
$t_1 = 1.00$	-0.193	-0.012	0.065	0.977	1
$t_2 = 2.43$	-0.193	-0.013	0.025	0.957	0.616
$t_3 = 5.13$	-0.193	-0.008	0.023	0.971	0.599
Scenario 1 with n = 5000					
$t_1 = 1.00$	-0.193	-0.001	0.009	0.949	0.362
$t_2 = 2.43$	-0.193	-0.001	0.004	0.952	0.259
$t_3 = 5.13$	-0.193	-0.001	0.004	0.957	0.247
Scenario 2 with n = 500					
$t_1 = 1$	-0.827	-0.067	1.607	0.99	5.081
$t_2 = 2$	-0.827	-0.058	0.812	0.988	3.684
$t_3 = 5$	-0.207	-0.009	0.418	0.98	2.533
Scenario 2 with n = 1000					
$t_1 = 1$	-0.827	-0.038	0.189	0.966	4.488
$t_2 = 2$	-0.827	-0.035	0.112	0.971	1.239
$t_3 = 5$	-0.207	-0.006	0.024	0.972	0.612
Scenario 2 with n = 5000					
$t_1 = 1$	-0.827	-0.013	0.017	0.968	0.515
$t_2 = 2$	-0.827	-0.006	0.009	0.955	0.375
$t_3 = 5$	-0.207	-0.001	0.004	0.959	0.257
Scenario 3 with n = 500					
$t_1 = 1.00$	-0.347	-0.016	0.069	0.985	2.306
$t_2 = 2.58$	-0.638	-0.018	0.052	0.979	0.89
$t_3 = 5.62$	-0.945	-0.021	0.077	0.977	1.09
Scenario 3 with n = 1000					
$t_1 = 1.00$	-0.347	0.010	0.038	0.982	2.206
$t_2 = 2.58$	-0.638	-0.006	0.022	0.968	0.584
$t_3 = 5.62$	-0.945	-0.013	0.024	0.964	0.602
Scenario 3 with n = 5000					
$t_1 = 1.00$	-0.347	-0.002	0.006	0.937	0.315
$t_2 = 2.58$	-0.638	-0.005	0.004	0.958	0.240
$t_3 = 5.62$	-0.945	-0.007	0.004	0.954	0.251

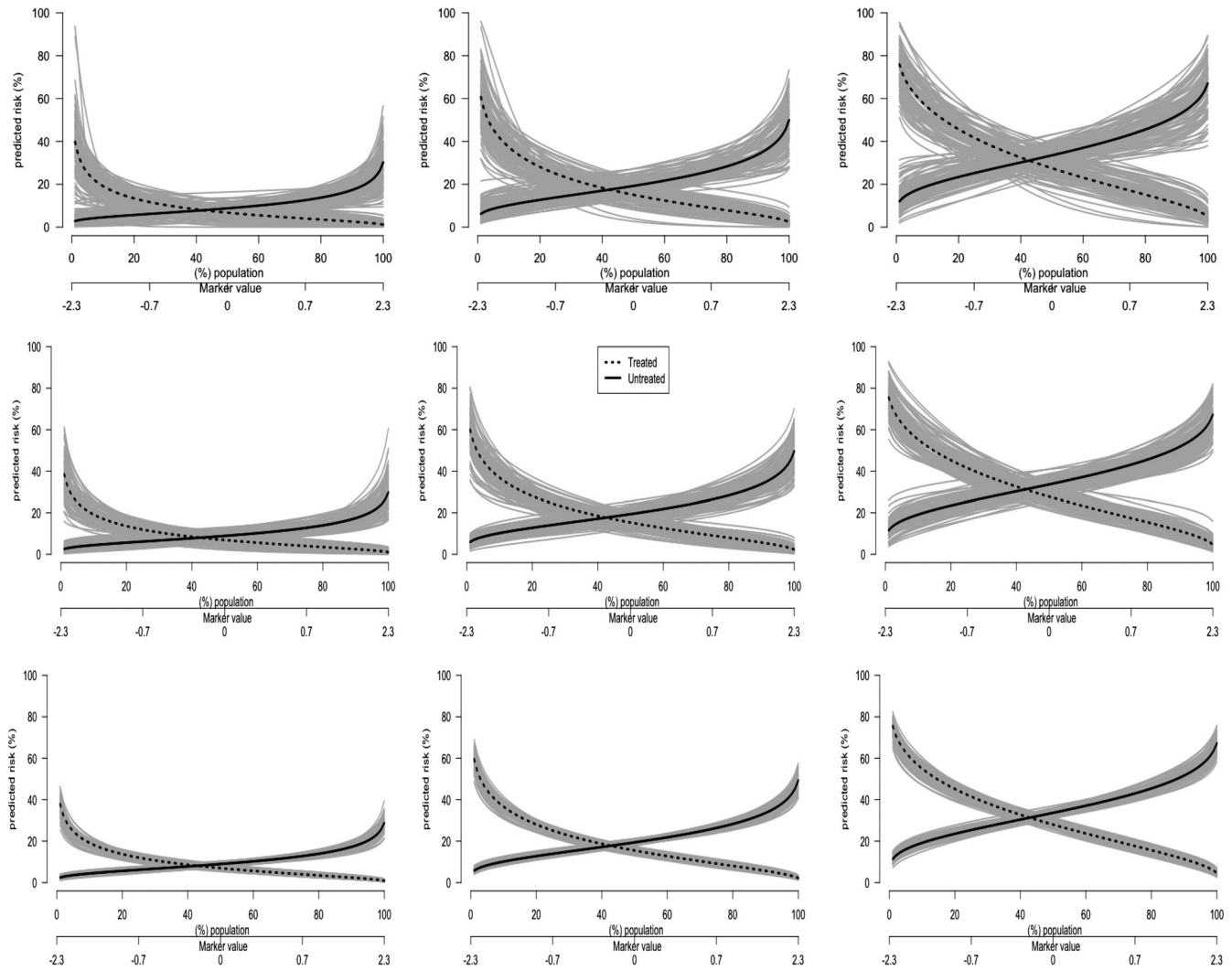


FIGURE 2 Time-dependent marker by treatment predictiveness curve at times t years for 1000 replications. For each replication, in gray, the predicted risk fitted with the time-dependent logistic model is provided. In black, the average of the predicted risks in the treated (dashed line) and untreated (full line) subjects is illustrated. We consider the sample size $n = 500$ in the curves at the top, $n = 1000$ in the curves at the middle and $n = 5000$ in the curves at the bottom for each scenarios. Scenario 1: the effect of the treatment is constant with time

5.1 | The prostate trial

The prostate trial data set comes from a randomized trial in patients with advanced prostate cancer. The data have been analyzed by Byar and Green¹⁵ and others. The primary endpoint was overall survival.

The sample includes 475 observations of which 29% are right censored. The treatment consisted of placebo and three doses of the synthetic oestrogen drug diethyl stilboestrol. For similar reasons to those given by Byar and Green,¹⁵ the placebo and the lowest dose level have been combined into a “placebo” arm ($n = 237$) and the higher doses into a “treatment” arm ($n = 238$). The biomarker that we evaluate, after log-transformation, is the serum prostatic acid phosphatase levels in King-Armstrong units. The values range from 0 to 8.7.

Our aim here is to estimate the threshold of the biomarker with $t = 14.5$ months and $t = 32$ months, which correspond to the 25% and 75% quantiles of the observed failure times. The probability of dying before 14.5 months is estimated as $\hat{\rho}_0(14.5) = 0.23$ in the placebo arm and $\hat{\rho}_1(14.5) = 0.24$ in the treatment arm. At 32 months, it is $\hat{\rho}_0(32) = 0.51 > \hat{\rho}_1(32) = 0.44$. Thus, this suggests a standard decision to not treat anybody when considering the 14.5 months outcome and to treat all patients when considering the 32 months outcome.

We now investigate whether the use of the biomarker could indicate that some people could benefit from the treatment while some others should avoid it, when considering either the 14.5 or 32 months time-horizon. The time-dependent

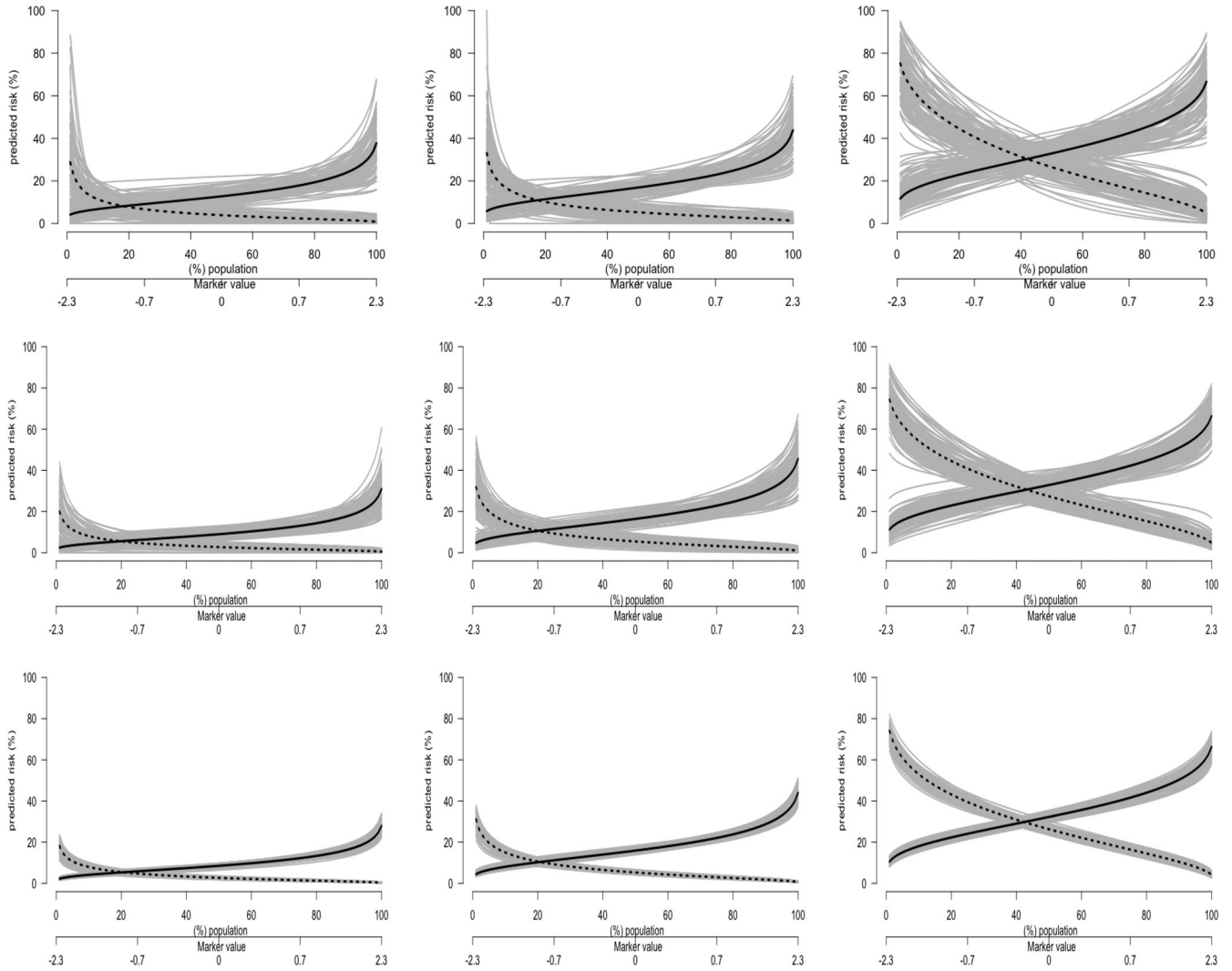


FIGURE 3 Scenario 2: the effect of the treatment is increasing with time

marker by treatment predictiveness curves is depicted in Figure 1. For time $t = 14.5$ months, the P -value of the interaction test ($\mathcal{H}_0 : \beta_3(t) = 0$) is $P = .004$, and the estimated threshold is $\hat{\gamma}(1.45) = 3.00$ (95% CI: (1.49, 4.46)). This means that the 30% of all patients who have a biomarker value above this threshold should benefit from the treatment. For $t = 32$ months, the P -value for the interaction is $P = .044$, and the estimated threshold is $\hat{\gamma}(32) = 1.38$ (95% CI: (0, 4.33)). This corresponds to the suggestion that 12% of patients should avoid the treatment.

5.2 | The JBR.10 trial

The JBR.10 is a randomized clinical trial where 133 patients in early-stage nonsmall-cell lung cancer were randomized either to an adjuvant chemotherapy ($n = 71$) or to observation only ($n = 62$). The primary endpoint was overall survival (54.5% of censoring). We aimed at evaluating the predictive capability of a malignancy-risk gene signature with values ranging from -138.2 to 315.6 . Our aim here is to estimate the threshold of the malignancy-risk gene signature at $t = 2$ years and $t = 6$ years which correspond to the 25% and 75% quantile of the failure times. The standard decision at time $t = 2$ (e.g. without the biomarker knowledge) is to treat all subjects since $\hat{\rho}_0(2) = 0.26 > \hat{\rho}_1(2) = 0.24$ and also at time $t = 6$ since $\hat{\rho}_0(6) = 0.46 > \hat{\rho}_1(6) = 0.37$. We considered the treatment and the malignancy-risk gene signature and their interaction into a time-varying effect logistic model.

The time-dependent marker by treatment predictiveness curves is depicted in the Figure 5. At time $t = 2$ years, the P -value of the interaction test is $P = .805$, and the estimated threshold is $\hat{\gamma}(2) = 7.8$ (95% CI: $[-132$ to $315]$).

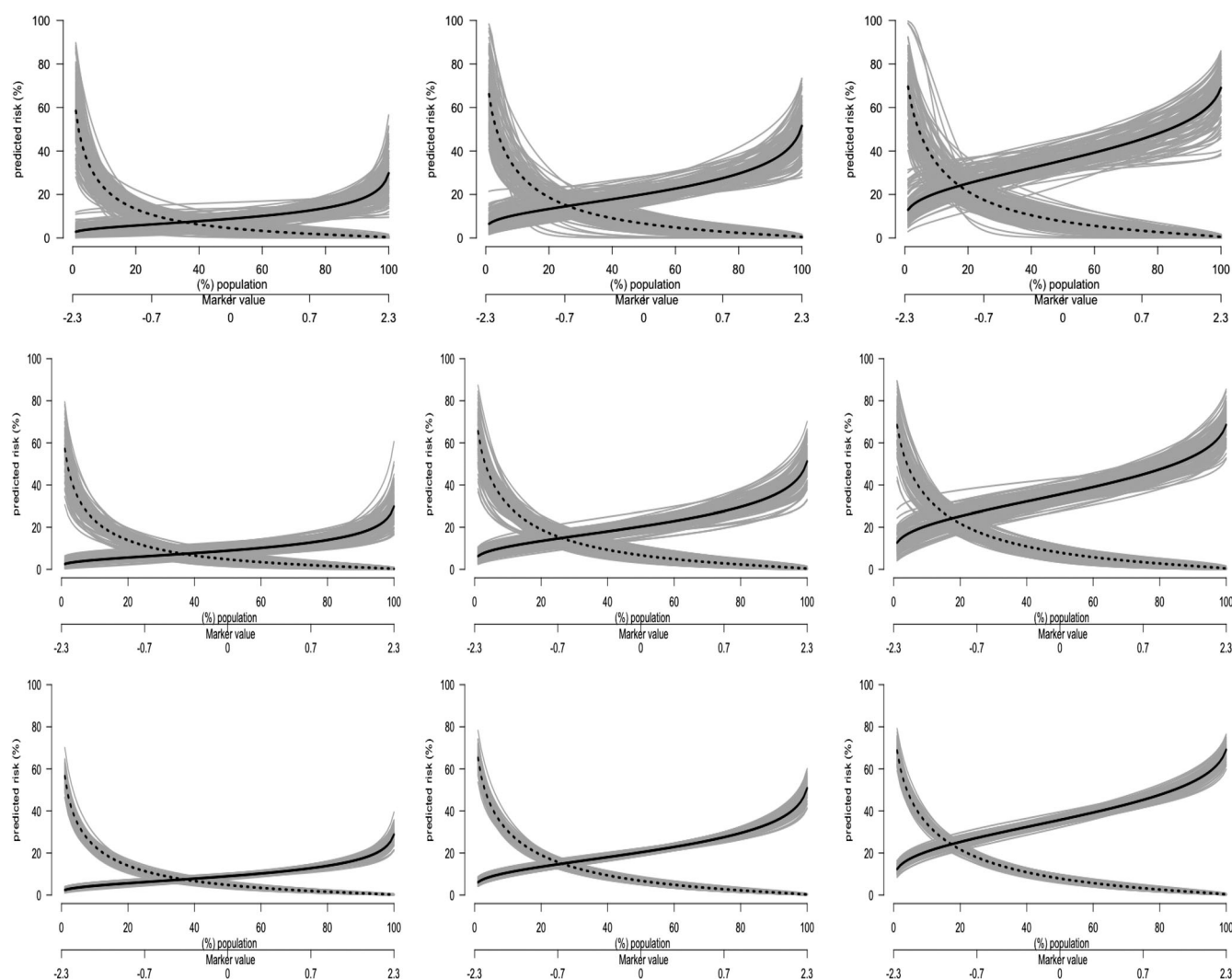


FIGURE 4 Scenario 3: the effect of the treatment is decreasing with time

		Cox model			
Times	(<i>t</i>)	Bias	MSE	CP	ACL
Scenario 1 with n = 5000					
<i>t</i> ₁ = 1.00	−0.193	−0.01	0.003	0.952	0.198
<i>t</i> ₂ = 2.43	−0.193	−0.01	0.003	0.952	0.198
<i>t</i> ₃ = 5.13	−0.193	−0.01	0.003	0.952	0.198
Scenario 2 with n = 5000					
<i>t</i> ₁ = 1	−0.827	0.472	0.238	0.043	0.477
<i>t</i> ₂ = 2	−0.827	0.472	0.238	0.043	0.477
<i>t</i> ₃ = 5	−0.207	−0.148	0.238	0.043	0.477
Scenario 3 with n = 5000					
<i>t</i> ₁ = 1.00	−0.347	−0.567	0.325	0	0.245
<i>t</i> ₂ = 2.58	−0.638	−0.276	0.325	0	0.245
<i>t</i> ₃ = 5.62	−0.945	0.031	0.325	0.891	0.245

TABLE 3 Mean bias, mean square error (MSE), coverage probability (CP), and average confidence interval length (ACL) of the threshold $\hat{\gamma}(t)$ when $\gamma(t)$ is estimated from a Cox proportionnal model at several time t

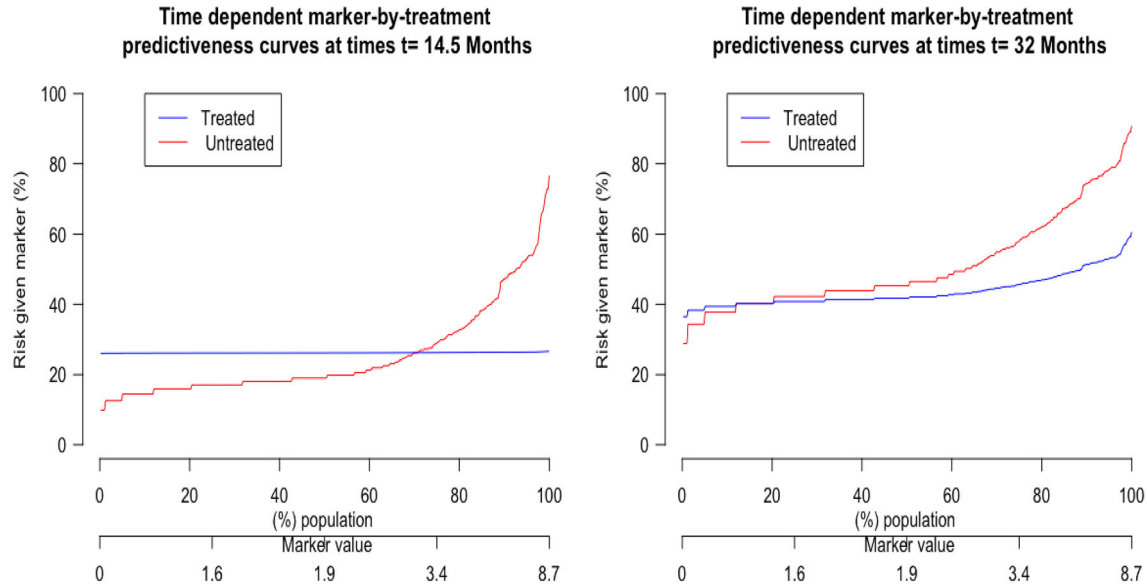


FIGURE 5 Application 1: prostate data

The interpretation of the pointwise estimate $\hat{\gamma}(2)$ indicates that 54.8% of patients must avoid the treatment. At time $t = 6$ years, the P -value for the interaction is $P = .437$ and we obtain $\hat{\gamma}(6) = -80.7$ (95% CI: $[-138$ to $332]$). Here only 12% of patients should avoid the treatment. These percentages represent the proportion of patients with biomarker value below the threshold. However, due to a small number of events, the CI covers the whole range of the gene signature.

5.3 | The CSL trial

The Copenhagen Study group for Liver diseases (CSL) trial dataset come from a randomized clinical trial where, in the period 1962 to 1969, 488 patients with liver cirrhosis were treated with either active drug prednisone (251 patients) or placebo (237 patients). The primary endpoint of the trial was the overall survival (40% of censoring). After randomization, patients were followed to either death, drop-out, or end of study (1974).

Suppose that we are interested in the predictive value of the thrombin biomarker, expressed as a percentage, with values between 12 and 135. The probability of dying before time $t = 5$ years $\hat{\rho}_0(5) = 0.57$ in the group placebo is higher than the probability of dying in the treated arm $\hat{\rho}_1(5) = 0.50$. Thus, the standard decision is to treat all patients with prednisone. We now investigate if the use of the thrombin as a predictive biomarker might save some patients from prednisone treatment considering either 2 years or 5 years as the time horizon. At time $t = 2$ years, the P -value of the interaction test is $P = .318$, and the estimated threshold is $\hat{\gamma}(2) = 61.52$ (95% CI: $[12-134]$). This corresponds to 40% of patients must avoid treatment. At 5 years, the P -value for the interaction is $P = .319$ and $\hat{\gamma}(5) = 26$ (95% CI: $[12-72.59]$) corresponding to 2% should avoid the treatment (Figure 6).

Thus, the decision to treat based on the thrombin measurement is very hazardous since its CI covers the whole range of the biomarker. This is in line with what was observed in the simulation for comparable sample size and number of observed events.

6 | DISCUSSION

In this work, we have proposed a statistical framework for evaluating the capacity of a biomarker for treatment selection with time-to-event data. Determining a biomarker threshold value beyond which treatment should be indicated is meaningful in clinical practice. In subgroups of patients with an unclear prognosis, it could help not to overtreat patients with low risk of recurrence and not to undertreat patients with high risk of recurrence. To map a continuous

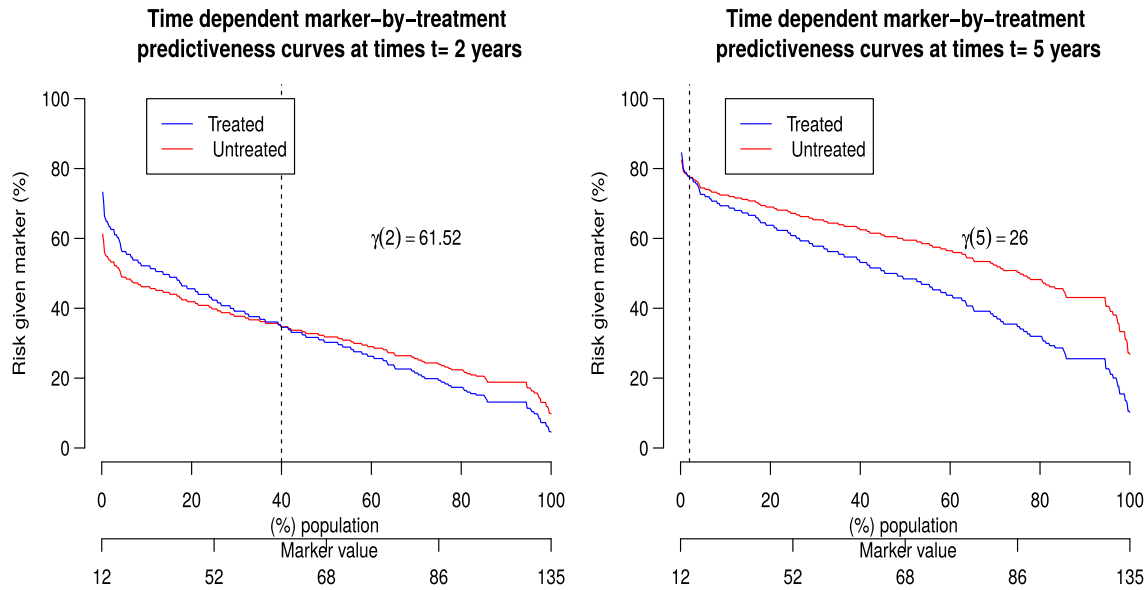


FIGURE 6 Application 3: CSL data

variable into a clinical decision, it is necessary to find a cutoff point for stratified groups requiring different types of treatments. Mean or median values of biomarker had been widely used across studies,¹⁸ but it is often desirable to determine cutoff points by using rigorous methods to obtain an optimal threshold.

We have shown that if the effect of the treatment varies with the time horizon t defining the response, then the optimal threshold of the biomarker varies with this time horizon t . This result complements the proposal of Janes et al.^{1,2} Our simulations showed that when the effect is constant over time, the time-dependent logistic model and the proportional hazards model provide similar estimations. However, the use of a time-dependent logistic model makes it possible to consider both time-dependent and constant coefficients, unlike the Cox model used by Janes et al.² As exemplified in our three case studies, we saw that the estimate of the threshold $\hat{\gamma}(t)$ is likely to be inaccurate and its CI is much too large to be useful for guiding the treatment decision. Even though some very large clinical trials assessing the clinical utility of genomic signature were conducted, they were not powered to show a significant treatment by biomarker interaction (Cardoso et al.,¹⁹ Sparano et al.).²⁰ One possible method to gain power could be to conduct a meta-analysis of individual patient data from randomized clinical trials that evaluate the predictive capacity of the same biomarker (Faron et al.,²¹ Kang et al.).²²

The estimated threshold $\hat{\gamma}(t)$ that defines the treatment rule does not directly translate into an estimate of the average clinical benefit in the population selected by the threshold. However, one could estimate this quantity by the average value of $\hat{\Delta}(t, Y_i)$ among subjects i selected when using the threshold $\hat{\gamma}(t)$. Other interesting related quantities could be estimated similarly, as previously pointed out by Janes et al.²

The method that we have proposed relies on the correct specification of the risk model being a time-dependent logistic model. If it is suspected that this model is not flexible enough, quadratic terms or splines could be introduced to model more subtle associations between the biomarker and the risk. Although this approach might reduce some bias, it would however lead to even wider CIs. A fully nonparametric approach based on a nearest-neighbor estimator,²³ similar to that of Yang et al.,²⁴ would share similar pros and cons.

We have presented methods to evaluate a single biomarker without assuming that other relevant predictor variables are observed. However, sometimes several candidate predictor variables are available for the analysis. In that case, other methods can be used instead of those presented in this manuscript. We refer Lipkovich et al.,²⁵ Janes et al.³ and references therein for interesting approaches for this context.

As pointed out in Lipkovich,²⁵ sometimes one may prefer to solve the equation $\hat{\Delta}(t, Y) = \eta$, with $\eta > 0$, rather than $\hat{\Delta}(t, Y) = 0$ as assumed in this manuscript. This is because using $\eta > 0$ could enable us to define a treatment benefit which is large enough to compensate, for example, its cost and side effects. This would require a related but different approach, for which a direct modeling of the risk difference $\Delta(t, Y)$ seems more appropriate than the logistic model that we have used. This is because the logistic model models odds ratios, which are sufficient to know

whether risk differences are positive or negative, but not their value (see, for example, Mehrotra²⁶). Hence, the method introduced by Richardson et al.²⁷ could be appealing to model the risk difference $\Delta(t, Y)$ when the interest lies in solving $\hat{\Delta}(t, Y) = \eta$. However, how to extend the method to handle right censored data has not been clarified yet.

Finally, extension of the proposed methodology to observational data could be feasible by employing inverse probability of treatment weighting, under usual assumptions for causal inference. With a binary endpoint, a method to estimate the probability of being a responder given a set of baseline covariates, using the potential outcomes framework, was recently used by Porcher et al.²⁸ As they (nicely) pointed out, the time-dependent logistic model that we have used could handle right censored endpoint in their context too.

ACKNOWLEDGEMENTS

B.M. would like to thank the “GDR Statistiques & Santé” for funding his internship.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Bassirou Mboup  <https://orcid.org/0000-0002-1156-9134>

Paul Blanche  <https://orcid.org/0000-0003-1415-7976>

Aurélien Latouche  <https://orcid.org/0000-0001-9218-0333>

REFERENCES

- Janes H, Pepe MS, Bossuyt PM, Barlow WE. Measuring the performance of markers for guiding treatment decisions. *Ann Intern Med*. 2011;154(4):253-259.
- Janes H, Brown MD, Huang Y, Pepe MS. An approach to evaluating and comparing biomarkers for patient treatment selection. *Int J Biostat*. 2014;10(1):99-121.
- Janes H, Brown MD, Crager MR, Miller DP, Barlow WE. Adjusting for covariates in evaluating markers for selecting treatment, with application to guiding chemotherapy for treating estrogen-receptor-positive, node-positive breast cancer. *Contemp Clin Trials*. 2017;63:30-39.
- Barlow WE. Design of a Clinical Trial for Testing the Ability of a Continuous Marker to Predict Therapy Benefit. In: Crowley J, Hoering A, Ankerst D, eds. *Handbook of Statistics in Clinical Oncology*. Washington, United States: Chapman and Hall/CRC Press; 2005.
- Dmitrienko A, Tamhane AC. Gatekeeping procedures with clinical trial applications. *Pharmaceutical Stat*. 2007;6(3):171-180.
- Pepe MS, Feng Z, Huang Y, et al. Integrating the predictiveness of a marker with its performance as a classifier. *Am J Epidemiol*. 2008;167(3):362-368.
- Bennett S. Analysis of survival data by the proportional odds model. *Stat Med*. 1983;2(2):273-277.
- Zheng Y, Cai T, Feng Z. Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. *Biometrics*. 2006;62(1):279-287.
- Scheike TH, Zhang MJ, Gerds TA. Predicting cumulative incidence probability by direct binomial regression. *Biometrika*. 2008;95(1):205-220.
- Scheike TH, Zhang MJ. Flexible competing risks regression modeling and goodness-of-fit. *Lifetime Data Anal*. 2008;14(4):464-483.
- Grøn R, Gerds TA. Binomial Regression Models. In: Klein JP, van Houwelingen HC, Ibrahim JG, Scheike TH, eds. *Handbook of Survival Analysis*. USA: CRC Press; 2013.
- Scheike TH, Zhang MJ. Analyzing competing risk data using the R timereg package. *J Stat Softw*. 2011;38(2):1-15.
- Janes H, Brown M. *TreatmentSelection: Evaluate Treatment Selection Biomarkers*. Vol 1. R Package Version 2. New York: De Gruyter; 2017:1.
- Martinussen T, Scheike TH. *Dynamic Regression Models for Survival Data*. New York, NY: Springer; 2006.
- Byar D, Green S. The choice of treatment for cancer patients based on covariate information. *Bull Cancer*. 1980;67(4):477-490.
- Zhu CQ, Ding K, Strumpf D, et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol*. 2010;28(29):4417-4424.
- Schlichting P, Christensen E, Andersen PK, et al. Prognostic factors in cirrhosis identified by Cox's regression model. *Hepatology*. 1983;3(6):889-895.
- Chen DT, Hsu YL, Fulp WJ, et al. Prognostic and predictive value of a malignancy-risk gene signature in early-stage non-small cell lung cancer. *J Natl Cancer Inst*. 2011;103(24):1859-1870.
- Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol*. 2008;26(5):729-735.
- Sparano JA. TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer*. 2006;7(4):347-350.

21. Faron M, Pignon JP, Paoletti X. Meta-analysis of prognostic and predictive factors: towards individual participant data? *Eur J Cancer*. 2018;104:224-226.
22. Kang C, Janes H, Tajik P, et al. Evaluation of biomarkers for treatment selection using individual participant data from multiple clinical trials. *Stat Med*. 2018;37(9):1439-1453.
23. Akritas MG. Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann Stat*. 1994;22:1299-1327.
24. Yang H, Tang R, Hale M, Huang J. A visualization method measuring the performance of biomarkers for guiding treatment decisions. *Pharm Stat*. 2016;15(2):152-164.
25. Lipkovich I, Dmitrienko A. Others. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Stat Med*. 2017;36(1):136-196.
26. Mehrotra DV. Stratification issues with binary endpoints. *Drug Inf J*. 2001;35:1343-1350.
27. Richardson T, Robins J, Wang L. On modeling and estimation for the relative risk and risk difference. *J Am Stat Assoc*. 2017;112(519):1121-1130.
28. Porcher R, Jacot J, Wunder JS, Biau DJ. Identifying treatment responders using counterfactual modeling and potential outcomes. *Stat Methods Med Res*. 2019;28:3346-3362.

How to cite this article: Mboup B, Blanche P, Latouche A. On evaluating how well a biomarker can predict treatment response with survival data. *Pharmaceutical Statistics*. 2020;1–14. <https://doi.org/10.1002/pst.2002>

3.3 Compléments

3.3.1 Une amélioration des intervalles de confiance

Une amélioration des intervalles de confiance dans les jeux de données réelles pour la cirrhose du foie et pour le cancer du poumon consistait à étendre la taille de l'échantillon.

3.3.2 Data extension

Les applications ont révélé de larges intervalles de confiance pour le seuil estimé du biomarqueur, car la taille de l'échantillon est petite pour détecter une interaction. Nous avons donc décidé d'augmenter le nombre d'événements (en augmentant la taille de l'échantillon à $n = 5000$) pour avoir des intervalles de confiance raisonnables.

Nous nous intéressons au délai observé, l'indicateur de censure, un indicateur de traitement binaire et un biomarqueur continu. La méthode utilisée pour générer chaque nouvelle observation est décrite ci-dessous :

- la valeur du biomarqueur est générée en inversant la fonction de distribution cumulative empirique. C'est-à-dire que chaque nouvelle valeur est tirée parmi les valeurs observées du biomarqueur avec équiprobabilité.
- le traitement randomisé est généré à l'aide d'une expérience de Bernoulli avec une probabilité de 0.5
- pour générer le temps d'événement : puisque notre objectif principal est la prédiction à un horizon donné, nous considérons que la stratégie suivante permet de générer le temps d'événement observé,
 - D'abord , nous fixons deux horizons de prédictions $t_1 < t_2$ et calculons les risques prédits en utilisant le modèle logistique dépendant du temps ajusté en fonction du traitement et du biomarqueur précédemment générés $risk(t_1), risk(t_2)$ i.e $\hat{P}(T \leq t | Treat, Y)$
 - Ensuite, une interpolation linéaire est utilisée pour fournir $risk(t)$ pour tout $t_1 < t < t_2$, i.e
$$risk(t) = risk(t_1) + (t - t_1) \frac{risk(t_2) - risk(t_1)}{t_2 - t_1}$$
 - Enfin, les temps d'événement t sont obtenus en résolvant numériquement l'équation $risk(t) = U$ avec $U \sim \text{Uniform}(0, 1)$ en utilisant la fonction `R uniroot`.
- pour générer le temps de censure :
 - nous sélectionnons et ordonnons les temps de censure à partir de la série de données réelles sans doublons et inversons la fonction de répartition estimée du temps

censuré pour générer le temps de censure équivalent à tirer le temps censuré, parmi ceux observés dans l'ensemble des données réelles, avec une probabilité égale au saut de l'estimateur de Kaplan-Meier de la distribution de censure.

- l'estimateur de Kaplan-Meier n'atteignant pas toujours zéro, nous avons ajouté le temps maximum observé (c'est-à-dire au-delà du dernier temps observé dans l'ensemble de données réel), qui est tiré avec une probabilité de un moins la somme des sauts.

Vous trouverez dans l'annexe A une illustration étape par étape de cette section avec du code R.

3.3.3 Les résultats obtenus

Pour les données JBR.10, nous avons considéré $t_1 = 2$ années et $t_2 = 6$ années. Le seuil estimé à 2 ans était de $\hat{\gamma}(2) = 40.4$ (95%CI : $[-66.1; 135.1]$) et à 6 ans, $\hat{\gamma}(6) = -60.1$ (95%CI : $[-106.4; -8.5]$).

Pour les données CSL, nous avons considéré $t_1 = 2$ ans et $t_2 = 5$ ans. Le seuil estimé à 2 ans était de 64.35 (95%IC : $[41.7; 86.6]$) et à 5 ans de 38.9 (95%IC : $[13.9; 69.5]$).

Cela montre que les capacités prédictives d'un biomarqueur ne peuvent être évaluées que dans le cadre de très grands essais cliniques afin de fournir des seuils et des intervalles de confiance cliniquement utiles. Ces résultats sont cohérents avec les développements de [Janes et al., 2015] sur le calcul du nombre de sujets nécessaire pour établir la capacité prédictive d'un biomarqueur.

Bibliographie

- Janes, H., Brown, M. D., Huang, Y., and Pepe, M. S. (2014). An approach to evaluating and comparing biomarkers for patient treatment selection. *Int J Biostat*, 10(1) :99–121. 24, 25
- Janes, H., Brown, M. D., and Pepe, M. S. (2015). Designing a study to evaluate the benefit of a biomarker for selecting patient treatment. *Stat Med*, 34(27) :3503–3515. 42
- Scheike, T. H., Zhang, M. J., and Gerds, T. A. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95(1) :205–220. 25
- Zheng, Y., Cai, T., and Feng, Z. (2006). Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. *Biometrics*, 62(1) :279–287. 24

Chapitre 4

Test de l'hypothèse de calibration d'un modèle de prédiction avec des données censurées

Sommaire

4.1	Introduction	46
4.2	Statistique de Test	47
4.3	Simulations	48
4.3.1	Performance du test T_B	48
4.3.2	Influence de la mauvaise calibration sur le seuil du biomarqueur	49
4.4	Application	50

4.1 Introduction

Un élément clé de la médecine de précision est de disposer d'un modèle prédictif "performant". En effet, il permet de déterminer le risque c'est à dire la probabilité qu'a un patient de subir un événement au cour d'une période de temps donnée en fonction de ses caractéristiques et permettent d'aider les cliniciens à prendre des décisions de traitement. La calibration est une dimension importantes dans l'évaluation des performances des modèles de prédiction car un modèle mal calibré fournit un risque prédit invalide et peut introduire des erreurs dans la prise de décision [Gail and Pfeiffer, 2005; Pepe and Janes, 2013]. On dit qu'un modèle est calibré pour un risque d'événement au sens faible si, pour tout pourcentage x , on peut s'attendre à ce que x patients sur 100 subissent l'événement parmi les patients qui ont un risque prédit par le modèle de $x\%$. Plusieurs études ont été menées sur la calibration d'un modèle de prédiction de risque. La plupart des études proposent une évaluation graphique plutôt qu'un test de l'hypothèse de calibration. En effet, Gerds et al. [2014] proposent une courbe de calibration pour risque concurrents fondée sur les pseudo-valeurs. Blanche et al. [2019] proposent un outil graphique appelé "Wally plots" pour évaluer la calibration des modèles. Austin et al. [2020] proposent une méthode graphique ainsi qu'une métrique permettant de mesurer l'écart de risque observé et prédit au niveau de certains quantiles.

Certains auteurs proposent des tests de calibration pour la quantification. La méthode classique dans le cas non censuré est le test de Hosmer-Lemeshow qui a comme défaut principal de dépendre fortement de la taille de l'échantillon. En effet le test est presque certainement significatif si l'échantillon est suffisamment important et non significatif si n est suffisamment petit. Il est à noter que de nombreux auteurs ont tenté d'étendre cette statistique de test pour des données de survie [Crowson et al., 2016; D'agostino and Nam, 2003; Demler et al., 2015]. En effet, Demler et al. [2015] proposent un test fondé sur la statistique de test proposée par D'agostino and Nam [2003] mais il est limité à un taux de censure à droite allant de 40% à 50%. Ces taux de censures ne sont pas compatible avec les domaines d'applications où l'on observe 80 voire 90% de censure, notamment pour le cancer du sein et prostate.

Par ailleurs, la méthode d'estimation du seuil d'un biomarqueur prédictif que nous avons proposée repose sur l'hypothèse forte les modèles de prédictions de risque sont bien calibrés [Mboup et al., 2020]. Il nous a paru intéressant d'étudier l'influence d'un modèle mal calibré sur la détermination du seuil d'un biomarqueur. En d'autres termes, la mauvaise calibration influence-t-elle le seuil du biomarqueur estimé ?

Nous proposons dans ce chapitre un test de calibration avec un critère de jugement censuré dont l'estimation du risque observé est fondée par la pondération par l'inverse de la probabilité de censure (IPCW) pour aller au-delà de 70% de censure. Dans ce contexte, nous allons

réestimer le seuil du biomarqueur pour répondre à la question ci-dessus.

Nous allons d'abord présenter la statistique de test dans la section 4.2. Ensuite, une étude de simulation sera réalisée à la section 4.3. Ce travail est en cours, nous terminons par les perspectives de l'application à la section 4.4.

4.2 Statistique de Test

Soient T le temps d'événement, C le temps de censure, X est le vecteur des covariables (Age, traitement, Biomarqueur, ect...), $\tilde{T} = \min(T, C)$ le temps d'événement observé, $\delta = \mathbb{I}_{(T \leq C)}$ l'indicateur de survenue d'événement et $D(t) = \mathbb{I}_{(T \leq t)}$ l'indicateur d'événement à l'horizon de prédiction t qui n'est pas toujours observé à cause de la censure. On suppose que le risque $\text{risk}(t, X)$ a été préalablement estimé en utilisant un autre échantillon indépendant de celui employé pour le test. Nous observons alors l'ensemble de données suivante $\{\tilde{D}_i(t), \tilde{T}_i, \delta_i, \text{risk}_i(t, X_i)\}$ pour tout $i = 1, \dots, n$. avec $\tilde{D}_i(t) = \mathbb{I}(\tilde{T}_i \leq t, \delta_i = 1)$. On note

$$\begin{cases} V_i(t) = 0 \text{ si } \tilde{T}_i < t \text{ et } \delta_i = 0 \\ V_i(t) = 1 \text{ sinon} \end{cases}$$

à cause de la censure $\tilde{D}_i(t) = \mathbb{I}(T_i \leq t)$ que si $V_i(t) = 0$ car $\tilde{D}_i(t) = (1 - V_i(t))D_i(t)$.

Un modèle est bien calibré au sens faible si

$$P(\tilde{D}(t) = 1 | \text{risk}(t, X) = r) \approx r \quad \forall r \in [0, 1] \quad (4.1)$$

Les hypothèses testées sont :

(H_0) $P(\tilde{D}(t) = 1 | \text{risk}(t, X) = r) = r$ contre (H_1) $P(\tilde{D}(t) = 1 | \text{risk}(t, X) = r) \neq r$.

La statistique de test que nous proposons, T_B , ressemble à celle proposée par D'agostino and Nam [2003], DG , qui estime le risque observé par un estimateur non paramétrique classique de Kaplan-Meier ($KM_g(t)$) mais ne tolère pas une censure à droite de plus de 15% [Demler et al., 2015].

$$DG = \sum_{g=1}^G \frac{n_g \left(KM_g(t) - \overline{\text{risk}(t, X)}_g \right)^2}{\overline{\text{risk}(t, X)}_g \left(1 - \overline{\text{risk}(t, X)}_g \right)}.$$

Nous proposons d'estimer le risque observé $Obs(t)$ en adaptant l'idée de l'approche de pondération par l'inverse de la probabilité de censure (IPCW).

$$T_B = \sum_{g=1}^G \frac{n_g \left(\overline{Obs(t)}_g - \overline{\text{risk}(t, X)}_g \right)^2}{\overline{\text{risk}(t, X)}_g \left(1 - \overline{\text{risk}(t, X)}_g \right)} \quad (4.2)$$

admet, sous H_0 , une loi asymptotique $\chi^2(G - 1)$ où $\overline{Obs(t)}_g = \hat{P}(\tilde{D}(t) = 1 | \text{risk}(t, X)_g)$ et $G = 10$. En partant de l'équation 4.1, l'idée consiste à diviser l'échantillon en différents

groupe de patients g ayant des risques prédits au temps t similaire et à comparer les risques moyens $\overline{\text{risk}(t, X)}_g$ dans chaque groupe aux risques d'événement observés conditionnelle à la moyenne des risques.

En reformulant $T_B = \sum_{g=1}^G \frac{(n_g \overline{\text{Obs}(t)}_g - n_g \overline{\text{risk}(t, X)}_g)^2}{n_g \overline{\text{risk}(t, X)}_g (1 - \overline{\text{risk}(t, X)}_g)}$ et en remarquant que sous l'hypothèse H_0 , $(n_g \overline{\text{Obs}(t)}_g - n_g \overline{\text{risk}(t, X)}_g) / \sqrt{n_g \overline{\text{risk}(t, X)}_g (1 - \overline{\text{risk}(t, X)}_g)}$ est la version centrée réduite de la variable aléatoire $n_g \overline{\text{Obs}(t)}_g$. Or, la loi marginale de $n_g \overline{\text{Obs}(t)}_g$ est la loi Binomiale $\mathcal{B}(n_g, \overline{\text{risk}(t, X)}_g)$. En effet, posons $p_g = \overline{\text{risk}(t, X)}_g$ et $N = (N_1 = n_1 \overline{\text{Obs}(t)}_1, N_2 = n_2 \overline{\text{Obs}(t)}_2, \dots, N_G = n_G \overline{\text{Obs}(t)}_G)$ un vecteur aléatoire répartis en G groupes de proportion p_g , $g = 1$ à G . Par conséquent, la loi de N est la loi multinomiale à $G - 1$ paramètres car $\sum_{g=1}^G p_g = 1$. Soit $T_g = \frac{N_g - n_g p_g}{\sqrt{n_g p_g (1 - p_g)}}$, $g = 1$ à G , les variables aléatoire, T_g , sont centrées et réduite. Asymptotiquement, cette variable aléatoire converge vers une loi $\mathcal{N}(0, 1)$ et son carré une loi $\chi^2(1)$. Alors la statistique $\sum_{g=1}^G T_g^2 = T_B$ suit asymptotiquement une loi $\chi^2(G - 1)$.

4.3 Simulations

Nous avons mené une étude de simulation pour évaluer les performances du test T_B . Pour générer les données, nous supposons que $\text{risk}(t, X)$ a été estimé sur un autre modèle de risque indépendamment de notre modèle. Nous avons simulés des données censurés à partir d'un modèle de Cox comportant 2 prédicteurs $X = (Y, Z)$ où Y est le biomarqueur continu et le traitement Z est binaire. La variable Z a été généré à partir d'une binomiale de paramètre $p = 0.5$, la variable Y suit une normale centré réduite, C le temps de censure suit une loi exponentiel de paramètre α , $\tilde{T} = \min(T, C)$ le temps de suivi observé, $\delta = \mathbb{I}_{\{T \leq C\}}$ l'indicatrice de survenue de l'évènement. Les coefficients $\beta_1 = \log(3)$, $\beta_2 = \log(0.5)$ représentent respectivement l'effet Y , l'effet de la variable Z .

4.3.1 Performance du test T_B

Le risque prédit est estimé par la relation suivante

$$\widehat{\text{risk}}(t, X) = 1 - \exp \left\{ -\widehat{\lambda}_0(t) \exp(\widehat{\beta}X) \right\}$$

où posant $\widehat{\lambda}_0(t) = at$ est la fonction de risque de base, $a = 0.1$ une constante et $\widehat{\beta} = (\beta_1, \beta_2)$. Nous estimons le risque observé $\text{Obs}_g(\cdot)$ en utilisant le modèle logistic dépendant du temps dont l'estimation repose sur une pondération par l'inverse des probabilités de censure. Nous obtenons alors

$$\widehat{\text{Obs}}_g(t) = P(\tilde{D}(t) = 1 | \widehat{\text{risk}}_g(t, X)) = \text{logit}^{-1}(\alpha_0(t) + \alpha_1 \widehat{\text{risk}}_g(t, X))$$

Nous avons considéré une taille d'échantillon de $n = 1000$ et un taux de censure supérieur à 60%. L'erreur de type 1 a été évaluée au niveau de signification de 0.05 en utilisant 1000 réplifications de simulation et en supposant que le modèle est bien calibré. Pour estimer la valeur T_B et la p -valeur du test, nous avons utilisé ces réplifications, qui ont fourni une précision suffisante pour le seuil de signification de 0.05. En effet, la Figure 4.1 est une illustration de la distribution de la statistique de test T_B . Nous avons conclu que le test T_B proposé permet de contrôler l'erreur de type 1 et la distribution de la statistique de test est la même que celle d'un χ^2 à $G - 1$ ddl.

Pour évaluer la puissance du test des tailles d'échantillon de $n = 600, 1000$ et 5000 ont été considérées et la puissance de T_B est comparé avec celle du test suivant proposé par [Demler et al., 2015]

$$DM = \sum_{g=1}^G \frac{\left(KM_g(t) - \overline{\text{risk}}_g(t, X)\right)^2}{\text{var}(KM_g(t))}$$

où $\text{var}(KM_g(t))$ est la variance de survie de Greenwood. Cependant, nous considérons que le risque

$$\widehat{\text{risk}}(t, X)_\omega = (1 - \omega)\widehat{\text{risk}}(t, X) + \omega\widehat{\text{risk}}(t, X)^2$$

où $\omega \in [0, 1]$ est un paramètre de réglage pour obtenir un modèle mal calibré. Lorsque $\omega > 0$, la prédiction du risque surestime le risque réel et vice versa. Le taux de censure est fixé à 65%. Nous avons généré 10 valeurs de ω comprise entre 0 et 1. Pour une valeur ω , nous avons répété $B = 1000$ fois et compté le nombre de fois que l'hypothèse nul est rejeté (p -value ≤ 0.05) pour les deux tests T_B et DM . Les résultats de la simulation ont démontré que le test T_B était plus puissant que le test DM pour détecter un modèle mal calibré avec des taux de censure dépassant 60% (voir Figure 4.2).

4.3.2 Influence de la mauvaise calibration sur le seuil du biomarqueur

Dans un premier temps, nous générons les données ci-dessus avec $n = 1000$ et $\beta_3 = 1.5$ l'interaction entre Y le biomarqueur et Z le traitement avec un taux de censure à 80%. Nous réalisons notre test pour assurer que notre modèle est bien calibré. Le risque prédit est estimé avec un modèle de Cox à l'horizon de 4 ans. Nous ne rejetons pas l'hypothèse nulle et le seuil de Y estimé est de $-0.36 [-0.39, -0.33]$. Le seuil trouvé est optimal car il est dans les valeurs du support du biomarqueur et permet d'attribuer ou non le traitement.

Dans un second temps, nous générons les données sans interaction entre Y et Z avec le même taux de censure 80%. Cependant, pour estimer le seuil du biomarqueur sous l'hypothèse d'un modèle mal calibré, nous avons ajouté l'interaction entre Y et Z dans le modèle pour

estimer le risque à l'horizon de 4 ans. En effet, nous réalisons notre test T_B de calibration pour assurer que le modèle est mal calibré. Nous trouvons $T_B = 31.284$ ($p = 0.000264 < 0.05$). Le test est très significatif alors nous rejetons l'hypothèse nul d'où le modèle est mal calibré. Par conséquent, nous estimons le seuil du biomarqueur ainsi que son intervalle de confiance sous mauvaise calibration. Le seuil de Y trouvé est de $-3.47 [-4.07, -3.19]$. Une valeur qui n'appartient pas dans le support de Y . Ce seuil n'est pas optimal.

Par cette approche nous illustrons qu'un modèle de prédiction mal calibré influe sur la détermination du seuil optimal du biomarqueur pour prendre une décision de traiter ou pas.

4.4 Application

Ce travail étant en cours, nous envisageons d'appliquer notre test sur les données de cancer de la prostate du chapitre 3. Les données sur la prostate provient d'un essai randomisé sur des patients atteints d'un cancer avancé de la prostate. Les données ont été analysées par Byar and Green [1980], entre autres. Le critère d'évaluation principal était la survie globale. L'échantillon comprend 475 observations et 15 variables. Le traitement consistait en un placebo et trois doses de diéthyl stilboestrol, un œstrogène synthétique. Le placebo et la dose la plus faible ont été combinés dans un groupe "placebo" ($n = 237$) et les doses plus élevées dans un groupe "traitement" ($n = 238$). Le biomarqueur est le taux de phosphatase acide prostatique sérique dans les unités King-Armstrong. Les valeurs du biomarqueur après transformation logarithmique sont comprise entre 0 et 8.7. Dans un premier temps nous allons construire un modèle pronostique sur un sous-échantillon de l'essai. Puis, nous effectuerons le test de calibration sur le reste de l'échantillon. Nous comparerons notre test à une méthode graphique.

Distribution of statistic test under H0

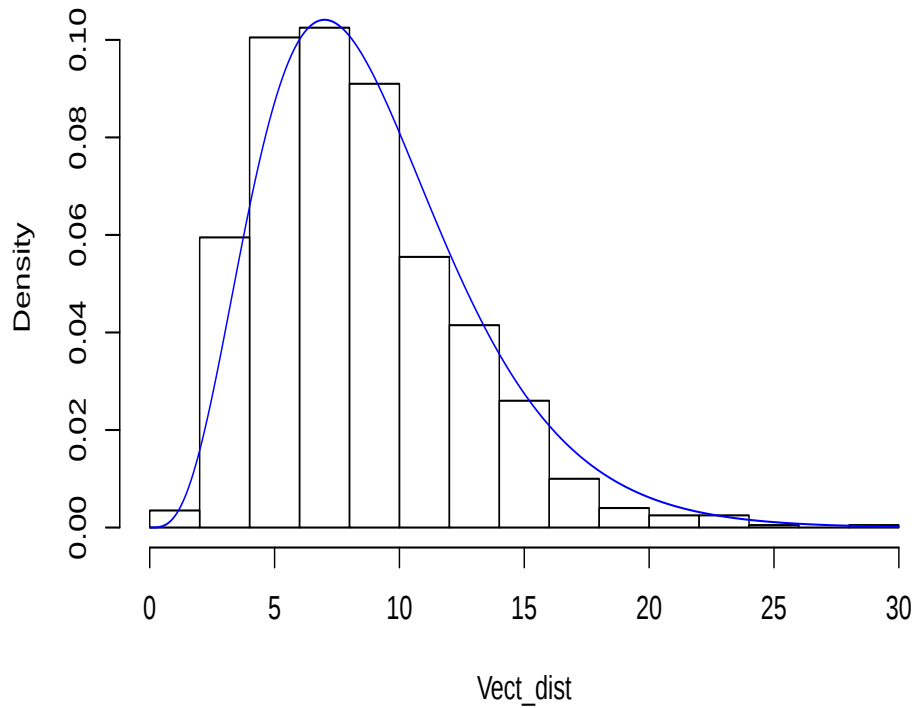


FIGURE 4.1 – Distribution de la statistique de test sous l’hypothèse nulle, en Bleu la distribution de la loi théorique du χ^2 à 9 ddl.

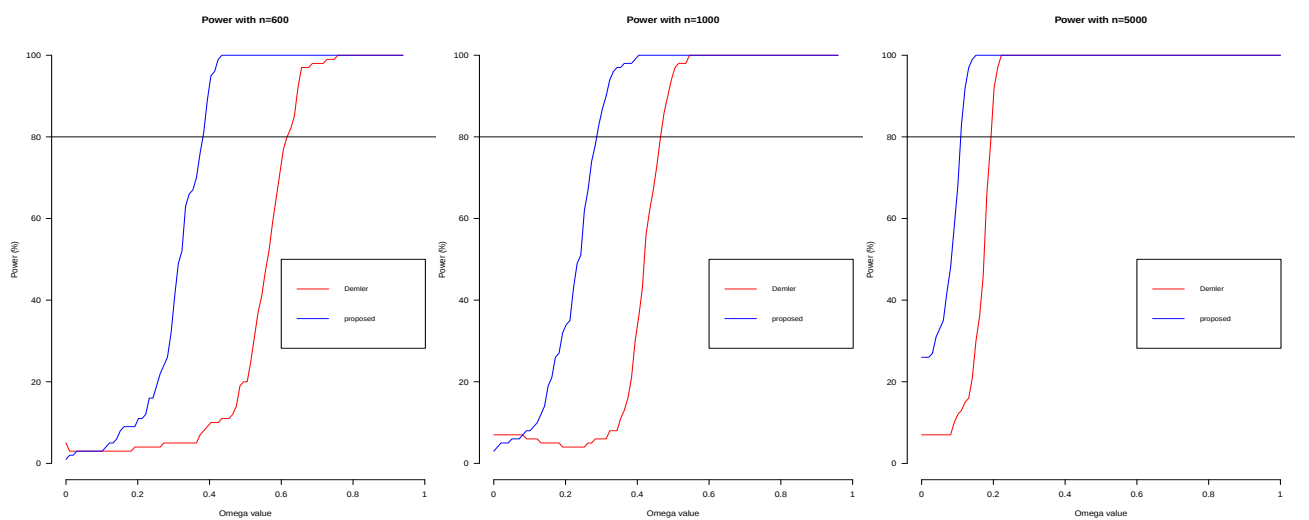


FIGURE 4.2 – Puissance des tests sous le risque mal spécifié sur une taille d’échantillon respective $n = 600$, $n = 1000$ et $n = 5000$

Bibliographie

- Austin, P. C., Harrell Jr, F. E., and van Klaveren, D. (2020). Graphical calibration curves and the integrated calibration index (ici) for survival models. *Statistics in Medicine*. 46
- Blanche, P., Gerds, T. A., and Ekstrøm, C. T. (2019). The wally plot approach to assess the calibration of clinical prediction models. *Lifetime data analysis*, 25(1) :150–167. 46
- Byar, D. and Green, S. (1980). The choice of treatment for cancer patients based on covariate information. *Bulletin du cancer*, 67(4) :477–490. 50
- Crowson, C. S., Atkinson, E. J., and Therneau, T. M. (2016). Assessing calibration of prognostic risk scores. *Statistical methods in medical research*, 25(4) :1692–1706. 46
- D’agostino, R. and Nam, B.-H. (2003). Evaluation of the performance of survival analysis models : discrimination and calibration measures. *Handbook of statistics*, 23 :1–25. 46, 47
- Demler, O. V., Paynter, N. P., and Cook, N. R. (2015). Tests of calibration and goodness-of-fit in the survival setting. *Statistics in medicine*, 34(10) :1659–1680. 46, 47, 49
- Gail, M. H. and Pfeiffer, R. M. (2005). On criteria for evaluating models of absolute risk. *Biostatistics*, 6(2) :227–239. 46
- Gerds, T. A., Andersen, P. K., and Kattan, M. W. (2014). Calibration plots for risk prediction models in the presence of competing risks. *Statistics in medicine*, 33(18) :3191–3203. 46
- Mboup, B., Blanche, P., and Latouche, A. (2020). On evaluating how well a biomarker can predict treatment response with survival data. *Pharmaceutical Statistics*. 46
- Pepe, M. and Janes, H. (2013). Methods for evaluating prediction performance of biomarkers and tests. In *Risk assessment and evaluation of predictions*, pages 107–142. Springer. 46

Chapitre 5

Contribution au développement d'un biomarqueur pronostique pour le cancer de l'ovaire

Sommaire

5.1	Introduction	56
5.2	Indexe des centrosomes par noyau (ICN)	57
5.3	Matériels et Méthodes	57
5.3.1	Échantillon d'étude	57
5.3.2	Analyses statistique	58
5.4	Résultats	59
5.5	Conclusion	60
5.6	Résumé de l'article en français	60
5.7	Manuscrit publié dans BioRxiv	61

5.1 Introduction

Les cancers épithéliaux de l’ovaire (CEO) représentent plus de 90% des cancers de l’ovaire de l’adulte et est le cancer gynécologique le plus létal [Berns and Bowtell, 2012; Konstantinopoulos and Awtrey, 2012]. Un diagnostic tardif et les options thérapeutiques limitées, malgré l’utilisation des nouveaux traitements tels que les inhibiteurs de l’angiogenèse¹ ou les voies de réparation de l’ADN, entraînent le taux de mortalité élevé [Konstantinopoulos et al., 2015; Pujade-Lauraine et al., 2014]. La classification histologique comprend principalement les carcinomes séreux, endométrioïdes, mucineux et à cellules claires. Parmi eux, le plus courant est le séreux de haut grade (HGSOC) qui répond au moins initialement à la chimiothérapie mais présente un pronostic global plus défavorable [Ramalingam, 2016]. Les carcinomes transcriptomiques² [Mateescu et al., 2011; Tothill et al., 2008] et le profilage protéogénomique [Gentric et al., 2019; Zhang et al., 2016] du HGSOC ont suggéré tout un spectre de diversité moléculaire qui peuvent être liés à la survie des patients, sans pour autant permettre une compréhension approfondie du mécanisme conduisant à une rechute [Lloyd et al., 2015]. En outre, jusqu’à 50% des HGSOC présentent des défauts dans les voies de recombinaison homologue³ [Network et al., 2011]. Cependant, le centrosome est le principal centre d’organisation des microtubules des cellules animales. Il se compose d’une paire de centrioles entourée d’une matrice de protéines appelée matériel péricentriolaire. Les anomalies de centrosomes sont associées à un grand nombre de pathologies humaines telles que le cancer. C’est ainsi qu’un biomarqueur continu conçu à partir des centrosomes et des noyaux tumoraux a été créé et appelé le nombre moyen de centrosome par noyau ou indice des centrosomes par noyau (ICN).

Notre contribution dans ce travail était d’évaluer la capacité pronostique du ICN et déterminer le seuil optimal de l’ICN afin de classer les patients à haut et bas risque de rechute du CEO. Il a été constaté que 80% des tumeurs contenaient un ICN élevé par rapport aux tissus sains suggérant qu’un nombre augmenté de centrosomes est présent dans les tumeurs ovariennes. De plus, différentes anomalies structurelles du centrosome qui semblent également être associées à un niveau d’ICN élevé ont été décrites. L’intégration des données cliniques montre que le niveau d’ICN a un impact avec la survie globale des patientes : un faible ICN semble être lié à une survie plus faible et à une rechute plus rapide. Notre étude a montré

1. L’angiogenèse est le processus de croissance de nouveaux vaisseaux sanguins à partir de vaisseaux pré-existants

2. La transcriptomique est l’étude de l’ensemble des ARN messagers produits lors du processus de transcription d’un génome

3. La recombinaison homologue est un type de recombinaison génétique où les séquences de nucléotides sont échangées entre des molécules d’ADN identiques

que l'augmentation du nombre de centrosomes peut impacter positivement la progression tumorale du cancer épithélial de l'ovaire.

Dans ce chapitre, nous allons détailler la méthodologie utilisée pour déterminer le seuil de l'ICN après avoir évalué sa valeur pronostique.

5.2 Indexe des centrosomes par noyau (ICN)

L'analyse commence par la quantification des anomalies du nombre de centrosomes dans les tumeurs. Les tissus tumoraux semblent très désorganisés et il est difficile de déterminer le nombre de centrosomes par cellule car, dans de nombreux cas, les centrosomes ne sont pas étroitement associés au noyau. Pour quantifier sans ambiguïté le nombre de centrosomes et pouvoir comparer toutes les tumeurs et tous les tissus sains, un décompte visuel du nombre de noyaux et du nombre de centrosomes dans chacun des dix champs choisis au hasard a été réalisé en utilisant l'immunofluorescence⁴ et la microscopie de pointe. L'indice des centrosomes par noyaux (ICN) est déterminé en divisant le nombre de centrosomes par le nombre de noyaux.

5.3 Matériels et Méthodes

5.3.1 Échantillon d'étude

Cette étude porte sur une cohorte rétrospective de 100 patientes présentant un cancer épithélial de l'ovaire (CEO) de stade avancé, lors de chirurgies effectuées et avant chimiothérapie à l'Institut Curie. Les patientes ont préalablement signé un formulaire de consentement pour autoriser l'utilisation des échantillons biologiques à des fins de recherche. Le stade de la maladie a été déterminé au moment de la chirurgie. L'histopathologie ainsi que le grade des tissus tumoraux ont été évalués, par la suite, par des pathologistes et selon les critères de classification de la Fédération Internationale de Gynécologie et Obstétrique (FIGO). Des données cliniques de ces patientes sont présentées dans le tableau 5.1. Pour chaque tumeur, l'indice des centrosomes par noyaux (ICN), a été évalué par l'équipe de Renato Basto. Étonnamment, nous avons constaté que la fréquence d'amplification des centrosomes était moins importante que ce qui est prédit dans la littérature, lorsqu'elle était définie par la présence de plus de deux centrosomes dans une cellule..

4. L'immunofluorescence est une technique d'immuno-marquage, qui utilise des anticorps ainsi que des fluorochromes

Echantillons de tumeur d'ovaires	100 échantillons
tumeurs primaires	100%
Age au diagnostic	
Mediane	60
Min-Max	31-87
Grades	
Low	3 (3%)
High	97 (97%)
FIGO	
I	10 (10%)
II	8 (8%)
III	59 (59%)
IV	12 (12%)
NA	11 (11%)
Types histologiques	
Séreux	90 (90%)
Endometrioïde	3 (3%)
Mucineux	4 (4%)
Cellules claire	3 (3%)
Nombre de tumeurs avec donnée génomique	81 (81%)

TABLEAU 5.1 – Quelques paramètres cliniques

5.3.2 Analyses statistique

Une partie cruciale de ce projet consiste à déterminer le seuil optimal pour l'ICN afin d'avoir deux groupes homogènes comme tumeurs à ICN faible ou élevé. Nous avons employé la courbe de prédictivité car cette méthode graphique est facilement compréhensible par les biologistes. Elle combine à la fois des concepts issus de la modélisation des risques et de l'analyse des performances des populations telles que la sensibilité et la spécificité, les valeurs prédictives et les courbes de ROC [Pepe et al., 2008]. De plus, elle est directement reliée

à une mesure de discrimination globale comme l'aire sous la courbe de ROC [Viallon and Latouche, 2011]. Elle correspond au tracé du risque estimé à partir d'un modèle pronostique en fonction de la valeur ordonnée du biomarqueur continu candidat (ICN). Pour construire une courbe de prédictivité, nous devons classer par ordre croissant les valeurs du risques et représenter graphiquement les valeurs ordonnées (voir figure 5.1). Nous avons utilisé un modèle de régression de Cox pour une réponse censuré à droite (rechute jusqu'à 6 mois) pour estimer ce risque. Le seuil optimal pour la valeur de l'ICN, correspond à l'intersection (si elle existe) du risque sans les informations de l'ICN (i.e. la prévalence d'une rechute à 6 mois dans l'échantillon étudié) et du risque donné avec l'ICN. Nous avons estimé les intervalles de confiance du seuil en utilisant des méthodes de rééchantillonnage.

5.4 Résultats

L'analyse a porté sur un total de 653627 noyaux, 874766 centrosomes provenant de 1174 champs, avec une moyenne de 5248 noyaux comptés par tumeur. Dans les tumeurs, l'ICN était beaucoup plus variable. En moyenne, 1.43 ± 0.038 , le minimum étant de 0,88 et le maximum de 2.45. Il est intéressant de noter que 89% des 100 tumeurs présentaient un ICN supérieur au ICN moyen des tissus sains et seulement 9% des tissus tumoraux présentaient une amplification des centrosomes, lorsqu'elle était définie par la présence de plus de deux centrosomes dans une cellule. L'utilisation des courbes de prédictivité permet d'obtenir la valeur optimale du seuil de l'ICN est de 1.456 ($95\%CI = [1.22 - 1.76]$, voir figure 5.1). En ce qui concerne la robustesse des estimations, nous sommes rassurés car un nombre suffisant d'événements a été observé dans notre analyse précédente (70 rechutes et 88 décès dans les échantillons HGSOc) et une censure administrative à 6 mois a été utilisé.

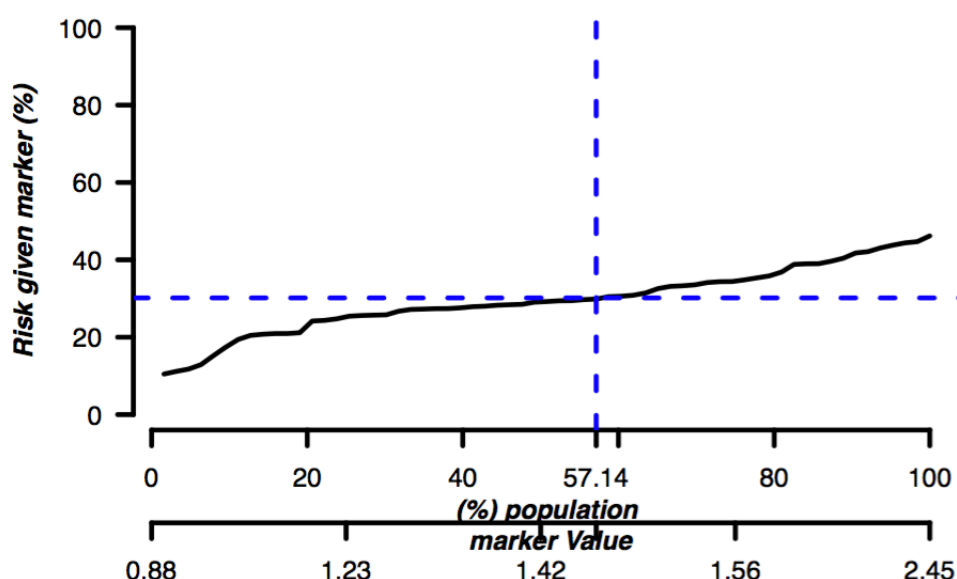


FIGURE 5.1 – Courbe de prédictivité pour la détermination du seuil de l’index des centrosomes par noyaux (ICN)

5.5 Conclusion

La courbe de prédictivité est très utile dans l’évaluation des marqueurs pronostiques. Cette méthodologie de détermination du seuil est optimale en ce sens que le seuil estimé minimise les taux d’erreur de faux positifs et faux négatifs [McIntosh and Pepe, 2002]. L’obtention du seuil de l’ICN à 1.45 permet de classer les patients à haut risque de rechute à 6 mois ($ICN > 1.45$) et bas risque ($ICN < 1.45$).

La prochaine étape porte sur l’automatisation du calcul de l’ICN afin de pouvoir répliquer notre analyse sur d’autres cohortes rétrospectives.

Le travail présenté dans ce chapitre n’est qu’une partie et le travail dans sa globalité est réalisé sous forme d’article. Ci-dessous l’article publié dans BioRxiv.

5.6 Résumé de l’article en français

Le centrosome est le principal centre d’organisation des microtubules des cellules animales. Il se compose d’une paire de centrioles entourée d’une matrice de protéines appelée matériel péricentriolaire (PCM). Les anomalies de centrosomes sont associées à un grand nombre de pathologies humaines telles que la microcéphalie ou le cancer. Le but de ce travail est d’analyser le type et la fréquence des anomalies de centrosomes dans le cancer épithélial de l’ovaire (CEO) qui le cancer gynécologique le plus létal dans le monde. Nous avons commencé notre étude en analysant le nombre moyen de centrosome par noyau (Index Centrosome/Noyau =

ICN) et par tumeur dans une cohorte de 100 patientes atteintes du cancer de l’ovaire. Nous avons constaté que 80% des tumeurs contenaient un ICN élevé par rapport aux tissus sains suggérant qu’un nombre augmenté de centrosomes est présent dans les tumeurs ovariennes. De plus, nous avons également décrit différentes anomalies structurales du centrosome qui semblent également être associées à un niveau d’ICN élevé. L’intégration aux données cliniques montrent que le niveau d’ICN a un impact avec la survie globale des patientes : un faible ICN semble être lié à une survie plus faible et à une rechute plus rapide. En utilisant des xénogreffes (PDXs) et des modèles cellulaires de cancer de l’ovaire permettant de manipuler le nombre de centrosomes, nous avons constaté qu’un ICN élevé pouvait avoir un impact positif sur la réponse à la chimiothérapie et inhiber la dissémination des cellules cancéreuses. Ces résultats montrent que l’augmentation du nombre de centrosomes peut impacter positivement la progression tumorale.

5.7 Manuscrit publié dans BioRxiv

Centrosome amplification favours survival and impairs ovarian cancer progression

Jean-Philippe Morretton¹, Aurélie Herbette², Camille Cosson², Bassirou Mboup³, Aurélien Latouche³, Pierre Gestraud⁴, Tatiana Popova⁵, Marc-Henri Stern⁵, Fariba Nemati^{2,6}, Didier Decaudin^{2,6}, Guillaume Bataillon⁷, Véronique Becette⁷, Didier Meseure⁷, André Nicolas⁷, Odette Mariani⁸, Claire Bonneau⁹, Jorge Barbazan¹⁰, Anne Vincent-Salomon⁷, Fatima Mechta-Grigoriou¹¹, Sergio Roman Roman², Roman Rouzier^{3,9}, Xavier Sastre-Garau^{7,12}, Oumou Goundiam^{2,13} and Renata Basto^{1,13}

1- Biology of centrosomes and genetic instability, CNRS, UMR144, Institut Curie, PSL Research University, 12 rue Lhomond, 75005 Paris, France.

2- Department of Translational Research, Institut Curie, PSL University, 26 rue d'Ulm, F-75248 Paris Cedex 05, France.

3- Statistical Methods for Precision Medicine, INSERM U900, Institut Curie, 35 Rue Dailly, 92210 Saint-Cloud, France.

4- Bioinformatics and Computational Systems Biology of Cancer, Mines Paristech, INSERM U900, Institut Curie, PSL University, 26 rue d'Ulm, F-75248 Paris Cedex 05, France.

5- DNA Repair & Uveal Melanoma (D.R.U.M.), INSERM U830, Institut Curie, PSL Research University, 26 rue d'Ulm, F-75248 Paris Cedex 05, France.

6- Laboratory of Preclinical Investigation, Department of Translational Research, Institut Curie, PSL University, 26 rue d'Ulm, F-75248 Paris Cedex 05, France.

7- Department of pathology, Institut Curie, 26 rue d'Ulm, F-75248 Paris Cedex 05, France.

8- Biological Resource Center, Department of pathology, Institut Curie, PSL Research University, BRIF BB-0033-00048, 26 rue d'Ulm, 75248 Paris, France.

9- Department of surgery, Institut Curie, 35 Rue Dailly, 92210, Saint-Cloud, France.
UFR Simone Veil - Santé, Université Versailles Saint Quentin, Université Paris
Saclay, Montigny le Bretonneux, France.

10- Migration and invasion Laboratory, CNRS, UMR144, Institut Curie, PSL
Research University, 12 rue Lhomond, 75005 Paris, France.

11- Stress and Cancer Laboratory, INSERM U830, Institut Curie, PSL Research
University, 26 rue d'Ulm, Paris, F-75005, France.

12- Present address: Laboratory of pathology, Intercommunal Hospital Center of
Creteil, 40 avenue de Verdun, 94010 Creteil Cedex

13- authors for correspondence: oumou.goundiam@curie.fr and
renata.basto@curie.fr

Abstract

Centrosome amplification has been described as a common feature of human cancers and it is known to promote tumorigenesis when induced in animals. However, little is known about the real status of centrosome numbers in human cancers and whether numerical alterations are solely associated with poor prognosis. To address this question, we have analyzed a large cohort of human epithelial ovarian cancers (EOCs) from 100 patients using state-of-the-art microscopy to determine the Centrosome-Nucleus Index (CNI) of each tumor. We found that EOCs are highly heterogeneous, with infrequent but strong centrosome amplifications leading to higher CNI than in healthy tissues. Strikingly, while a correlation between CNI and genomic alterations, such as aneuploidy or chromosome rearrangements could not be established, we found that high CNI correlates with increased patient survival and sensitivity to chemotherapy. Using ovarian cancer cellular models to manipulate centrosome numbers and Patient-Derived Xenografts (PDXs), we found that higher CNIs can positively impact the response to chemotherapy and inhibit cell dissemination. Our findings highlight a novel paradigm linking centrosome amplification to the inhibition of tumor progression.

Introduction

The centrosome is the main microtubule (MT) -organizing center of animal cells. Each centrosome is composed of two centrioles surrounded by pericentriolar material (PCM), which is the site of MT nucleation. The centrosome facilitates the accuracy of chromosome segregation during mitosis and influences cell polarity and migration ¹⁻⁴. Centrosome duplication is normally tightly controlled to ensure that each centrosome duplicates only once per cell cycle ^{5,6}. The presence of more than two centrosomes in a cell, centrosome amplification, has long been associated with tumorigenesis, with T. Boveri ⁷ proposing a link between extra centrosomes, multipolar divisions and consequent aneuploidy. When induced through the manipulation of the centrosome duplication machinery, centrosome amplification, was sufficient to drive tumor formation *in vivo* in a variety of tissues from different animal models ⁸⁻¹². Interestingly, although the consequences of centrosome amplification have been normally associated with abnormal cell division and the generation of aneuploidy ^{7,9,11,12}, centrosome amplification can also impact cellular homeostasis in alternative ways. When induced in breast epithelial cells, centrosome amplification resulted in the assembly of invasion-like features, which were RAC1-dependent ¹³. More recently, it has been shown that centrosome amplification can drive invasion in a non-cell autonomous manner through increased oxidative stress ¹⁴. Non-cell autonomous detachment of mitotic tumor cells has also been described in organoids containing increased levels of Ninein-like protein, which induces centrosome structural defects ¹⁵⁻¹⁷. Importantly however, even if many studies have described numerical centrosome defects in cultured cancer cells only a limited number of studies tumors has analyzed centrosome number alterations *in situ*.

Epithelial ovarian cancers (EOCs) are the most lethal gynecologic malignancies ^{18,19}. The high mortality rate is a result of delayed diagnosis and limited therapeutic options despite the use of new drugs, such as the inhibitors of angiogenesis or DNA repair pathways ^{20,21}. 75% of EOC patients are diagnosed at advanced disease stages, resulting in a 5-year overall survival rate that has recently been improved from 30 to ~47% ^{22,23}. Histological classification includes mainly serous, endometrioid, mucinous and clear cells carcinomas. The most common EOCs subtype is high-grade serous (HGSOC) which responds at least initially to chemotherapy but presents a worse overall prognosis ²⁴. Transcriptomic ²⁵⁻²⁷ and

proteogenomic profiling^{28,29} of HGSOC suggested a whole spectrum of molecular diversity that can be linked to patient survival, without yielding a deep understanding of the mechanism leading to relapse³⁰. Moreover, up to 50% of HGSOC exhibit defects in homologous recombination (HR) pathways²⁶. HR deficient (HRD) patients with germline or somatic mutations in *BRCA1/2* genes are known to be more sensitive to platinum-based chemotherapy and Parp inhibitors than non-BRCA-mutated tumors^{19,21,31} more broadly defined as the HR proficient (HRP) patients. Since EOCs are characterized by high level of genomic alterations^{26,32} and centrosome numerical defects are associated with aneuploidy, we characterized a large cohort of 100 naive EOCs, comprising 88 HGSOCs, using immunofluorescence and state-of-the-art microscopy. For each tumor, we established the centrosome-nucleus index (CNI) as a proxy to compare centrosome numbers among our cohort. Surprisingly, we found that the frequency of centrosome amplification was less important than what is predicted from the literature. Integration of CNI data with genomic and clinical data revealed a striking association between centrosome amplification and patient outcome. Using patient-derived xenografts (PDXs) and cell line models, we showed that centrosome amplification can positively influence the response to chemotherapy, while it can also inhibit tumor cell dissemination through the mesothelium. Our results demonstrated for the first time that centrosome amplification is not associated with a worse prognosis, but more surprising, they show that decreased centrosome numbers, translate in poorer response to chemotherapy and increased capacity of tumor cell invasion. Overall this study identifies decreased centrosome numbers, but not centrosome amplification, as a condition that favours ovarian cancer progression.

Results

Characterization of centrosome defects in human epithelial ovarian cancer (EOC) tissues

In order to analyze centrosomes in human epithelial ovarian cancers (EOCs), we designed a strategy where 20µm frozen tissue sections were obtained by the pathology department of Institut Curie. These were previously categorized as healthy tissues (corresponding to healthy ovaries from prophylactic oophorectomy or hysterectomy removal) or tumor tissues, which enclosed a mix of serous (90%), endometrioid (3%), mucinous (4%) and clear cell carcinoma (3%) (methods and Supplementary Table 1). Importantly, all tumors were naïve, obtained after surgery without previous neo-adjuvant chemotherapy. Tissues were methanol fixed and processed for immunostaining with two different antibodies CDK5RAP2 and pericentrin (PCNT), two PCM components, to unambiguously identify centrosomes through co-localization. Confocal microscopy was used to obtain optical sections from ten random fields in the entire tissue (Figure 1A). Analysis of healthy tissues allowed us to identify centrosomes through the co-localization of the two centrosomes markers (Figure 1B). We also noticed the presence of structures that only contained one of the two centrosome markers (Figure 1B), and importantly, these were not considered as centrosomes. To further characterize and confirm the centrosomal configurations described above, we used 3D structural illumination microscopy (3D-SIM) of ovarian tissues immunostained with the centriolar marker-Cep135 and PCNT, allowing higher resolution for both centrioles and PCM (Figure 1D). To our knowledge this represents the first centrosome super resolution analysis performed in human tissues and tumors. We found that in healthy tissues, each centrosome contained two centrioles and that PCNT surrounded one of the two centrioles, presumably the mother centriole (Figure 1D), as expected³³.

Analysis of tumor tissues revealed the presence of highly heterogeneous conditions with several centrosome abnormalities. Extra centrosomes were easily identified by the presence of multiple CDK5RAP2-PCNT positive co-localizing structures associated with one nucleus (Figure 1C). In certain cells, ECs were isolated and spread away from each other (Figure 1C- top panel) and these were named isolated centrosomes. In other cells, extra centrosomes were clustered together- clustered centrosomes (Figure 1C, middle panel). Interestingly, we also

observed a configuration that to our knowledge has never been described before, where ECs were tightly associated in a single structure appearing very tightly clustered and hence named super-clusters (Figure 1C lower panel). SIM analysis of these tumors, with the markers described above confirmed the aberrant extra centrosome morphologies (Figure 1E).

We next quantified the frequency of these defects in a cohort of 19 healthy tissues and 100 tumor tissues. We imaged ten random different fields, corresponding to different regions of the tumor. Importantly, we only imaged and analyzed regions corresponding exclusively to the tumor tissue, excluding regions of stromal tissue that normally surround the tumor. Interestingly, while the majority of tumors (60%) presented at least one of the defects described above in terms of centrosome number, whereas the remaining 40% of tumors did not show any of these defects (Figure 1F). This type of centrosome numerical aberration was never observed in healthy tissues. Still considering only the different type of ECs configurations found in tumors, we observed that 18% of the tumors (n=18, from 100) presented the three categories: isolated, clustered and super-clustered (Figure 1G).

All together, the methodology employed to analyse 100 ovarian tumors and comparison with ovarian healthy tissues revealed the presence of centrosome number abnormalities in a large fraction of EOCs.

Extra centrosomes are present in the large majority of EOCs, but high levels of centrosome amplification are infrequent

We next focused our analysis in the quantification of centrosome number abnormalities in tumors. Tumor tissues appeared very disorganized and it was difficult to ascertain the number of centrosomes per cell as in many cases, centrosomes were not closely associated with the nucleus. To unambiguously quantify centrosome number and to be able to compare all tumors and healthy tissues, we visually counted the number of nuclei and the number of centrosomes in each of the ten randomly chosen fields, and determined the Centrosome Nuclei Index (CNI) by dividing the number of centrosomes by the number of nuclei. It is important to mention that we tried to automatize centrosome and nuclear segmentation followed by quantification. This approach was however far from reproducing the manual counting, with a strong bias towards considering unrelated

structures as centrosomes. The data we present therefore results from manual counting.

Overall our analysis comprised a total of 653627 nuclei, 874766 centrosomes from 1174 fields, with an average of 5248 nuclei counted per tumor. In healthy tissues, the average CNI was 1.02 ± 0.02 and it was relatively stable, varying from 0.81 to 1.16 (Figure 2A). In tumors, however, the CNI was much more variable. On average, 1.43 ± 0.038 , with the minimum 0.61 and maximum at 2.55. Interestingly, 89% (n=89 out of 100) of the tumors presented a CNI superior to the average CNI found in healthy tissues (Figure 2A, yellow dashed line and Supplementary Figure 1A). However, only 9% (n=9 out of 100) of tumor tissues exhibited centrosome amplification (Figure 2A, green dashed line), when defined by the presence of more than two centrosomes in a cell ^{7,34,35}. We also investigated the frequency of extra centrosomes clusters and super-clusters per nuclei, and found that they were extremely uncommon ($0.71\% \pm 0.09$ for clusters and $0.72\% \pm 0.08$ for super-clusters) (Supplementary Figure 1B), confirming the low frequencies of extra centrosomes in these tumors.

We next dichotomized our population in two groups using Classification And Regression Trees (CART) methods ³⁶, restricting the analysis to the high-grade serous ovarian cancers (HGSOCs) within our cohort. These represented the majority of the tumors- 88%, which is also the case worldwide for EOCs ²⁴. This resulted in the categorization of the cohort into low CNI (≤ 1.45) and high CNI (> 1.45), with 55 tumors falling into the low CNI category, while 33 were placed in the high CNI category (Figure 2A, red line).

We first investigated whether the dichotomization of our tumor cohort in low and high CNI identified any preference for the different extra centrosome categories (isolated, cluster and super-cluster) identified by microscopy. Using multivariate analysis, we recognized a significant trend for isolated centrosomes and clusters ($p=0.021$ and $p=0.035$ respectively, Supplementary Figure 1C) to be associated with high CNI tumors. Interestingly however, even if not statistically significant, super-clusters were associated with low CNI tumors ($p=0.0788$, ns). To gain more information about the distribution of structures containing extra centrosomes, we plotted the number of clusters and super-clusters (as the structures that contain more centrosomes) in parallel to the CNI analysis (Figure 2B). We found that certain

tumors with low CNI (placed at the bottom of the graph) contained clusters and super-clusters at similar frequencies as tumor tissues placed at the other end of the graph. We hypothesized that low CNI tumors containing the same frequencies of extra centrosomes than high CNI tumors, should contain cells without centrosomes. Corroborating this hypothesis, we could easily identify regions without centrosomes (Supplementary Figure 1D) in tumors with low CNI values.

We concluded that EOCs are highly heterogeneous in terms of centrosome numbers, and surprisingly only a small population of tumor cells display extra centrosomes.

The CNI does not correlate with proliferation, mitotic indexes, genomic alterations or transcriptomic changes in HGSOCs

We next wanted to study the possible correlation between CNI and different molecular and clinical parameters. We analyzed whether the CNI status correlated with proliferation. We used two indicators, the mitotic index (MI) and, the proliferation marker Ki67 by H&E staining and immunochemistry respectively. Interestingly, we did not find any correlation between CNI and MI or Ki67 (Supplementary Figure 2A-B), suggesting that both CNI low and CNI high tumors show similar proliferative and mitotic indexes.

Genomic alterations are frequently found in HGSOCs ^{26,32}. Centrosome defects can lead to mitotic errors, chromosome instability and aneuploidy ^{37,38}. In order to identify a possible link between centrosome number and genomic alterations, we used high resolution Cytoscan arrays and GAP tools ³⁹. With these tools, we analyzed information related with chromosome content (ploidy) and the presence of small and/or large DNA structural rearrangements. Importantly, we did not find any correlation between CNI status and ploidy, chromosome number and DNA structural rearrangements (Supplementary Figure 2C-F). We concluded that small or large chromosome breaks were not associated with low and high CNI tumors.

Different pan-cancer studies ^{40,41} have shown that whole genome duplications (WGD) precedes many different types of genomic alterations. WGDs might represent a mechanism to generate aneuploidy, leading to chromosome number reduction, as shown in a mouse ovarian cancer model ⁴². WGD-positive (near tetraploid) tumors contain a ploidy of 3.31 on average, while ploidy is closer to ~1.99 (near diploid) for

WGD-negative tumors⁴¹. Centrosome amplification and WGD are hallmarks that have been associated, as both can be produced via the same mechanisms such as cytokinesis failure⁴⁰. We therefore examined if CNI correlated with ploidy in our tumor cohort. We found that this was not the case (Supplementary Figure 2G), even if we noticed that tumors with low CNI contained twice more near tetraploid (67%, n=26 out of 39 total tumors), than near diploid karyotypes (33%, n=13 out of 39 total tumors). In tumors with high CNI however, the distribution was similar for near tetraploid (46%, n=13 out of 28 total tumors) and for near diploid (54%, n=15 out of 28 total tumors), (Supplementary Figure 2G). This suggests that cytokinesis failure is not the only mechanistic explanation for extra centrosome accumulation, or that there are yet unidentified centrosome reduction mechanisms at play. Moreover, even if not statistically significant, low CNI seems to be associated with WGD and hence with worse clinical prognosis⁴⁰.

We also analyzed the transcriptome of our cohort using Affymetrix U133Plus2.0 microarray technology with the aim of identifying altered transcriptomic signatures. Interestingly, we did not find any major differences in gene expression between low and high CNI tumors, even when considering the extreme low or high CNI HGSOCS (not shown). We next compared with published HGSOCS transcriptome signatures, which have described fibrosis and stress profiles due to the expression of mesenchymal and oxidative stress genes, respectively²⁵. We found that while high CNI tumors display equivalent distributions between fibrosis and stress tumors, a tendency for fibrosis (63%, n=29 out of 46) was found in low CNI tumors (Supplementary Figure 2H). We conclude that no major transcriptomic alterations correlate with CNI. Interestingly however, low CNI tumors seem to display characteristics typical of fibrosis type, which are of poor prognosis.

High CNI correlates with better overall survival and response to chemotherapy in HGSOCS

The results described above suggested that low CNI is associated with worse prognosis. To independently test this possibility, we plotted patient survival curves according to the CNI status. We found that low CNI was associated with worse overall survival (Figure 3A, Log-rank test: p=0.018, HR=1.931, 95% CI=[1.14-3.28]) and furthermore with an increased risk of relapse, (Figure 3B, Log-rank test: p=0.018, HR=1.706, 95% CI=[1.06-2.75]). In contrast, high CNI was associated with

better overall survival. To avoid any bias due to tumour stage in the prognostic value of CNI, we thus investigated whether the CNI status reflected FIGO staging ⁴³). Importantly, we found that both low and high CNI tumors could be identified at all stages (I to IV) (Supplementary Figure 3A). Interestingly, the majority of the cases in our cohort correspond to stage III (59.0%, Supplementary Table 1) and these comprise once more low and high CNI tumours. We concluded that the association between high CNI and patient survival does not depend on tumor stage.

Relapse is a challenging situation in HGSOCs. Patients are categorized according to their response to chemotherapy and women who relapse within 6 months after the completion of the first line of chemotherapy are defined as platinum resistant ⁴⁴. We tested if the CNI parameter can be used as an indicator of relapse, defining early and late as before or after 6 months. We used predictiveness curves ⁴⁵ to evaluate the performance (robustness) of the CNI as a classifier and the optimum threshold allowing to stratify patients according to relapse. We performed this analysis taking into consideration the presence of extra centrosomes, since 63% (n=56 out of 88) of our HGSOC cohort harbour these defects. Using bootstrap resampling process, our predictiveness curves showed that the optimum CNI value is 1.456 (95% CI=[1.22-1.76], Supplementary figure 3B). This optimum value confirmed the threshold of 1.45 established previously to define low and high CNI tumors. Taken together our study shows for the first time that centrosome numbers can be used as an indicator of disease recurrence in HGSOC. Furthermore, these results clearly demonstrate the prognostic value of CNI status in HGSOC for patient survival and response to treatment. Unexpectedly, they also show that low CNI, and so reduced centrosome numbers are associated with worse prognosis.

High CNI tumors include more cases with homologous recombination deficiency

Mutations in genes encoding members of the DNA damage repair (DDR) pathway such as BRCA1 and BRCA2 (BRCA1/2), which are involved in homologous recombination (HR) lead to increased risk of breast and ovarian cancers ⁴⁶. Interestingly however, patients harbouring HR deficiency (HRD) are more sensitive to platinum, one of the two main chemotherapy for EOCs. We investigated whether there was an association between HRD and CNI status using the Large-scale transition (LST) genomic signature ^{47,48}. This signature is based on the presence of

large-scale chromosome breakpoints of at least 10Mb, which is an indicator of HRD. While low CNI status contained similar distributions of HRD and HR proficient (HRP) tumors (45% and 55% respectively), high CNI was mainly associated with HRD tumors (74% HRD and 26% HRP, respectively $p = 0.024$) (Figure 3C). Importantly, analysis of HRD patient overall survival did not show any significant association with the CNI status (Figure 3D, $p = 0.648$, $HR = 1.229$ and $95\% CI = [0.49-3.17]$). However, HRP patients showed significant differences according to CNI (Figure 3E $p = 0.0372$, $HR = 2.644$ and $95\% CI = [1.11-6.2]$), suggesting that this index can be used to stratify HRP patients. Overall, these results show that an increase in the number of centrosomes within a tumor can be beneficial for the patient. Further, while the CNI status does not seem to be a parameter to take into consideration in HRD patient survival (the ones that respond better to treatment), it can differentiate less sensitive HRP patients.

Investigating the effect of chemotherapy according to CNI

The treatment of EOCs relies on a combination of platinum and taxane derivative agents, which target DNA integrity and the microtubule cytoskeleton respectively ⁴⁴. We wondered whether there was an association between the CNI and the response to chemotherapy, which could explain the different response to treatment in patients with low and high CNI. We first tested the effect of a combination of carboplatin and paclitaxel in cell lines, where centrosome number can be easily manipulated. In order to increase centrosome number, we generated iOVCAR8-Plk4 and iSKOV3-Plk4 stable cell lines, where the over-expression of Plk4 (Plk4OE), the master centriole duplication kinase can be induced with Doxycycline (Dox) (hence referred to as Plk4OE+), a strategy previously used to amplify centrosomes (Figure 4A) ⁴⁹. Centrinone, a Plk4 inhibitor ⁵⁰ was used to decrease centrosome numbers (Figure 4A), and it will be referred to as centrinone cells. Treatment of either cell line with Dox and centrinone effectively impacted the CNI (Figure 4B and Supplementary Figure 4A-B). Although proliferation was decreased in Plk4OE+ and centrinone treated cells, these still proliferated (Supplementary Figure 4C-D) and the levels of apoptosis were only mildly increased (Supplementary Figure 4E). OVCAR8 and SKOV3 are HGSOC cell lines containing mutations in p53 ⁵¹, explaining the lack of response to centrosome number alterations, in contrast to diploid untransformed cell lines ^{50,52,53}.

We next determined the IC50 relative to carboplatin or paclitaxel for each cell line according to CNI status. In order to define efficient concentrations required to induce optimum cell growth inhibition, different concentrations of each drug including IC50 were used for drug combination optimization. Treatment of iOVCAR8-Plk4OE+ cells significantly impacted cell viability, while there was no additional effect on iOVCAR8 with reduced centrosome numbers, which was similar to controls (Ctrls) or to cells treated with DMSO (Figure 4C-D). Interestingly, treatment of iSKOV3 cells did not impact their viability (Figure 4E-F). We concluded that only iOVCAR8-Plk4OE+ (manipulated to contain a higher CNI- centrosome amplification) showed increased sensitivity to combined chemotherapy.

We next investigated the effect of chemotherapy *in vivo*, using two Patient-Derived Xenografts (PDXs) derived from two tumors from our cohort with distinct CNIs (Figure 4G-I). Importantly, to avoid any bias due to platinum sensitivity known for HRD tumors, we selected in our study two PDXs models derived from HRP tumors. Chemotherapies were administered intraperitoneally either every three weeks (carboplatin) or weekly (paclitaxel) during 6 weeks. Tumor growth was assessed until the tumor reached a volume of 2500 mm³, as stipulated by ethical regulations. The effect of chemotherapy on tumor growth was determined by the median period of time required to reach a 4 fold increase in tumor volume (RTVx4). Analysis of the median time showed that chemotherapy inhibited tumor proliferation of both ovarian PDXs. However, while the delay to reach RTVx4 comprised 20 days in the low CNI PDX OV014 compared to control (Figure 4J, Log-rank test: p=0.0013, HR=3.552, 95% CI=[2.47-20.3]), it was extended to 51 days in the high CNI PDX OV026 compared to the control (Figure 4K, Log-rank test: p=0.0005, HR=4.306, 95% CI=[3.925-58.50]). These results suggest that combined chemotherapy delays more significantly the growth of PDXs with higher CNI.

Low CNI ovarian cancer cells cause more efficient mesothelial cell clearance

It has been shown that centrosome amplification induces invasive oncogenic-like features in a 3-D culture mammary cell (MCF10A) model, both in cell or non-cell autonomous manner^{13,14}. Importantly, the levels of an activated form of RAC1, a small GTPase with described oncogenic signalling properties¹³, were increased. We thus tested whether centrosome number alterations lead to RAC1 activation in ovarian cancer cell lines. We did not observe any significant difference in activated

RAC1 levels in response to Dox or centrinone treatment (Supplementary Figure 5A-D). Thus, the differences observed between these two experimental condition seem to be justified by differential tissue specific responses to centrosome numerical alterations, as already described in flies and mice ^{8,9,11,12,35,54}.

EOCs undergo a particular mode of motility, where tumor cell invade the peritoneal cavity through a process called tumor dissemination ⁵⁵. Indeed tumor cells detach from the primary tumor site, adhere and go through mesothelial cells that enclose peritoneal organs ⁵⁶. Since centrosome number alterations have been shown to impact cell migration and invasion ^{3,13,57}, we hypothesized that differences in centrosome number might influence the capacity of ovarian cancer cells to invade the mesothelial barrier. We performed *in vitro* mesothelial clearance assays using ovarian cancer spheroids derived from iOVCAR8-Plk4 and iSKOV3-Plk4 described above (Figure 5A). Cells were treated for 4 days with Dox or centrinone to induce centrosome number alterations. In the last two days, they were plated on polyHEMA, to induce spheroid assembly. Importantly, we verified that this treatment did not influence centrosome number in any of the tested conditions (Supplementary Figure 6C-E). To perform the clearance assays we differentially labelled tumor spheroids, and mesothelial cells previously plated on collagen-coated surfaces. The capacity to clear and invade this layer of cells was measured over time (Figure 5B-C), as described in ⁵⁸. Both iOVCAR8 or iSKOV3 cells with or without DMSO showed comparable mesothelial clearance capacity. Interestingly, iOVCAR8-Plk4OE+ or iSKOV3-Plk4OE+, with extra centrosomes showed decreased capacity to clear. Strikingly, centrinone treatment increased the clearance capacity in both cell lines, with a more pronounced effect of iSKOV3 spheroids (Figure 5D, Supplementary Figure 6A-B and Supplementary videos 1-8). Statistical analysis (methods) revealed a significant difference in clearance capacity between Ctrl and Plk4OE+ cells. Importantly, and surprisingly, centrinone treated cells, showed increased clearance capacity when compared with DMSO treated cells.

Our results showed that tumor spheroids with reduced centrosome number displayed increased clearance capacity, while the presence of extra centrosomes induced the opposite behaviour. We wanted to compare heterogeneous spheroids that contained a mixed cell population in terms of centrosome number. We generated two types of spheroids for iOVCAR8 and iSKOV3 cells containing either a mix of Ctrl cells with Plk4OE+ or Ctrl cells with centrinone treated cells. In each

spheroid, cells were differentially labelled with green and red cell dyes before plating them on mesothelial cells labelled in blue. We confirmed that each treatment resulted in the expected alteration of centrosome number of the mixed spheroids (Supplementary Figure 6H-I). We then analyzed clearance capacity as described above. Interestingly, mixed spheroids containing centrinone treated cells cleared less efficiently than spheroids containing exclusively centrinone-treated cells. Remarkably however, these mixed spheroids remained more capable of clearing mesothelial cells than any spheroids that did not contain cells without centrosomes, highlighting the capacity for ovarian cancer cells with reduced centrosome numbers to drive clearance (Figure 5E, Supplementary Figures 6F-G and Supplementary videos 9-12).

Together, the mesothelial clearance assays described above shows that ovarian cancer cells with extra centrosomes seem to display a disadvantage in terms of cancer cell dissemination through mesothelial cells. Strikingly, our results also show that low CNI spheroids have a significant advantage (presented by two distinct cell lines), suggesting that low CNI cells might be the ones that considerably contribute to the metastatic process in HGSOCs.

Discussion

Whereas centrosome number alterations, namely centrosome amplification, has been shown to be sufficient to initiate tumorigenesis in animals ⁸⁻¹², the frequency of centrosome amplification in human tumors has remained under investigated. Here, we analyzed a large cohort of human ovarian tumors and we found that centrosome amplification is less frequent than what has been found in most human cancer cell lines, including ovarian cancer cell lines (⁵⁹ and this study- where CNI determination has been used for HGSOCs). Importantly, our study shows a higher heterogeneity in the number of centrosomes in ovarian tumors, with certain nuclei presenting over 15 centrosomes, while others even lack centrosomes. Tumor heterogeneity has been reported for many different tumor features and this work shows that heterogeneity of centrosome numbers is a hallmark of ovarian tumors.

Centrosome amplification has been correlated with aneuploidy and chromosome instability, however, a correlation between centrosome number increase and aneuploidy was not found in this cohort. These results suggest that even if centrosome amplification might contribute to chromosome number alterations, this is not the only mean by which ovarian tumors become aneuploid. Furthermore, our work also suggests that centrosome numbers do not influence the global rate of proliferation as revealed by Ki67 and mitotic index analyses. In vertebrates, p53 inhibits the proliferation of cells lacking centrosomes ⁵⁰ and in cells with extra centrosomes ^{49,60}. However since p53, is found mutated in the large majority of ovarian tumors ²⁶, centrosome number alterations most likely do not influence cell cycle progression or trigger cell cycle arrest.

Although still limited to a few tumor types, observations in certain tumors or in cancer cell lines derived from advanced tumors grades have suggested that centrosome amplification is correlated with advanced tumor stages and worse tumor prognosis ⁶¹. Surprisingly, however, in the EOC cohort analyzed in this study, which comprises in its large majority high-grade tumors, centrosome amplification was not frequent. And more surprisingly, increased centrosome numbers in high CNI tumors correlated with better prognosis. Patient overall survival was increased in tumors with high CNI, and time to relapse, was also found to be increased in the same group of patients. In contrast, patients with low CNI presented decreased survival and showed short-term relapse (Figure 6).

It is possible that many different factors contribute to the positive association between high CN1 and patient outcome in EOCs identified in this study. Our work has identified two important contributions. The first one is related with chemotherapy. We found that both in PDXs and in one ovarian cell line (OVCAR8), centrosome amplification delays tumor growth and decreases cell viability. This chemosensitivity is not due to HRD status since we used two PDXs models derived from HRP tumors, the ones that present worse prognosis. The fact that centrosome amplification did not influence chemotherapy in SKOV3 cell lines, suggests the contribution of differences in the genetic background of each cell line. Additional experiments should address the mechanisms underlying chemotherapy sensitivity in ovarian cancer cells with extra centrosomes. Since taxol targets and stabilizes the microtubule cytoskeleton, it is tempting to speculate that the combination of extra centrosomes and taxol might inhibit centrosome clustering, leading to multipolar cell division and so decreasing tumor cell viability ⁶².

The findings that ovarian cancer cell lines containing extra centrosomes showed decreased mesothelial cell clearance capacity was surprising in light of the results found in breast cells ^{13,14}, suggesting once more that tissue specific properties influence the consequences of altered centrosome numbers. Ovarian tumors do not show typical invasion features seen in other tumor types ⁶³. Indeed peritoneal metastases via transcoelomic dissemination appear to be a characteristic of HGSOCs ^{55,64}, where tumor cells detach from the primary tumor, using the ascitic fluid as a carrier to reach mesothelial cells that line the peritoneal cavity ^{56,65}. Importantly however, we show that decreased centrosome number confers an advantage to ovarian cancer cell dissemination (Figure 6). These results, together with the results described above, support the importance of characterizing centrosome numbers and its association with survival and relapse in light of cancer treatment possibilities. It has been suggested that inhibition of centrosome clustering or centrosome duplication might represent treatment opportunities to inhibit the proliferation of cancer cells with extra centrosomes ^{34,35,66}. Our results suggest that, at least in ovarian cancer, this type of therapy might inhibit the response to chemotherapy or even fuel the capacity of cancer cells to invade, with catastrophic outcome to the patient.

The results showing increased chemosensitivity of cancer cells and the PDX with high CN1 should be consider within the context of possible chemotherapy side

effects. If in heterogeneous primary tumors, cells with increased centrosome numbers are more susceptible of being eliminated after chemotherapy, this might result in the selection of cells with normal or low centrosome numbers. These cells show increased invasion capacity and so might represent an additional burden as they will favor metastatic behavior and relapse.

Another important finding of this work is the possibility of using centrosome number and in particular the CNI, as a prognostic tool. This might be quite advantageous to clinically manage patients according to CNI status, in particular in light of emerging novel therapies such as Parp inhibitors or others⁶⁷. It is possible that the use of CNI, in addition to genomic signatures might be beneficial in HGSOCs, which lack bona fide biomarkers and limited therapeutic options⁶⁸.

Figure legends

Figure 1. Characterization of centrosome numbers in EOCs

(A) Schematic diagram of the workflow used to analyze ovarian tissue sections. Frozen healthy or tumor ovarian tissues were sectioned into 20 μ m thickness sections and methanol fixed. These were subsequently immunostained for two centrosomes markers and nuclei were labeled with DAPI. Ten random fields were imaged through the entire Z-stack using confocal microscopy. Each field was analyzed and centrosomes and nuclei were quantified visually. (B-C) On the left, representative micrographs of low magnification views of healthy (B) and tumor tissues (C) immunostained with antibodies against pericentrin (PCNT) and CDK5RAP2, show in red and green respectively. DNA in blue. The white dashed squares represent the regions shown in higher magnification on the right. One centrosome was considered as such when PCNT and CDK5RAP2 signals co-localized. Lack of co-localization was noticed (B) and discarded for quantification. In tumors (C), extra centrosomes could be noticed and were present in three different configurations: isolated centrosomes (top), when more than two centrosomes were present and easily distinguished as well separated entities; clusters when extra centrosomes were present and remained closely associated and super-clusters when extra centrosomes were tightly packed and found close to each other. (D-E) Super resolution microscopy of healthy tissues (D) and tumor tissues (E) immunostained for the centriole marker Cep135 and PCNT. In normal tissues, two centrioles showing asymmetric PCM localization are detected, while in tumors (E), the three different configurations described above for extra centrosomes can be seen, with two centrioles forming each centrosome in the isolated centrosome category, many clustered centrioles in the cluster category and even more in the super-clustered configuration. (F) Graph bar showing the percentage of tissues with and without extra centrosomes. (G) Venn diagram illustrating the overlap of the different extra centrosome categories in tumors.

Figure 2. Characterization of the Centrosome Nuclei Index (CNI) in healthy and tumor tissues

(A) Plot showing the CNI value of all 100 tumors (blue) positioned in ascending value and 19 healthy tissues (yellow) analysed. The yellow dash line represents the average CNI value of all normal tissues analysed (1.02) and the red line the threshold of CNI value defining HGSOEs as low or high CNI tumors (1.43). The green dash line represents centrosome amplification as defined by the literature (>2 centrosomes in the cell). (B) Plot showing the total number of clusters and super-clusters identified in tumors. Note that the order of the tumors is conserved between the two plots to allow for comparison between the CNI and the number of extra centrosomes within the same tumor.

Figure 3. Higher CNI correlates with increased survival and time to relapse

(A-B) Kaplan-Meier curves showing patient overall survival (A) and the percentage of patients without relapse, after the first line of chemotherapy (B) according to CNI status. Statistical significance was assessed with Log-rank test for group comparison. (C) Contingency table showing the distribution of HR proficient (HRP) or deficient (HRD) in low and high CNI tumors. p value from Fisher's exact test. (D-E) Kaplan-Meier curves showing patient overall survival according to CNI status in HRD (D) and HRP (E) patients. Statistical significance was assessed with Log-rank test.

Figure 4. Chemotherapy inhibits growth of ovarian cancer cell lines and PDXs with extra centrosomes

(A) Schematic diagram of the generation of iOVCAR8 and iSKOV3 cells lines. OVCAR8 and SKOV3 cells lines were infected with lentiviral vector expressing the full length Plk4 construct. After selection, positive clones of inducible iOVCAR8 and iSKOV3 were isolated. These cells are referred to as control (Ctrl) cells. Addition of doxycycline (Dox+) generates Plk4OE+ cells. Treatment of Ctrl cells with centrinone reduces centrosome numbers and DMSO was used as control. (B) Graph bars plotting the CNIs of iOVCAR8 (left) and iSKOV3 (right) after Dox+, DMSO or Centrinone treatments. p value from one-way ANOVA (C-F). Graph bars plotting the percentage of viable cells in iOVCAR8 (C-D) and iSKOV3 (E-F) after the indicated chemotherapy treatment for 72 hours. Note that non-treated cells were used as

reference (100% viability), n=3 independent experiments. Statistical significance was assessed by a Wilcoxon test. (G) Schematic diagram of ovarian tumor engraftment in mouse and chemotherapy in PDXs. (H) Graph bar showing the CNIs of the tumors and corresponding PDXs. 10 fields from each PDX were quantified in a total of 9204 nuclei. Statistical significance was assessed with Mann-Whitney test. (I) Representative micrographs of tumors and the corresponding PDXs labeled with PCNT (red) and CDK5RAP2 (green) antibodies. (J-K) Kaplan-Meier curves illustrating the time elapsed to reach 4X the initial tumor volume in Ctrl or chemotherapy treated PDXs with low (J) or high (K) CNIs. The black line indicates the median time (in days) to reach RTVx4 under chemotherapy compared to Ctrl: 20 days for Low CNI PDXs and 51 days for High CNI PDXs. p value from Log-rank test.

Figure 5. Ovarian cancer cell lines with low CNI show improved capacity to clear through mesothelial cells

(A) Schematic diagram of workflow. Ovarian cancer cell lines were grown on polyHEMA to form spheroids and were labeled in red. Mesothelial cells were labeled in green and plated as monolayers on collagen I coated surfaces. The CNI was validated for each experiment by immunofluorescence microscopy. Ovarian spheroids were plated on mesothelial cells and imaged for 16hrs in time-lapse movies. The normalized clearance quantification was determined by dividing the hole size at different time points by the initial spheroid size. (B) Stills of a time-lapse movie of Ctrl spheroids (labeled in red in the merged figures and shown in grey on the top panel) and mesothelial cells (labeled in green in the merge figures and shown in grey on the middle panel). Time is shown in hours (h). (C) z- view of Ctrl cells as shown in B. Note the red cells at the beginning of the movie on top of the mesothelial layer while at later time points they have cleared through the mesothelial cells. (D) Graph bars of the normalized clearance in A.U. of iOVCAR8 (left) and iSKOV3 (right) spheroids after the indicated treatments. For each experimental condition 45 different spheroids were analyzed from three independent experiment. Statistical significance was assessed with ANCOVA test. (E) Graph bars of the normalized clearance in A.U. of iOVCAR8 (left) and iSKOV3 (right) spheroids after the indicated treatments or in mixed spheroids of the indicated treatments. For each experimental condition 30 different spheroids were analyzed from three independent experiments. Statistical significance was assessed with ANCOVA test.

Figure 6. Model of the impact of centrosome alterations in HGSOCs

In our cohort of HGSOCs we have identified two sub-populations of tumors. One with high and one with low CNI, which represent frequencies of centrosome number alterations. In both high and low CNI tumors, cells with extra centrosomes can be identified but the frequency of cells with one or zero centrosomes is higher in low CNI tumors. High CNI tumors (top purple cells) show decreased mesothelial clearance capacity and can show increased chemosensitivity. On the other hand, low CNI tumors, show increased mesothelial clearance capacity and low chemosensitivity, which accelerates tumor growth and peritoneal dissemination. Together, these conditions might facilitate tumor progression and relapse.

ACKNOWLEDGMENTS:

We are grateful to the patients that consent participating in this research and to the medical teams involved in their care. We thank S. Godinho, F. Gergely, S. McClelland, S. Taylor for sharing unpublished results, discussions and/or comments on the manuscript. We thank F. Edwards, V. Marthiens, S. Gemble, A. Goupil, D. Vargas and A. Simon for discussions and comments on the manuscript. We thank the Tissue Imaging (PICT-IBiSA) and Nikon Imaging Centre at Institut Curie, member of the French National Research Infrastructure France-BioImaging (ANR10-INBS-04). We thank A. Vieillefon, A. Rapinat and D. Gentien from the Genomics Platform of the translational research department at Institut Curie for cell line authentication.

The project was supported by a PIC project grant from the Labex CelTisphybio (2013, Institut Curie) and INCA PL-BIO grants (2015-PLBIO15-237) and the CNRS. The Basto lab is a member of the CelTisphybio labex.

AUTHOR CONTRIBUTION:

The project was initially designed and conceptualized by O.G., X. S. and R.B. with significant input from S.R.R. J.P.M. performed most experiments, including immunostaining, CNI quantifications of all the tissues and cell lines and the clearance assays; C.C was involved in setting the protocols for centrosome analysis in cells and tissues; A.H. help establishing the stable cell lines and performed *in vitro* chemotherapy experiments; B.M. and A.L. performed predictiveness curves and

provided expertise in statistical analysis. P.G. analysed gene expression levels and provided advice in statistical analysis. T.P. and M.H.S. help and advice on the HRD and ploidy status analysis. F.N. and D.D. performed all *in vivo* experiments. G.B. and V.B. did the pathological review of all the specimens; G.B. analyzed mitotic index; D.M and A.N. analysed ki67 in tumors; J.B. contributed to mesothelial clearance assays; O.G. performed activation assays. O.M. managed samples availability; X.S.-G. and A.V.S. provided human samples from pathology department of Institut Curie; C.B. and R.R. supplied surgical pieces to enlarge the cohort; R.R. provided expertise in ovarian cancers. F.M-G provided the OVCAR8 cell line, stress/fibrosis signature and advice in the methodology, as well as S.R.R. The work was supervised by O.G and R.B.

DECLARATION OF INTEREST:

M.H.S. and T.P. are inventors of the BRCAness/HRD (LST) genomic signature and Current exploitation of the patent is ongoing by Myriad Genetics. The other authors declare no competing interests.

REFERENCES:

- 1 Bettencourt-Dias, M. & Glover, D. M. Centrosome biogenesis and function: centrosomics brings new understanding. *Nature reviews. Molecular cell biology* **8**, 451-463, doi:10.1038/nrm2180 (2007).
- 2 Bornens, M. The centrosome in cells and organisms. *Science* **335**, 422-426, doi:10.1126/science.1209037 (2012).
- 3 Kushner, E. J. *et al.* Excess centrosomes disrupt endothelial cell migration via centrosome scattering. *J Cell Biol* **206**, 257-272, doi:10.1083/jcb.201311013 (2014).
- 4 Ogden, A., Rida, P. C. & Aneja, R. Heading off with the herd: how cancer cells might maneuver supernumerary centrosomes for directional migration. *Cancer Metastasis Rev* **32**, 269-287, doi:10.1007/s10555-012-9413-5 (2013).
- 5 Bornens, M. Centrosome composition and microtubule anchoring mechanisms. *Curr Opin Cell Biol* **14**, 25-34 (2002).
- 6 Gonczy, P. Centrosomes and cancer: revisiting a long-standing relationship. *Nat Rev Cancer* **15**, 639-652, doi:10.1038/nrc3995 (2015).
- 7 Boveri, T. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J Cell Sci* **121 Suppl 1**, 1-84, doi:10.1242/jcs.025742 (2008).
- 8 Basto, R. *et al.* Centrosome amplification can initiate tumorigenesis in flies. *Cell* **133**, 1032-1042, doi:10.1016/j.cell.2008.05.039 (2008).
- 9 Sabino, D. *et al.* Moesin is a major regulator of centrosome behavior in epithelial cells with extra centrosomes. *Curr Biol* **25**, 879-889, doi:10.1016/j.cub.2015.01.066 (2015).
- 10 Coelho, P. A. *et al.* Over-expression of Plk4 induces centrosome amplification, loss of primary cilia and associated tissue hyperplasia in the mouse. *Open Biol* **5**, 150209, doi:10.1098/rsob.150209 (2015).
- 11 Sercin, O. *et al.* Transient PLK4 overexpression accelerates tumorigenesis in p53-deficient epidermis. *Nature cell biology* **18**, 100-110, doi:10.1038/ncb3270 (2016).
- 12 Levine, M. S. *et al.* Centrosome Amplification Is Sufficient to Promote Spontaneous Tumorigenesis in Mammals. *Dev Cell* **40**, 313-322.e315, doi:10.1016/j.devcel.2016.12.022 (2017).
- 13 Godinho, S. A. *et al.* Oncogene-like induction of cellular invasion from centrosome amplification. *Nature* **510**, 167-171, doi:10.1038/nature13277 (2014).
- 14 Arnandis, T. *et al.* Oxidative Stress in Cells with Extra Centrosomes Drives Non-Cell-Autonomous Invasion. *Dev Cell* **47**, 409-424 e409, doi:10.1016/j.devcel.2018.10.026 (2018).
- 15 Schnerch, D. & Nigg, E. A. Structural centrosome aberrations favor proliferation by abrogating microtubule-dependent tissue integrity of breast epithelial mammospheres. *Oncogene* **35**, 2711-2722, doi:10.1038/onc.2015.332 (2016).
- 16 Casenghi, M. *et al.* Polo-like kinase 1 regulates Nlp, a centrosome protein involved in microtubule nucleation. *Dev Cell* **5**, 113-125 (2003).
- 17 Ganier, O. *et al.* Structural centrosome aberrations promote non-cell-autonomous invasiveness. *Embo j* **37**, doi:10.15252/embj.201798576 (2018).
- 18 Berns, E. M. & Bowtell, D. D. The changing view of high-grade serous ovarian cancer. *Cancer Res* **72**, 2701-2704, doi:10.1158/0008-5472.can-11-3911 (2012).

- 19 Konstantinopoulos, P. A. & Awtrey, C. S. Management of ovarian cancer: a 75-year-old woman who has completed treatment. *JAMA* **307**, 1420-1429, doi:10.1001/jama.2012.269 (2012).
- 20 Pujade-Lauraine, E. *et al.* Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *J Clin Oncol* **32**, 1302-1308, doi:10.1200/jco.2013.51.4489 (2014).
- 21 Konstantinopoulos, P. A., Ceccaldi, R., Shapiro, G. I. & D'Andrea, A. D. Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer. *Cancer Discov* **5**, 1137-1154, doi:10.1158/2159-8290.cd-15-0714 (2015).
- 22 Vaughan, S. *et al.* Rethinking ovarian cancer: recommendations for improving outcomes. *Nat Rev Cancer* **11**, 719-725, doi:10.1038/nrc3144 (2011).
- 23 Torre, L. A. *et al.* Ovarian cancer statistics, 2018. *CA Cancer J Clin* **68**, 284-296, doi:10.3322/caac.21456 (2018).
- 24 Ramalingam, P. Morphologic, Immunophenotypic, and Molecular Features of Epithelial Ovarian Cancer. *Oncology (Williston Park)* **30**, 166-176 (2016).
- 25 Mateescu, B. *et al.* miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response. *Nat Med* **17**, 1627-1635, doi:10.1038/nm.2512 (2011).
- 26 TCGA. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-615, doi:10.1038/nature10166 (2011).
- 27 Tothill, R. W. *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* **14**, 5198-5208, doi:10.1158/1078-0432.ccr-08-0196 (2008).
- 28 Gentric, G. *et al.* PML-Regulated Mitochondrial Metabolism Enhances Chemosensitivity in Human Ovarian Cancers. *Cell Metab* **29**, 156-173 e110, doi:10.1016/j.cmet.2018.09.002 (2019).
- 29 Zhang, H. *et al.* Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **166**, 755-765, doi:10.1016/j.cell.2016.05.069 (2016).
- 30 Lloyd, K. L., Cree, I. A. & Savage, R. S. Prediction of resistance to chemotherapy in ovarian cancer: a systematic review. *BMC Cancer* **15**, 117, doi:10.1186/s12885-015-1101-8 (2015).
- 31 Konstantinopoulos, P. A. & Matulonis, U. A. PARP Inhibitors in Ovarian Cancer: A Trailblazing and Transformative Journey. *Clin Cancer Res* **24**, 4062-4065, doi:10.1158/1078-0432.CCR-18-1314 (2018).
- 32 Goundiam, O. *et al.* Histo-genomic stratification reveals the frequent amplification/overexpression of CCNE1 and BRD4 genes in non-BRCAness high grade ovarian carcinoma. *Int J Cancer* **137**, 1890-1900, doi:10.1002/ijc.29568 (2015).
- 33 Conduit, P. T., Wainman, A. & Raff, J. W. Centrosome function and assembly in animal cells. *Nature reviews. Molecular cell biology* **16**, 611-624, doi:10.1038/nrm4062 (2015).
- 34 Godinho, S. A., Kwon, M. & Pellman, D. Centrosomes and cancer: how cancer cells divide with too many centrosomes. *Cancer Metastasis Rev* **28**, 85-98, doi:10.1007/s10555-008-9163-6 (2009).

- 35 Marthiens, V., Piel, M. & Basto, R. Never tear us apart--the importance of centrosome clustering. *J Cell Sci* **125**, 3281-3292, doi:10.1242/jcs.094797 (2012).
- 36 Breiman, L., Friedman, J., Olshen, R. & Stone, C. Classification and Regression Trees. *CRC Press* (1984).
- 37 Ganem, N. J., Godinho, S. A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460**, 278-282, doi:10.1038/nature08136 (2009).
- 38 Pihan, G. A. Centrosome dysfunction contributes to chromosome instability, chromoanagenesis, and genome reprogramming in cancer. *Front Oncol* **3**, 277, doi:10.3389/fonc.2013.00277 (2013).
- 39 Popova, T. *et al.* Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome Biol* **10**, R128, doi:10.1186/gb-2009-10-11-r128 (2009).
- 40 Bielski, C. M. *et al.* Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet* **50**, 1189-1195, doi:10.1038/s41588-018-0165-1 (2018).
- 41 Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45**, 1134-1140, doi:10.1038/ng.2760 (2013).
- 42 Lv, L. *et al.* Tetraploid cells from cytokinesis failure induce aneuploidy and spontaneous transformation of mouse ovarian surface epithelial cells. *Cell Cycle* **11**, 2864-2875, doi:10.4161/cc.21196 (2012).
- 43 Prat, J. Staging Classification for Cancer of the Ovary, Fallopian Tube, and Peritoneum: Abridged Republication of Guidelines From the International Federation of Gynecology and Obstetrics (FIGO). *Obstet Gynecol* **126**, 171-174, doi:10.1097/aog.0000000000000917 (2015).
- 44 Banerjee, S., Bookman, M. A. & Gore, M. in *Emerging Therapeutic Targets in Ovarian Cancer* (eds Stan Kaye, Robert Brown, Hani Gabra, & Martin Gore) 1-33 (Springer New York, 2011).
- 45 Huang, Y., Sullivan Pepe, M. & Feng, Z. Evaluating the predictiveness of a continuous marker. *Biometrics* **63**, 1181-1188, doi:10.1111/j.1541-0420.2007.00814.x (2007).
- 46 Chen, S. & Parmigiani, G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol* **25**, 1329-1333, doi:10.1200/jco.2006.09.1066 (2007).
- 47 Popova, T. *et al.* Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res* **72**, 5454-5462, doi:10.1158/0008-5472.can-12-1470 (2012).
- 48 Manie, E. *et al.* Genomic hallmarks of homologous recombination deficiency in invasive breast carcinomas. *Int J Cancer* **138**, 891-900, doi:10.1002/ijc.29829 (2016).
- 49 Holland, A. J. *et al.* The autoregulated instability of Polo-like kinase 4 limits centrosome duplication to once per cell cycle. *Genes & development* **26**, 2684-2689, doi:10.1101/gad.207027.112 (2012).
- 50 Wong, Y. L. *et al.* Cell biology. Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science* **348**, 1155-1160, doi:10.1126/science.aaa5111 (2015).
- 51 database, C. <https://portals.broadinstitute.org/ccle>.

- 52 Holland, A. J. *et al.* Polo-like kinase 4 controls centriole duplication but does not directly regulate cytokinesis. *Mol Biol Cell* **23**, 1838-1845, doi:10.1091/mbc.E11-12-1043 (2012).
- 53 Lambrus, B. G. *et al.* p53 protects against genome instability following centriole duplication failure. *J Cell Biol* **210**, 63-77, doi:10.1083/jcb.201502089 (2015).
- 54 Vitre, B. *et al.* Chronic centrosome amplification without tumorigenesis. *Proc Natl Acad Sci U S A* **112**, E6321-6330, doi:10.1073/pnas.1519388112 (2015).
- 55 Barbolina, M. V. Molecular Mechanisms Regulating Organ-Specific Metastases in Epithelial Ovarian Carcinoma. *Cancers (Basel)* **10**, doi:10.3390/cancers10110444 (2018).
- 56 Kipps, E., Tan, D. S. & Kaye, S. B. Meeting the challenge of ascites in ovarian cancer: new avenues for therapy and research. *Nat Rev Cancer* **13**, 273-282, doi:10.1038/nrc3432 (2013).
- 57 Cheng, H. W. *et al.* Centrosome guides spatial activation of Rac to control cell polarization and directed cell migration. *Life Sci Alliance* **2**, doi:10.26508/lisa.201800135 (2019).
- 58 Iwanicki, M. P. *et al.* Ovarian cancer spheroids use myosin-generated force to clear the mesothelium. *Cancer Discov* **1**, 144-157, doi:10.1158/2159-8274.cd-11-0010 (2011).
- 59 Marteil, G. *et al.* Over-elongation of centrioles in cancer promotes centriole amplification and chromosome missegregation. *Nature communications* **9**, 1258, doi:10.1038/s41467-018-03641-x (2018).
- 60 Marthiens, V. *et al.* Centrosome amplification causes microcephaly. *Nature cell biology* **15**, 731-740, doi:10.1038/ncb2746 (2013).
- 61 Zyss, D. & Gergely, F. Centrosome function in cancer: guilty or innocent? *Trends in cell biology* **19**, 334-346, doi:10.1016/j.tcb.2009.04.001 (2009).
- 62 Kwon, M. *et al.* Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. *Genes & development* **22**, 2189-2203, doi:10.1101/gad.1700908 (2008).
- 63 Weidle, U. H., Birzele, F., Kollmorgen, G. & Rueger, R. Mechanisms and Targets Involved in Dissemination of Ovarian Cancer. *Cancer genomics & proteomics* **13**, 407-423 (2016).
- 64 Worzfeld, T. *et al.* The Unique Molecular and Cellular Microenvironment of Ovarian Cancer. *Front Oncol* **7**, 24, doi:10.3389/fonc.2017.00024 (2017).
- 65 Niedbala, M. J., Crickard, K. & Bernacki, R. J. Interactions of human ovarian tumor cells with human mesothelial cells grown on extracellular matrix. An in vitro model system for studying tumor cell adhesion and invasion. *Exp Cell Res* **160**, 499-513 (1985).
- 66 Mason, J. M. *et al.* Functional characterization of CFI-400945, a Polo-like kinase 4 inhibitor, as a potential anticancer agent. *Cancer Cell* **26**, 163-176, doi:10.1016/j.ccr.2014.05.006 (2014).
- 67 Pillay, N. *et al.* DNA Replication Vulnerabilities Render Ovarian Cancer Cells Sensitive to Poly(ADP-Ribose) Glycohydrolase Inhibitors. *Cancer Cell* **35**, 519-533 e518, doi:10.1016/j.ccell.2019.02.004 (2019).
- 68 Lisio, M. A., Fu, L., Goyeneche, A., Gao, Z. H. & Telleria, C. High-Grade Serous Ovarian Cancer: Basic Sciences, Clinical and Therapeutic Standpoints. *Int J Mol Sci* **20**, doi:10.3390/ijms20040952 (2019).

- 69 Gao, H. *et al.* High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* **21**, 1318-1325, doi:10.1038/nm.3954 (2015).

Figure 1 - Morretton et al.

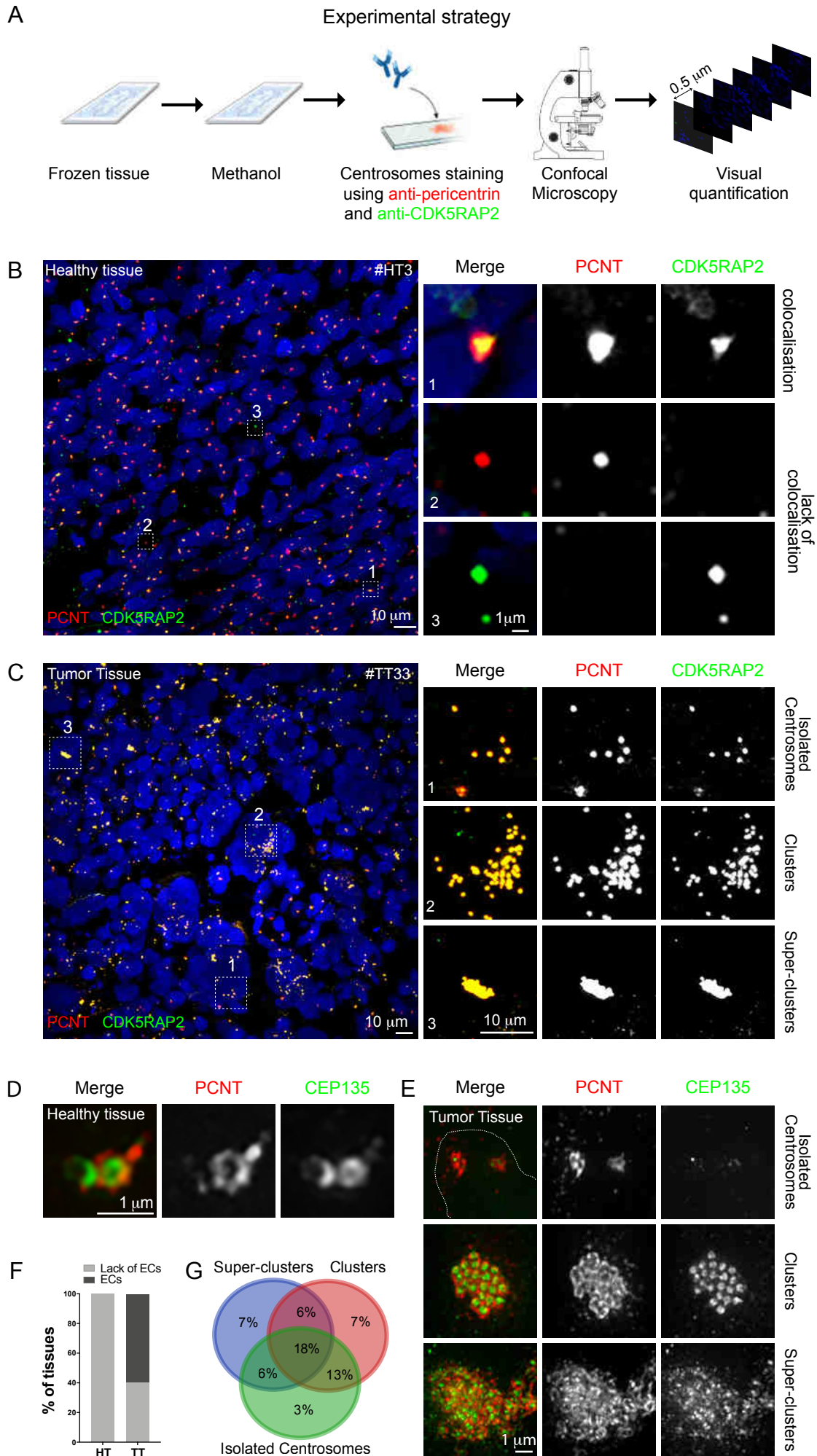


Figure 2- Morretton et al.

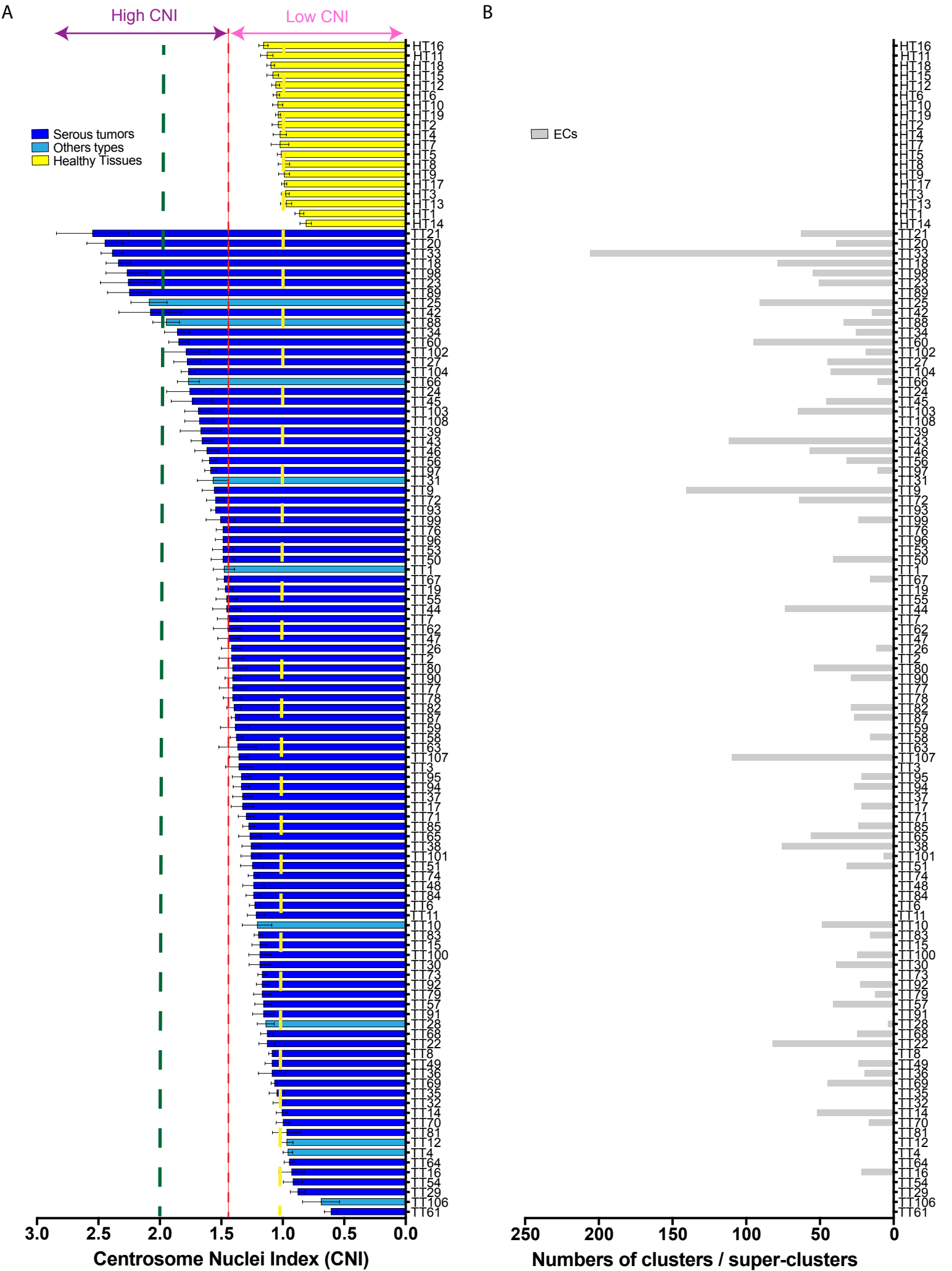


Figure 3 - Morretton et al.

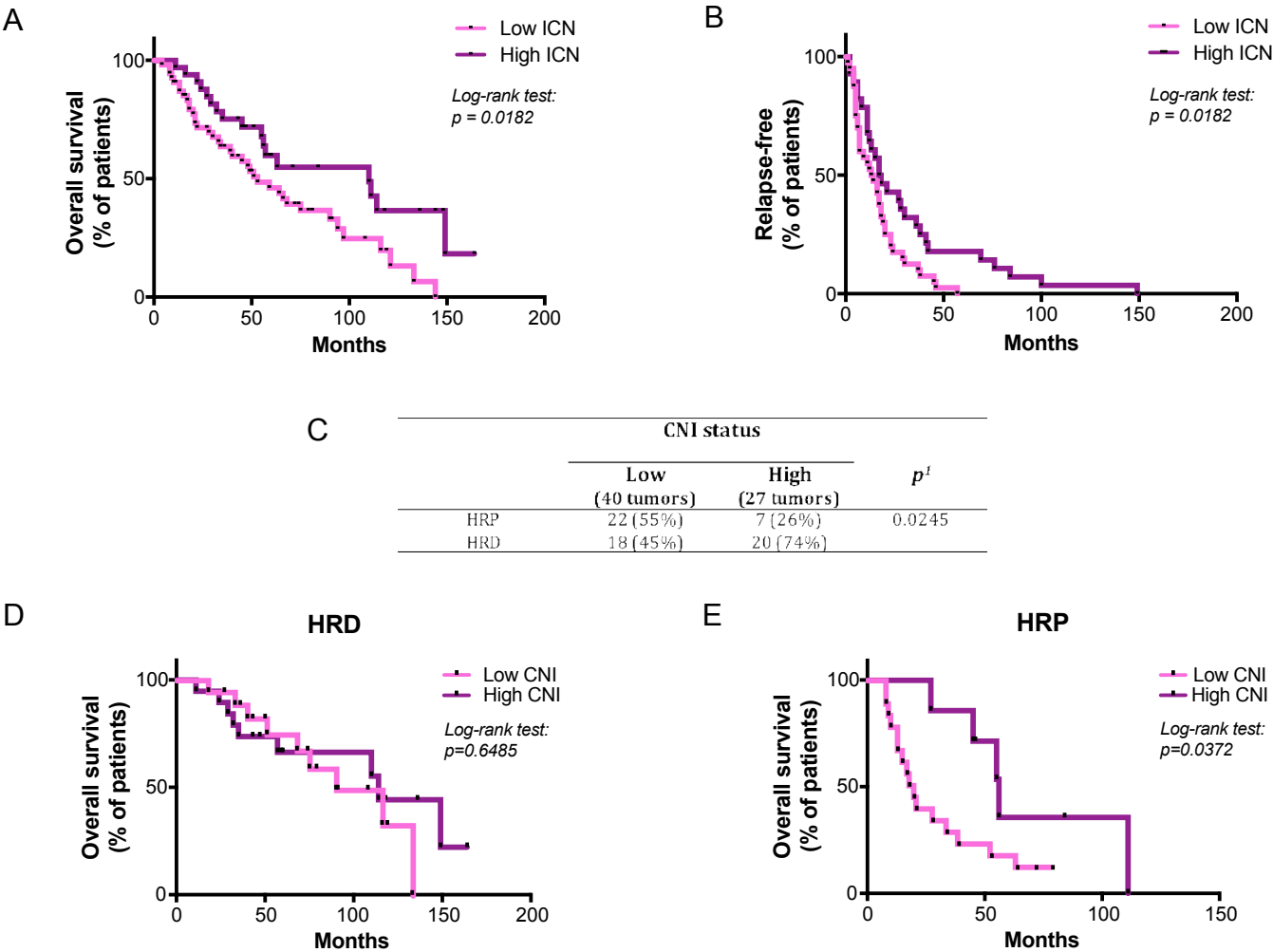
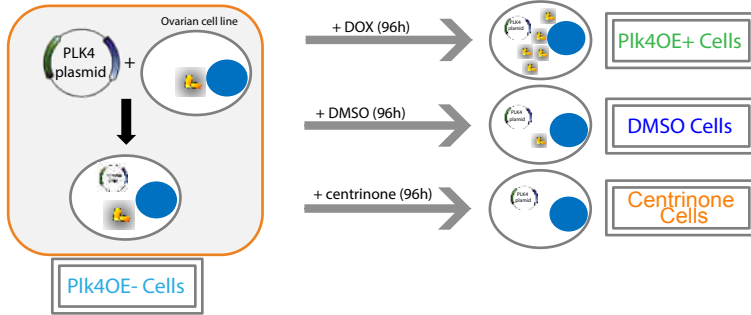
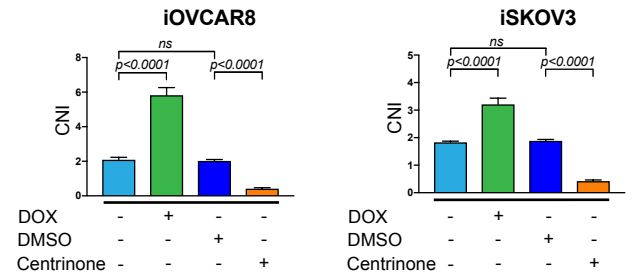


Figure 4 - Morretton et al.

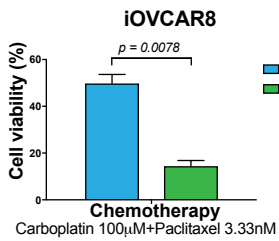
A Ovarian cancer stable cell lines with doxycycline inducible system (Plk4)



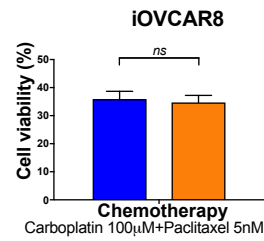
B



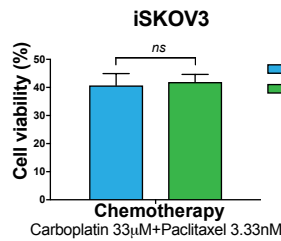
C



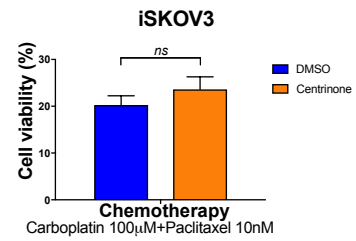
D



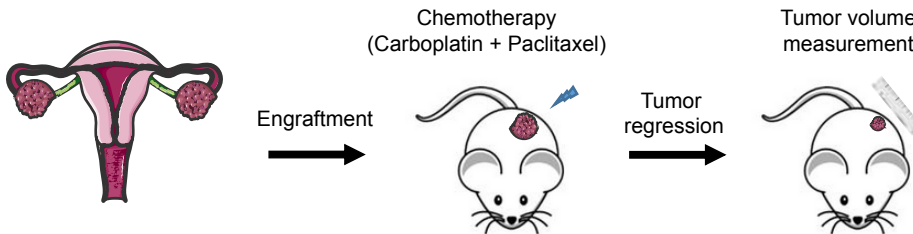
E



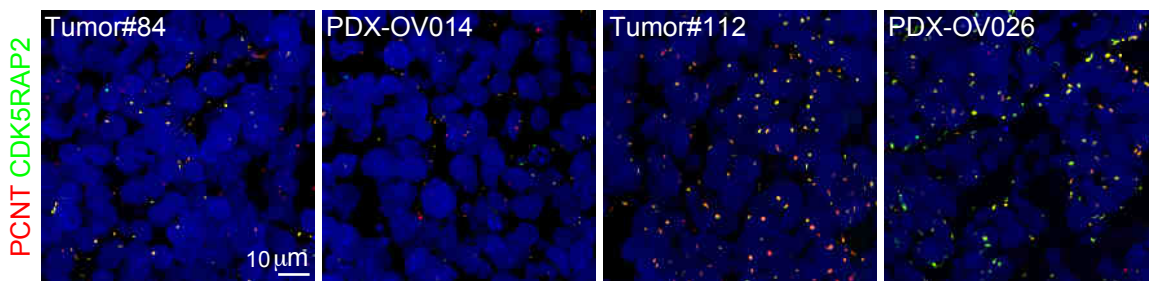
F



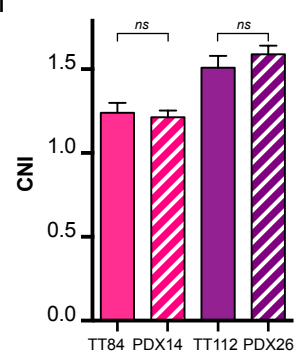
G



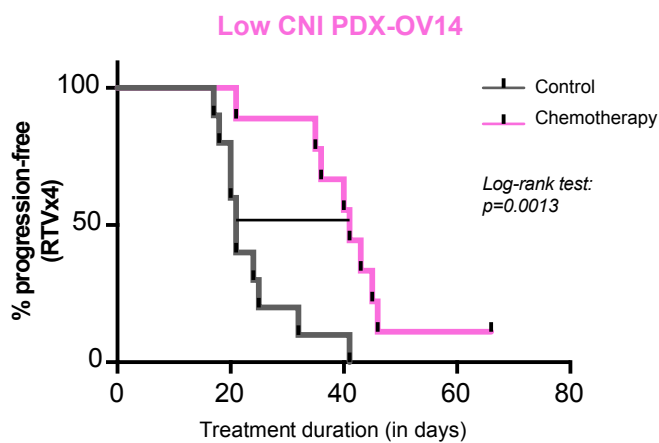
H



I



J



K

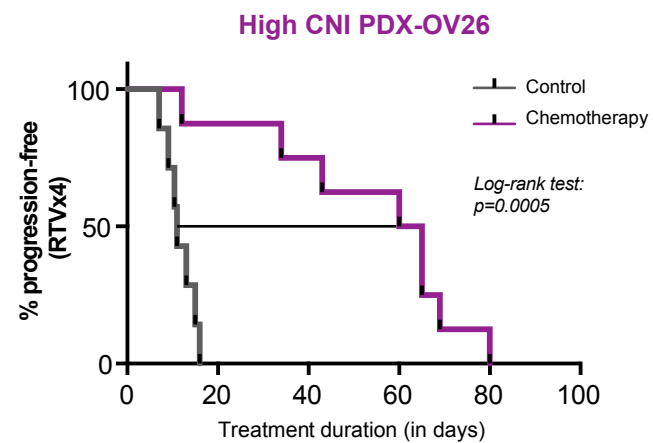


Figure 5 - Morretton et al.

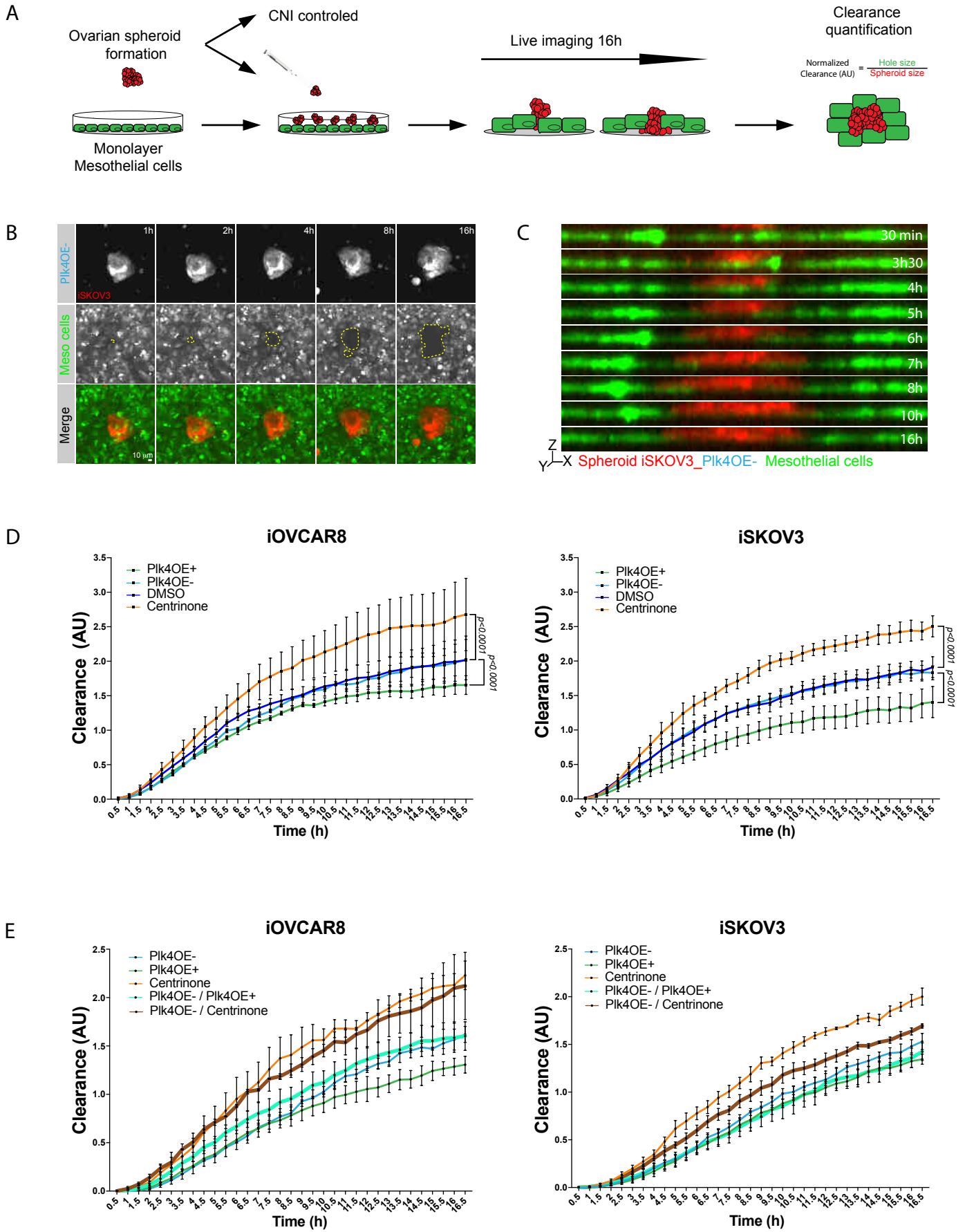
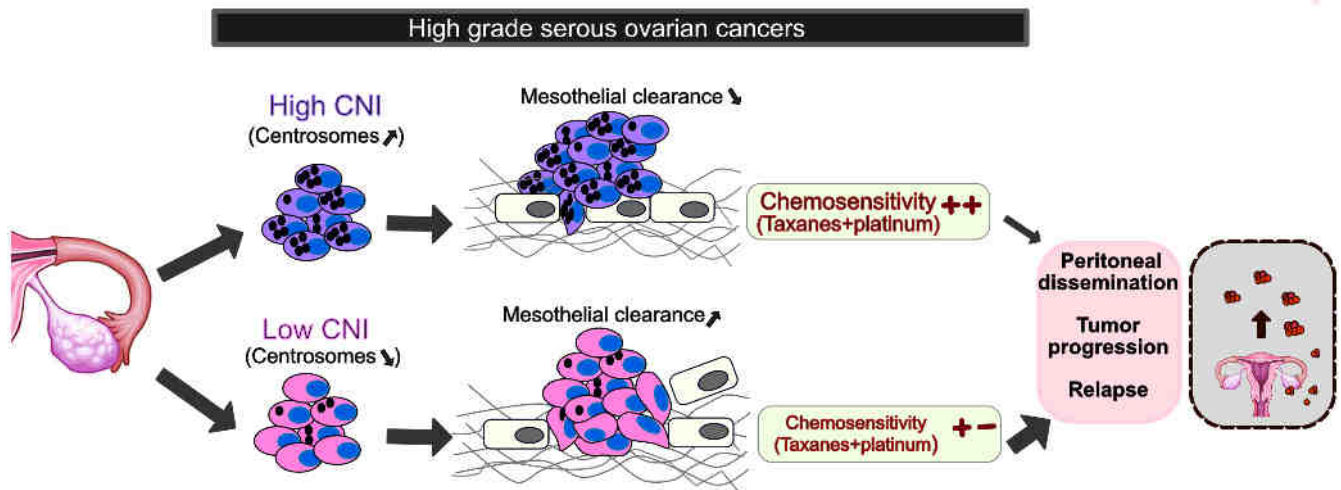
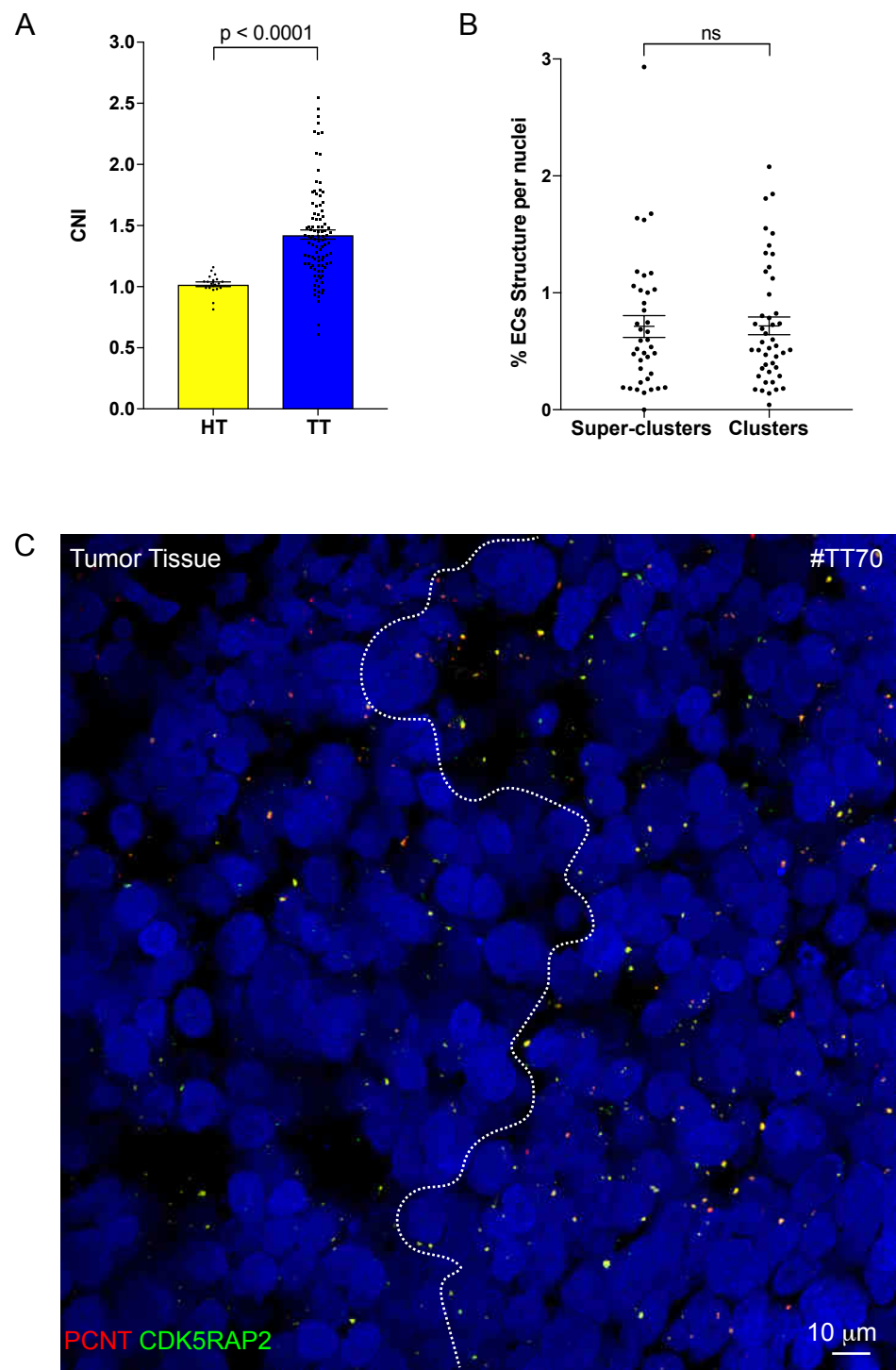


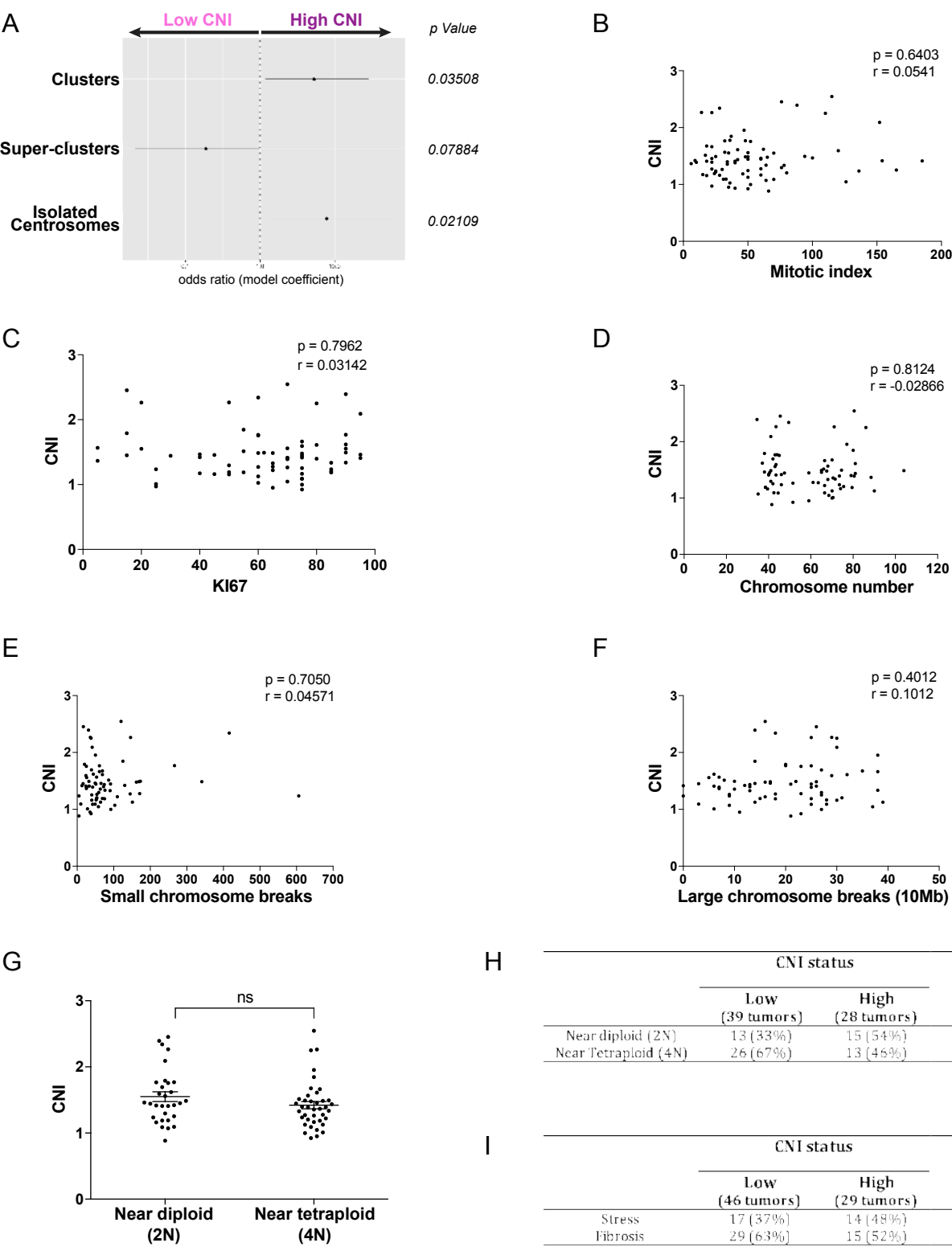
Figure 6 - Morretton et al.



Supplementary Figure 1- Morretton et al.

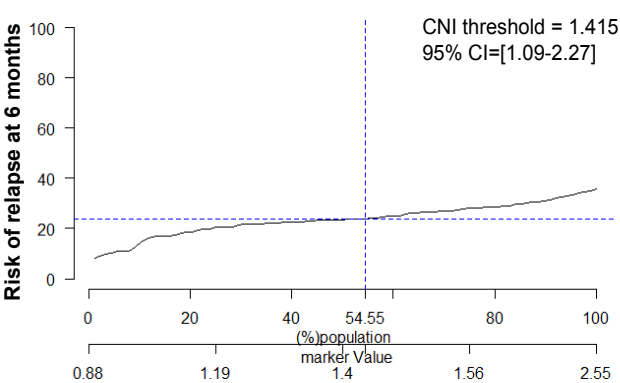


Supplementary Figure 2 - Morretton et al.

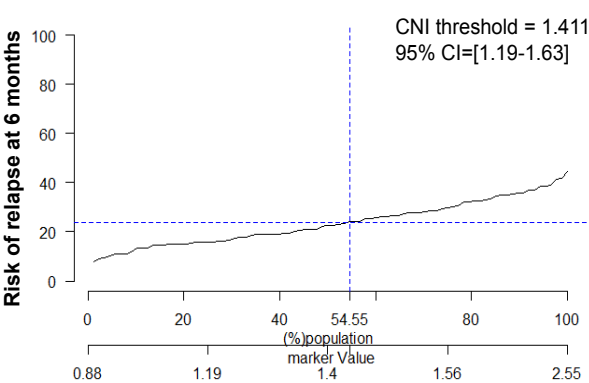


Supplementary Figure 3 - Morretton et al.

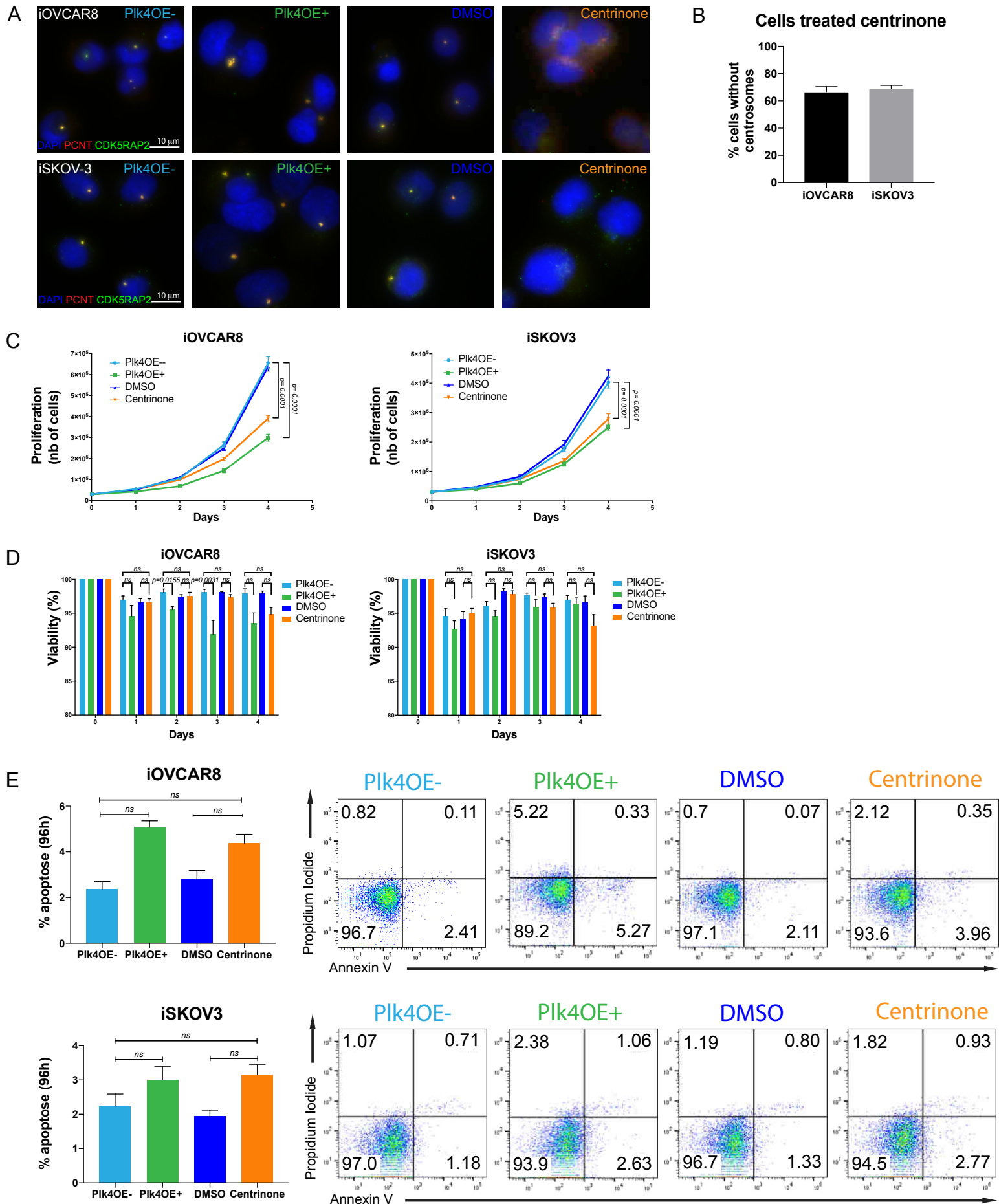
A



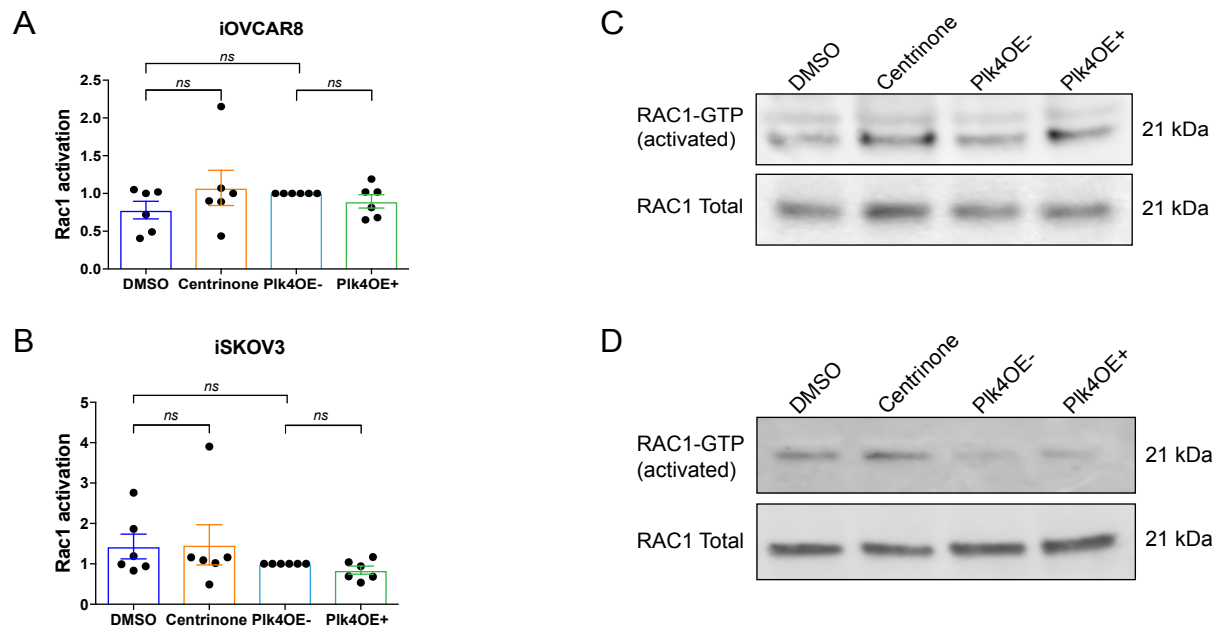
B



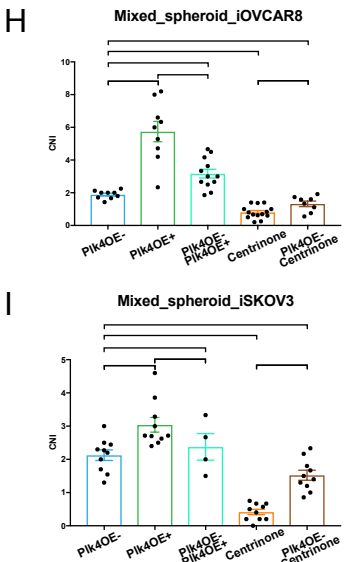
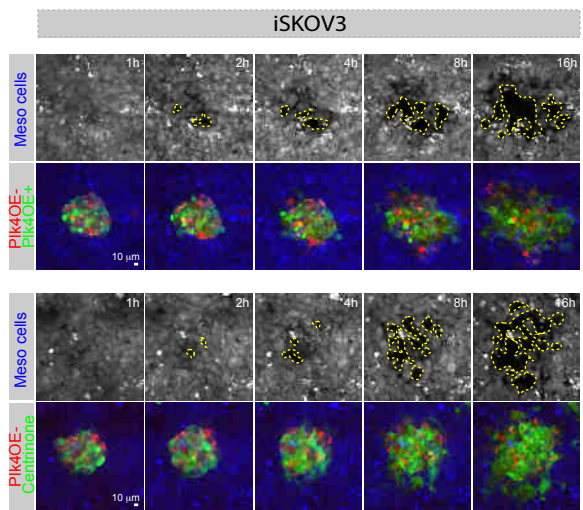
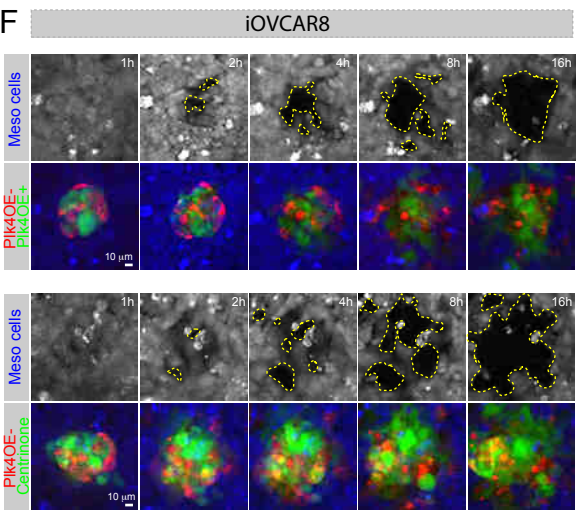
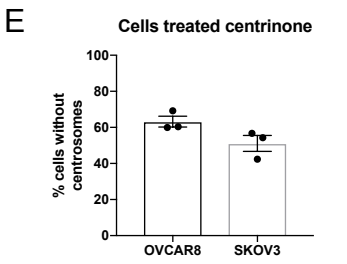
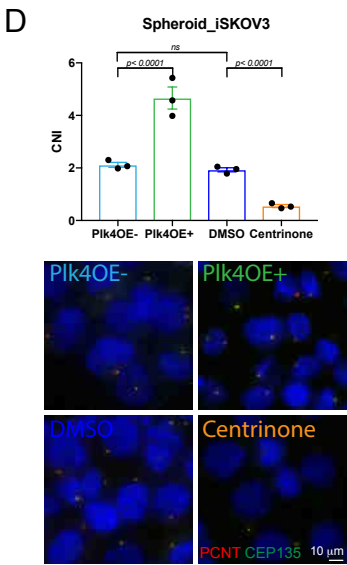
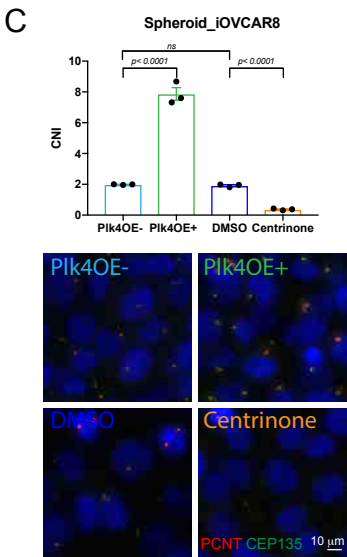
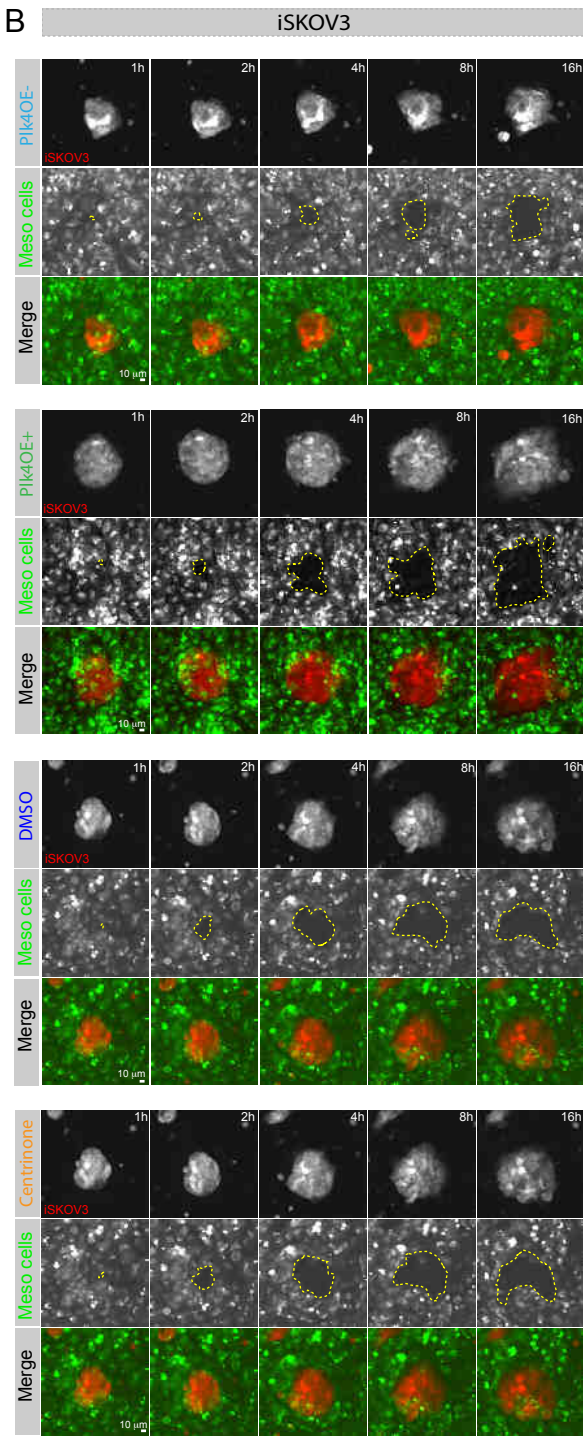
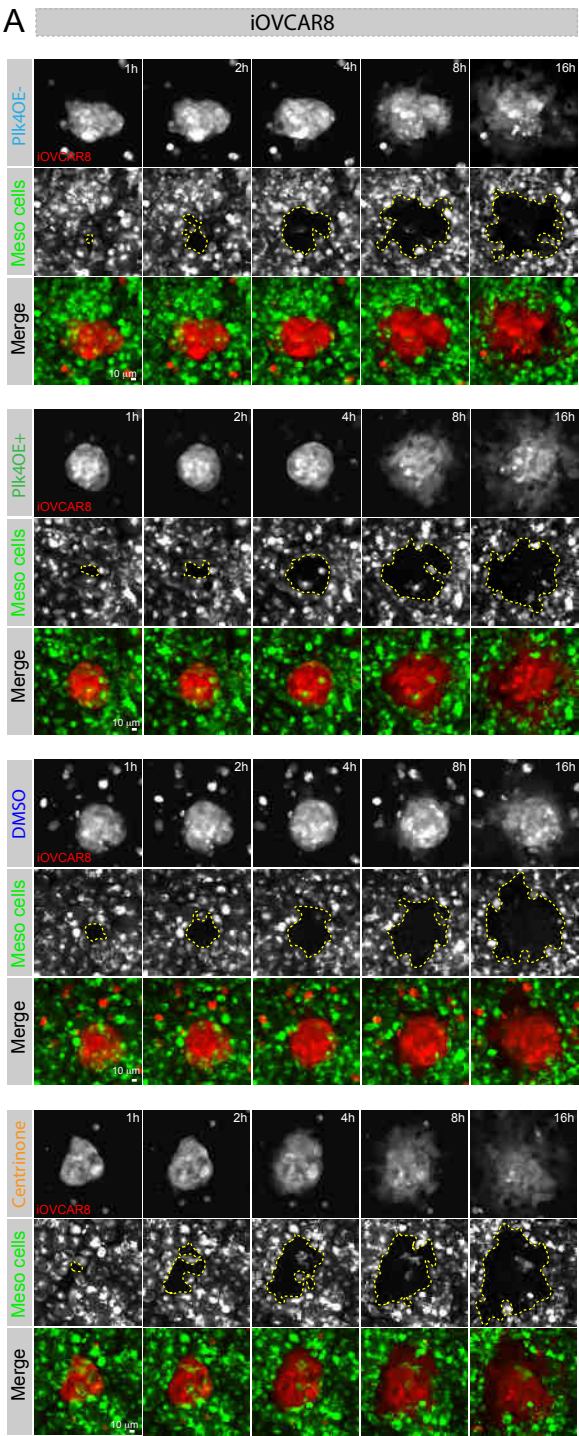
Supplementary Figure 4 - Morretton et al.



Supplementary Figure 5 - Morretton et al.



Supplementary Figure 6 - Morretton et al.



Bibliographie

- Berns, E. M. and Bowtell, D. D. (2012). The changing view of high-grade serous ovarian cancer. *Cancer research*, 72(11) :2701–2704. 56
- Gentric, G., Kieffer, Y., Mieulet, V., Goundiam, O., Bonneau, C., Nemati, F., Hurbain, I., Raposo, G., Popova, T., Stern, M.-H., et al. (2019). Pm-regulated mitochondrial metabolism enhances chemosensitivity in human ovarian cancers. *Cell metabolism*, 29(1) :156–173. 56
- Konstantinopoulos, P. A. and Awtrey, C. S. (2012). Management of ovarian cancer : a 75-year-old woman who has completed treatment. *JAMA*, 307(13) :1420–1429. 56
- Konstantinopoulos, P. A., Ceccaldi, R., Shapiro, G. I., and D’Andrea, A. D. (2015). Homologous recombination deficiency : exploiting the fundamental vulnerability of ovarian cancer. *Cancer discovery*, 5(11) :1137–1154. 56
- Lloyd, K. L., Cree, I. A., and Savage, R. S. (2015). Prediction of resistance to chemotherapy in ovarian cancer : a systematic review. *BMC cancer*, 15(1) :117. 56
- Mateescu, B., Batista, L., Cardon, M., Gruosso, T., De Feraudy, Y., Mariani, O., Nicolas, A., Meyniel, J.-P., Cottu, P., Sastre-Garau, X., et al. (2011). mir-141 and mir-200a act on ovarian tumorigenesis by controlling oxidative stress response. *Nature medicine*, 17(12) :1627–1635. 56
- McIntosh, M. W. and Pepe, M. S. (2002). Combining several screening tests : optimality of the risk score. *Biometrics*, 58(3) :657–664. 60
- Network, C. G. A. R. et al. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature*, 474(7353) :609. 56
- Pepe, M. S., Feng, Z., Huang, Y., Longton, G., Prentice, R., Thompson, I. M., and Zheng, Y. (2008). Integrating the predictiveness of a marker with its performance as a classifier. *Am. J. Epidemiol.*, 167(3) :362–368. 58
- Pujade-Lauraine, E., Hilpert, F., Weber, B., Reuss, A., Poveda, A., Kristensen, G., Sorio, R., Vergote, I., Witteveen, P., Bamias, A., et al. (2014). Bevacizumab combined with

- chemotherapy for platinum-resistant recurrent ovarian cancer : the aurelia open-label randomized phase iii trial. *Obstetrical & Gynecological Survey*, 69(7) :402–404. 56
- Ramalingam, P. (2016). Morphologic, immunophenotypic, and molecular features of epithelial ovarian cancer. *Oncology*, 30(2). 56
- Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., Johnson, D. S., Trivett, M. K., Etemadmoghadam, D., Locandro, B., et al. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical cancer research*, 14(16) :5198–5208. 56
- Viallon, V. and Latouche, A. (2011). Discrimination measures for survival outcomes : connection between the auc and the predictiveness curve. *Biometrical Journal*, 53(2) :217–236. 59
- Zhang, H., Liu, T., Zhang, Z., Payne, S. H., Zhang, B., McDermott, J. E., Zhou, J.-Y., Petyuk, V. A., Chen, L., Ray, D., et al. (2016). Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell*, 166(3) :755–765. 56

Chapitre 6

Quantification du bénéfice à long terme de l'immunothérapie à l'aide de la régression quantile censurée

Sommaire

6.1	Problématique	106
6.2	Modélisation et estimation	106
6.3	Article en révision dans la revue JCO Precision Oncology	108

6.1 Problématique

L'immunothérapie a fait ces derniers temps une entrée remarquée dans l'arsenal thérapeutique en oncologie [Rittmeyer et al., 2017]. Cette nouvelle approche thérapeutique est caractérisée par son bénéfice de survie qui se fait grâce à la durée de réponse prolongée obtenue chez les patients répondeurs. Son effet retardé dans le temps se traduit par une séparation tardive des courbes de survie entre les bras de traitements (par exemple immunothérapie vs chimiothérapie). L'hypothèse de proportionnalité des risques n'est donc pas vérifiée et c'est ce qui a conduit plusieurs auteurs à proposer de nouvelles méthodes pour quantifier le bénéfice de survie à long terme de l'immunothérapie [Horiguchi et al., 2018; Pons-Tostivint et al., 2019; Uno et al., 2014; Vivot et al., 2019]. Ces méthodes reposent pour la plupart du temps sur le bénéfice de survie restreint (Restricted Mean Survival Time) ou sur la médiane de survie qui n'est pas forcément atteinte dans les pathologies considérées.

Dans ce chapitre, nous proposons d'utiliser la régression quantile censurée pour quantifier le bénéfice de survie à long terme de l'immunothérapie. La régression quantile modélise directement les quantiles conditionnels du temps de survie. Plus important encore, la régression quantile permet aux covariables d'avoir des effets différents aux différents quantiles de la distribution de survie et est donc capable de saisir une importante hétérogénéité de la population [Koenker, 2005]. Parmi les méthodes d'estimation des coefficients de la régression quantile censurée, nous utilisons la méthode de Peng and Huang [2008] qui est fondée sur l'estimateur de Nelson-Aalen de la fonction de risque cumulé.

Dans la section suivante, nous détaillons le modèle et l'inférence proposés par Peng and Huang [2008].

6.2 Modélisation et estimation

Soient Z l'indicateur de traitement ($Z = 1$ si le traitement est immuno, $Z = 0$ sinon), T le délai d'événement et C le délai de censure. Soit $\tilde{T} = \min(T, C)$ le délai observé, $\delta = \mathbb{I}_{(T \leq C)}$ l'indicateur d'événement ($\delta = 1$ si l'individu a subi l'événement et $\delta = 0$ sinon). Les données observées consistent en n -triplet iid $(\tilde{T}, \delta, Z)$, noté par $\{\tilde{T}_i, \delta_i, Z_i\}, i = 1, \dots, n$. Nous définissons le quantile conditionnel comme

$$Q_T(\tau|Z) = \inf\{t : P(\tilde{T} \leq t|Z) \geq \tau\}, 0 \leq \tau \leq 1.$$

Nous considérons le modèle de régression quantile censuré suivant

$$Q_T(\tau|Z) = \beta_0(\tau) + \beta_1(\tau)Z$$

où $\beta(\tau) = (\beta_0(\tau), \beta_1(\tau))$ est le vecteur des coefficients inconnus de la régression quantile.

Pour estimer les coefficients $\beta(\tau)$ de la régression quantile censuré, plusieurs méthodes ont été proposées dans la littérature [Peng and Huang, 2008; Portnoy, 2003; Powell, 1986]. Nous utilisons la méthode de Peng and Huang [2008] qui est fondé sur l'estimateur de Nelson-Aalen de la fonction risque cumulé. Nous considérons les notations suivantes : $F(t|Z_i) = P(T_i \leq t|Z_i)$, $\Lambda(t|Z_i) = -\log(1 - F(t|Z_i))$ la fonction de risque cumulé, $N_i(t) = \mathbb{I}\{\tilde{T}_i \leq t, \delta_i = 1\}$ un processus de comptage et $M_i(t) = N_i(t) - \Lambda_i\{t \wedge \tilde{T}_i|Z_i\}$ la martingale associée avec le processus de comptage $N_i(t)$. Cependant, l'espérance du martingale conditionnel est nul, i.e. $\mathbb{E}(M_i(t)|Z_i) = 0$. Ils en déduisent l'équation d'estimation suivante :

$$\mathbb{E}(n^{-1/2} \sum_{i=1}^n x_i [N_i(\mathbf{x}'_i \beta(\tau)) - \int_0^\tau I(\tilde{T}_i \geq \mathbf{x}'_i \beta(\tau)) dH(u)]) = 0 \quad (6.1)$$

où $\mathbf{x}'_i = (1, Z_i)$ and $H(u) = -\log(1 - u)$ pour $u \in [0, 1]$. L'intégrale de l'équation d'estimation 6.1 a été approximé sur une grille $0 = \tau_0 < \tau_1 < \dots < \tau_L < 1$.

Comme indiqué dans Peng and Huang [2008], il est difficile de produire une estimation de la variance asymptotique des estimations par régression quantile, car la matrice de covariance du processus lié à $\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$ implique une fonction de densité inconnue. Nous avons utilisé l'approche du bootstrap pour estimer la variance ainsi que les intervalles de confiance de $\hat{\beta}(\tau) = (\hat{\beta}_0(\tau), \hat{\beta}_1(\tau))$. La procédure est décrite comme suit : nous avons estimé les coefficients $\beta(\tau)$ en utilisant la fonction crq du package quantreg R. Des estimateurs du bootstraps noté $\{\beta_b^*(\tau)\}_{b=1}^B$ pour chaque quantile τ , sont obtenus en répétant cette procédure $B = 1000$ fois. La variance de $\beta(\tau)$ peut être estimée par la variance de l'échantillon de $\{\beta_b^*(\tau)\}_{b=1}^B$. Les intervalles de confiance pour $\beta(\tau)$ sont basés sur les percentiles empiriques de $\{\beta_b^*(\tau)\}_{b=1}^B$. Cela signifie que la limite supérieure est le quantile à 97.5% de l'effet provenant du ré-échantillonnage et la limite inférieure est le quantile à 2.5%. Pour illustrer ce travail, nous avons reconstruit des données d'essai randomisé du cancer du poumon à partir des courbes de Kaplan Meier publiés en utilisant l'algorithme de Guyot et al. [2012].

Lors du reviewing de cet article, un re-lecteur a suggéré d'effectuer les diagnostics d'adéquation pour la régression quantile. Dans ce contexte, nous avons étudié la linéarité globale entre la covariable et le quantile conditionnel du modèle en utilisant les résidus du martingale. En effet, nous avons représenté les résidus de martingale i.e $N_i(t) - \Lambda_i\{t \wedge \tilde{T}_i|Z_i\}$ (en ordonnée) en fonction de la valeur du quantile ajusté i.e $\hat{Q}_T(\tau|Z) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)Z$ (en abscisse) pour vérifier graphiquement les hypothèses de linéarité du modèle de régression de quantile censuré. La Figure 6.1 montre que les résidus sont négatifs lorsque la valeur du quantile ajusté est faible, et positifs lorsque la valeur du quantile ajusté est importante. Une non-linéarité modérée est présente, en particulier pour les quantiles inférieurs et supérieurs. Dans la figure 1 de l'article ci-dessous, la courbe de survie se stabilise au niveau du quantile 0.72 sur le bras immunothérapie. Cela signifie que le quantile conditionnel dans ce bras de traitement

serait estimé à l'infini pour tout niveau de quantile supérieur à 0.72. Par conséquent, nous nous concentrons sur les niveaux de quantile de 0.1 à 0.60. Toutefois, au niveau du quantile, $\tau = 0.6$, l'hypothèse de linéarité est valide. Ainsi, la linéarité globale ne tient pas mais tient pour une plage de quantiles. Nous avons été surpris par l'absence de mise en œuvre d'une telle qualité d'ajustement.

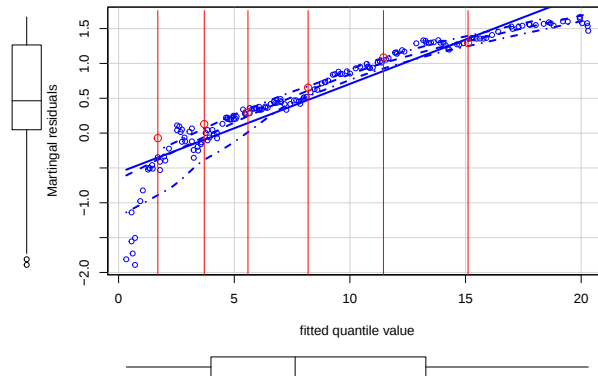


FIGURE 6.1 – Scatter plot des résidus de martingale par rapport aux valeurs de quantile ajustées. Les lignes verticales en rouges correspondent aux déciles (de 10% à 60%)

Dans la section suivante est présenté l'article principal provenant de ce travail

6.3 Article en révision dans la revue JCO Precision Oncology

Insights for quantifying the long-term benefit of immunotherapy using quantile regression

Bassirou Mboup^{*1,2}, Christophe Le Tourneau^{1,3}, and Aurélien Latouche^{1,2}

¹Institut Curie, PSL Research University, INSERM, U900, F-92210

²Conservatoire National des Arts et Métiers

³Department of Drug Development and Innovation (D3i), Institut Curie, Paris, France

November 5, 2020

Abstract

Purpose

Immunotherapy have been approved to treat many tumor types. However, one characteristic of this therapeutic class is that survival benefit is due to late immune response, which leads to a delayed treatment effect. Quantifying the benefit, if any, of such treatment, will thus require other metrics than the usual hazard ratio and different approaches have been proposed to quantify long-term response of immunotherapy.

Method

In this paper, we suggest to employ quantile regression for survival data to quantify the long-term benefit of immunotherapy. Our motivation is that this approach is not trial-specific and provides clinically understandable results without specifying arbitrary time points or the necessity to reach median survival, as it is the case with other methods. We use reconstructed data from publish Kaplan Meier curves to illustrate our method.

Result

On average, patients from the immunotherapy group have 60% of chance to survive 5.46 months (95% CI: (2.57; 9.02)) more than patients in the chemotherapy group.

^{*}Corresponding author: Bassirou Mboup
E-mail: bassirou.mboup@curie.fr

1 Introduction

Immunotherapy has been demonstrated to improve overall survival (OS) in several tumor types [1]. However, one characteristic of this new therapeutic class is that the improved OS is mainly due to the *durable* responses obtained in a minority of patients. Indeed, treatment effect may depend on time caused by late immune response. This delayed effect of immunotherapy results in a late separation of the survival curves. Thus, the usual assumption of proportional hazard between treatment arms does not hold anymore. This has led to debate on which was the more appropriate way to handle this non-proportionality [2].

We previously defined long-term responder patients as patients having an OS exceeding twice the median OS of the whole patient population. In a meta-analysis on aggregated data, the proportion of long-term responder patients using this criterion was significantly higher in patients treated with immunotherapy than those treated with other anticancer drugs [3]. Yet, our definition was arbitrary similar to other proposals that we detail thereafter. To quantify long-term survival benefit the area under the Kaplan-Meier curve from the time of separation (also termed *milestone*) to a given time-horizon was proposed [4]. Patients receiving immunotherapy lived in average 2.5 months longer than patients treated with chemotherapy. The area under the Kaplan-Meier curves is 8.3 and 10.8 within the chosen window of 36 to 72 months for chemotherapy and immunotherapy respectively. [5] used the ratio of the late-life expectancy, defined as the ratio of the area under the Kaplan-Meier OS curve from the median survival time to the end of follow-up in each arm, to quantify the long-term benefit of immunotherapy. These approaches are adequate when the proportional hazards assumption does not hold, but they require an arbitrary choice of a landmark time, a quantile or focus on median survival times, which is not always defined.

To complement these proposals, we propose to focus on the entire conditional distribution of survival times and estimate the effect of immunotherapy on each quantile of the survival times using censored quantile regression model [6]. We suggest to use quantile regression to identify the upper quantile of the survival curves and to assess the treatment effect on the difference scale. Notably, [7] suggested to employ the ratio or difference of percentile of the survival function between treatment arm. In fact, difference of percentile of the survival function, is in the time scale, facilitating the quantification of the benefit of a treatment arm over another. However, the method of [7] requires a choice of an arbitrary quantile. Censored quantile regression is an alternative to simply modeling conditional mean that can model the quantile of the response variable as a linear function of covariates and allows distinguishing among differential effects of covariates across conditional quantiles and thus is able to capture important population heterogeneity [8].

The remainder of this paper is organized as follows. First, we provide details of the quantile regression model for right censored data, the inference and the goodness of fit. Next, we illustrate the methodology with a reconstructed data set from a randomized trial in patients with non-small cell lung cancer and present the results of this application. We close with a discussion. The supplementary, which contains R code, reconstructed data, is available at <https://github.com/BMBOUP/immuno>.

2 Censored quantile regression

The τ -th quantile of the OS is the smallest time where the OS exceeds τ . With explanatory covariate, Z , the τ -th conditional quantile of T given Z , is $Q_T(\tau|Z) = \inf\{t|P(T \leq t|Z) \geq \tau\}$. The quantile regression model relates $Q_T(\tau|Z)$ linearly to Z for each $0 < \tau < 1$.

$$Q_T(\tau|Z) = \beta_0(\tau) + \beta_1(\tau) Z \quad (1)$$

where $\beta_0(\tau)$, $\beta_1(\tau)$ are respectively the intercept and the effects of the treatment on the τ -th quantile of T . With a binary covariate denoting the treatment arm (e.g. $Z=1$ for immunotherapy) the difference $Q_T(\tau|Z = 1) - Q_T(\tau|Z = 0) = \beta_1(\tau)$ will be our target of interest for quantifying the benefit, if any, of immunotherapy over standard of care.

In practice, T is not always observable because loss of follow-up. To account for the right censoring, two approaches were considered for the inference of model (1). [9] developed, under random censoring, a recursively reweighted estimation procedure which a generalization of the Kaplan-Meier estimator. [6] developed, under the weaker assumption of conditionally independent censoring, a method relying on the Nelson-Aalen estimator using counting processes and martingale theory. These two approaches have been implemented in the R package `quantreg` [10].

In the sequel, we will employ the Peng and Huang's estimator since their approach enables to conduct hypothesis testing, resampling inference and model diagnosis. In fact, testing the null hypothesis $H_0 : \beta_1(\tau) = 0$, $\tau \in [0; 1]$, is of interest in our setting since it corresponds to no benefit of the experimental arm at the τ -th quantile. Furthermore, one strong assumption of the model (1) is the global linear relationship between conditional quantile and the covariates. Following [6], the assumption of the global linearity between covariates and conditional quantile of the model can be assess using the martingale residuals. The plot of the fitted quantile value (X-axis) versus martingale residuals (Y-axis) are next used to visually check the assumptions of the linearity. We

were surprised by the lack of implementation of such graphical procedure. We thus provide a R function in the reproducible code.

The technical derivations of model (1) are deferred to the supplementary.

Finally, we should mention that in the case of a single binary covariate and a given quantile, the quantile regression is equivalent to the non-parametric approach that would compare the quantile of the survival function between two groups [11, 12]. This approach is currently implemented in the R package `surv2sampleComp` [13].

3 Illustration and results

We used the published Kaplan-Meier survival curves to simulate individual failure times. The algorithm developed by [14] enable to emulate the survival time from the two arms of the clinical trial of [1] that comprises 850 patients with metastatic non-small-cell lung cancer that compared immunotherapy versus chemotherapy. The reliability of this method as quantified by the hazard ratio (HR) is illustrated in Figure 1. The estimated HR from reconstructed data is compatible with the published HR. The primary endpoint was overall survival compared between treatment groups. On the reconstructed data, the average follow-up was 8.78 months, and during the follow-up, 555 deaths were observed leading to approximately 35% of right censored observations. This censoring rate is consistent with the 30% observed rate in the actual trial. Usually, the 0.5 quantiles of the survival (e.g. median) are reported in each arms : 14.13 months in the experimental arm and 9.78 months in the control arm.

In Figure 1, the survival curve levels off around the 0.72 quantile for the immunotherapy group which corresponds to 24 months. This means that the conditional quantile in the immunotherapy group at any level higher than 0.72 would be estimated to be infinity. Therefore, we focus on the quantile levels from 0.1 to 0.60. In our study, the upper bound of the *estimable* quantile of survival time is 60 % which corresponds to 18.15 months in the immunotherapy arm and 12.69 months in the chemotherapy arm. At the quantile level 0.1, we notice that the coefficient is almost null. Due to the overlap of the survival curves at that level of the survival time. We then estimate for each available quantile, the benefit in months of immunotherapy treatment using censored quantile regression.

Differences between treatment arms can be plotted at increasing quantile, for the illustrative purpose of a potential benefit. Yet, this plot does not account for multiple testing issue. We thus recommend to focus on a single attainable quantile (60% in our motivating example). The model diagnostic reveals that a moderate non-linearity is present, especially for lower and upper quantiles

(Figure is available in the supplementary). Yet, for the quantile of interest, e.g. $\tau = 0.6$, the linearity assumption holds. Thus, at the quantile level 0.6 of the survival time, patients treated with immunotherapy have a survival average benefit of 5.46 months compared to patients treated with chemotherapy (95% bootstrap confidence interval: (2.57; 9.02)). In other words, on average, patients from the immunotherapy group have 60% of chance to survive 5.46 months (95% CI: (2.57; 9.02)) more than patients in the chemotherapy group.

At quantile 0.6, the estimated benefit obtained with `surv2sample()` is similar to that of the quantile regression. To illustrate the usefulness of the quantile regression, we consider the setting where in addition to the effect of treatment on the quantile of the survival, the effect of the expression of the biomarker PD-L1 is investigated as well as their interaction. In fact, the hazard ratio of the overall survival by PD-L1 gene expression was estimated at 0.69 (95%CI: (0.57; 0.83)) in the supplementary Figure 2 of [1]. The details of the simulation are deferred to the supplementary and the estimated coefficients are summarized in the table 1. At the quantile, $\tau = 0.6$ the estimated effect for immunotherapy is 1.73 (95% CI: (0.40; 3.31)) indicates that immunotherapy patients with a PD-L1 value of 0 have, on average 60 % of chance to survive 1.73 months more than chemotherapy patients with the same PD-L1 value. Furthermore, patients treated with immunotherapy and with an increase of PD-L1 expression equal to 1 have, on average 60% of chance to survive 8.28 months $(1.73 + (6.55 \times 1))$.

4 Discussion

There had been a considerable effort to quantify the long-term survival in the presence of delayed or reverse treatment effect, relying notably on 2-samples comparison of median survival time or restricted mean survival time [7, 15, 4, 5].

In this work, we illustrate on a single randomized trial with reconstructed individual failure time, the merits of the quantile regression for survival data. This is a readily available tool for quantifying the long-term survival benefit of immunotherapy and provides clinically meaningful results without specifying arbitrary time points or the necessity to reach median survival. Furthermore, it informs of the treatment effect on the span of the OS. One appeal of this method is that it allows to have effect that are quantile-specific. Of note, our illustrative example, did not exhibited a plateau, where a survival curve level off. The quantile regression, can be useful to estimates the quantile that announce such a plateau.

The proposed quantile regression relates the outcome linearly to the explanatory variables. This can be a limitation and this assumption should be checked. Moreover, one of the limits of the

censored quantile regression is that it may fail when the censoring rate is high.

To improve the level of evidence of the benefit of immunotherapy over chemotherapy, it would be of interest, to investigate how to conduct a meta-analysis on individual patient data that would synthesis the treatment effect at the trial level on the difference scale using quantile regression. A first route would be to consider a fixed-effect approach that would weight properly the benefit of each trial. Yet, quantile regression for clustered censored data is not available. Thus a one stage inference, accounting for the correlation at the trial level, is not feasible for this regression model.

Conflict of interest statement

None declared.

References

- [1] Achim Rittmeyer, Fabrice Barlesi, Daniel Waterkamp, Keunchil Park, Fortunato Ciardiello, Joachim Von Pawel, Shirish M Gadgil, Toyooki Hida, Dariusz M Kowalski, Manuel Cobo Dols, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (oak): a phase 3, open-label, multicentre randomised controlled trial. *The Lancet*, 389(10066):255–265, 2017.
- [2] Bo Huang. Some statistical considerations in the clinical development of cancer immunotherapies. *Pharmaceutical statistics*, 17(1):49–60, 2018.
- [3] Elvire Pons-Tostivint, Aurélien Latouche, Pauline Vafard, Francesco Ricci, Delphine Loirat, Ségolène Hescot, Marie-Paule Sablin, Roman Rouzier, Maud Kamal, Claire Morel, et al. Comparative analysis of durable responses on immune checkpoint inhibitors versus other systemic therapies: a pooled analysis of phase iii trials. *JCO Precision Oncology*, 3:1–10, 2019.
- [4] Miki Horiguchi, Lu Tian, Hajime Uno, SuChun Cheng, Dae Hyun Kim, Deb Schrag, and Lee-Jen Wei. Quantification of long-term survival benefit in a comparative oncology clinical study. *JAMA oncology*, 4(6):881–882, 2018.
- [5] Alexandre Vivot, Perrine Créquit, and Raphaël Porcher. Use of late-life expectancy for assessing the long-term benefit of immune checkpoint inhibitors. *JNCI: Journal of the National Cancer Institute*, 111(5):519–521, 2019.
- [6] Limin Peng and Yijian Huang. Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649, 2008.

- [7] Hajime Uno, Brian Claggett, Lu Tian, Eisuke Inoue, Paul Gallo, Toshio Miyata, Deborah Schrag, Masahiro Takeuchi, Yoshiaki Uyama, Lihui Zhao, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of clinical Oncology*, 32(22):2380, 2014.
- [8] Roger Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005.
- [9] Stephen Portnoy. Censored regression quantiles. *Journal of the American Statistical Association*, 98(464):1001–1012, 2003.
- [10] Roger Koenker. *quantreg: Quantile Regression*, 2019. R package version 5.54.
- [11] L. Tian, L. Zhao, and L.J. Wei. Predicting the restricted mean event time with the subject’s baseline covariates in survival analysis. *Biostatistics*, 15(2):222–233, 2014.
- [12] Lihui Zhao, Lu Tian, Hajime Uno, Scott D Solomon, Marc A Pfeffer, Jerald S Schindler, and Lee Jen Wei. Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study. *Clinical trials*, 9(5):570–577, 2012.
- [13] Lu Tian, Hajime Uno, and Miki Horiguchi. *surv2sampleComp: Inference for Model-Free Between-Group Parameters for Censored Survival Data*, 2017. R package version 1.0-5.
- [14] Patricia Guyot, AE Ades, Mario JNM Ouwens, and Nicky J Welton. Enhanced secondary analysis of survival data: reconstructing the data from published kaplan-meier survival curves. *BMC medical research methodology*, 12(1):9, 2012.
- [15] Tai-Tsang Chen. Statistical issues and challenges in immuno-oncology. *Journal for immunotherapy of cancer*, 1(1):18, 2013.

$\tau = 0.6$	Value	Lower Bd	Upper Bd	Std Error	T Value	P-value
(Intercept)	1.43	-0.31	2.56	0.73	1.95	0.05
Treat	1.73	0.41	3.31	0.74	2.34	0.01
PDL1	1.92	1.51	2.32	0.21	9.27	0.00
Treat:PDL1	6.56	5.53	7.33	0.46	14.23	0.00

Table 1: Estimated coefficient for the quantile regression at the quantile 0.6 with treatment, PD-L1 and treatment by PD-L1 expression interaction

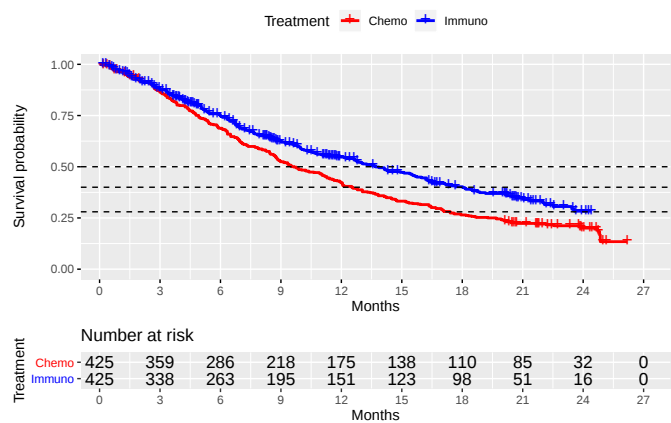


Figure 1: Reconstructed Kaplan-Meier curve. The dotted lines are the median, the 0.6 and the upper quantile of survival time in the immunotherapy arm.

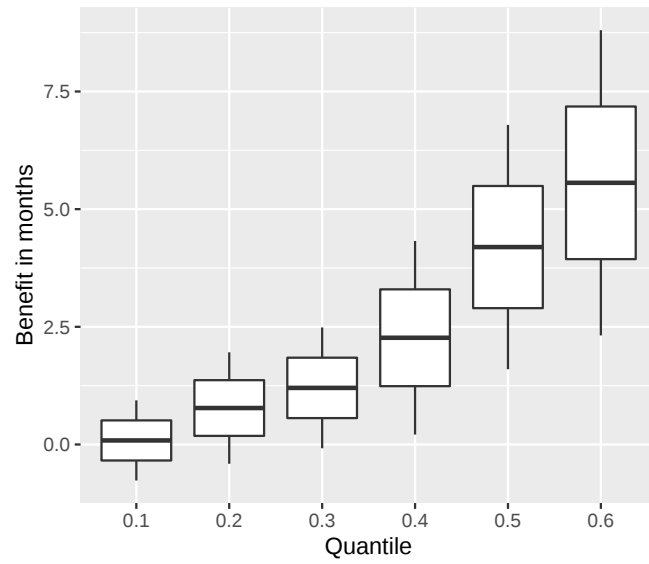


Figure 2: Boxplot of survival benefit (e.g difference of quantile between arms) in months for each quantile available

Bibliographie

- Guyot, P., Ades, A., Ouwers, M. J., and Welton, N. J. (2012). Enhanced secondary analysis of survival data : reconstructing the data from published kaplan-meier survival curves. *BMC medical research methodology*, 12(1) :9. 107
- Horiguchi, M., Tian, L., Uno, H., Cheng, S., Kim, D. H., Schrag, D., and Wei, L.-J. (2018). Quantification of long-term survival benefit in a comparative oncology clinical study. *JAMA oncology*, 4(6) :881–882. 106
- Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press. 106
- Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482) :637–649. 106, 107
- Pons-Tostivint, E., Latouche, A., Vaflard, P., Ricci, F., Loirat, D., Hescot, S., Sablin, M.-P., Rouzier, R., Kamal, M., Morel, C., et al. (2019). Comparative analysis of durable responses on immune checkpoint inhibitors versus other systemic therapies : a pooled analysis of phase iii trials. *JCO Precision Oncology*, 3 :1–10. 106
- Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, 98(464) :1001–1012. 107
- Powell, J. L. (1986). Censored regression quantiles. *Journal of econometrics*, 32(1) :143–155. 107
- Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., Von Pawel, J., Gadgeel, S. M., Hida, T., Kowalski, D. M., Dols, M. C., et al. (2017). Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (oak) : a phase 3, open-label, multicentre randomised controlled trial. *The Lancet*, 389(10066) :255–265. 106
- Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T., Schrag, D., Takeuchi, M., Uyama, Y., Zhao, L., et al. (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of clinical Oncology*, 32(22) :2380. 106

Vivot, A., Créquit, P., and Porcher, R. (2019). Use of late-life expectancy for assessing the long-term benefit of immune checkpoint inhibitors. *JNCI : Journal of the National Cancer Institute*, 111(5) :519–521. 106

Chapitre 7

Conclusion et perspectives

Résumé des travaux

L'un des objectifs de la médecine personnalisée est de prendre des décisions thérapeutiques fondées sur la mesure d'un biomarqueur. Il est donc souvent intéressant d'évaluer dans quelle mesure un biomarqueur peut prédire la réponse à un traitement. Les biomarqueurs prédictifs de la réponse au traitement peuvent potentiellement améliorer les résultats cliniques et diminuer les coûts médicaux en permettant le traitement uniquement aux personnes susceptibles d'en bénéficier.

Dans une première partie, nous nous sommes intéressé à l'évaluation de la capacité pronostique et prédictif d'un biomarqueur continu ainsi que la détermination du seuil optimal en présence de données censurées. Une méthodologie populaire consiste à utiliser un modèle de régression et à tester l'interaction entre le traitement et le biomarqueur. Cependant, l'existence d'une interaction n'est pas suffisante pour qu'un biomarqueur soit prédictif. Par conséquent, l'utilisation des courbes de prédictivité du biomarqueur par bras de traitement a été recommandée. En plus d'évaluer dans quelle mesure le biomarqueur prédit la réponse au traitement, il peut aider à définir un seuil optimal pour guider les décisions de traitement. Cette courbe représente le risque d'une réponse binaire en fonction des quantiles du biomarqueur, pour chaque groupe de traitement. Pour estimer cette courbe, des méthodes qui supposent une réponse binaire ou qui s'appuient sur un modèle de risque proportionnel pour données censurées ont été proposées. Dans notre travail, nous avons proposé quelques extensions pour les données censurées. Elles s'appuient sur un modèle logistique dépendant du temps et nous avons proposer d'estimer ce modèle via une pondération par l'inverse de la probabilité de censure (IPCW) pour définir des estimateurs consistants. Les résultats ont suggéré qu'un grand nombre d'événements doivent être observés pour définir un seuil avec une précision suffisante pour qu'il soit utile cliniquement. Si l'effet du traitement varie avec

le temps, le seuil optimal dépend aussi du temps.

D'autre part, nous avons proposé un test de l'hypothèse de calibration fondé par des pondérations de l'inverse de la probabilité de censure pour tenir compte des taux de censure élevés. La calibration du modèle de prédiction pour la détermination du seuil optimal est très important car nous avons montré qu'un modèle mal calibré fournit un seuil erroné et conduit à une mauvaise classification des patients.

Par ailleurs, notre contribution dans l'évaluation d'un nouveau biomarqueur pronostique pour le cancer de l'ovaire a également montré l'utilité de la courbe de prédictivité. L'utilisation de cette courbe a permis de déterminer un seuil optimal du biomarqueur et de classer les patients à haut risque de rechute du cancer de l'ovaire. La compréhension de cette courbe au niveau des biologistes a été très facile.

Dans une seconde partie, nous nous sommes intéressé à la quantification du bénéfice à long terme de l'immunothérapie. Il existe de nombreuses approches dans la littérature. Nous avons proposé d'utiliser le modèle de régression quantile censuré pour quantifier ce bénéfice. Parmi les méthodes d'estimation des coefficients de ce modèle, nous avons utilisé la méthode Peng and Huang [2008] qui est fondé sur l'estimateur de Nelson Aalen. Les résultats dans notre application ont suggéré un bénéfice de 5 mois aux patients traités par immunothérapie pour le quantile d'ordre 60%.

Perspectives

Les méthodes développées dans le chapitre 3 font l'hypothèse que les données sont issues d'un essai randomisé. Étendre notre méthode à des données observationnelles serait un plus pour étudier la capacité prédictive de biomarqueur en vie réelle. Des méthodes de pondération par l'inverse de la probabilité d'être traité (IPTW) pourraient être envisagées et couplées par la pondération par l'inverse de la probabilité de censure (IPCW) qui avait été employée pour tenir compte de la censure et de définir des estimateurs consistants. Pour illustrer nos développements, nous disposons de données de 3310 patientes atteintes d'un cancer du sein, évaluant la capacité pronostique du marqueur de prolifération Ki67 et de l'index mitotique vis à vis de la rechute [Rossi et al., 2015].

Récemment, Blangero et al. [2020a] utilisent la différence de l'aire qui sépare les deux courbes ROC spécifiques au bras de traitement pour évaluer et tester la capacité prédictive d'un biomarqueur continu et proposent une méthode bayésienne en passant par une fonction d'utilité pour estimer le seuil du biomarqueur [Blangero et al., 2020b]. Un travail intéressant à mener serait de déterminer le seuil du biomarqueur en fonction du ratio entre le coût du

traitement et le coût de la maladie à différent horizon de prédiction. Nous comptons nous appuyer sur le travail de Huang et al. [2015] qui avait développé une nouvelle mesure appelée bénéfice attendu qui étend la mesure du bénéfice net pour quantifier la réduction de la somme des coûts de la maladie et du traitement en utilisant le biomarqueur.

Pour améliorer le niveau de preuve du bénéfice de l'immunothérapie par rapport à la chimiothérapie développé au chapitre 6, il serait intéressant d'étudier comment mener une méta-analyse sur les données individuelles des patients qui synthétiserait l'effet du traitement au niveau de l'essai sur l'échelle de différence en utilisant la régression quantile. Une première voie consisterait à envisager une approche à effet fixe qui pondérerait correctement le bénéfice de chaque essai.

Bien que certains outils graphiques aient été proposés pour le diagnostic de la régression quantile [Buuren, 2007], le diagnostic de la régression quantile censurée est encore très peu développé. La conception d'outils efficaces de diagnostic de modèles pour la régression quantile censurée mérite des recherches plus approfondies.

Conclusion générale

Dans cette thèse, nous avons étendu une méthode d'évaluation de la capacité prédictive d'un biomarqueur et proposé un outil pour la quantification du bénéfice à long terme d'un traitement adapté aux données censurées. L'ensemble de notre travail a été implémenté en R. La diffusion des méthodes et simulations développées sont disponibles sur https://github.com/BMBOUP/Optimal_threshold et <https://github.com/BMBOUP/immuno>. Ceci améliore la reproductibilité de nos travaux et leur ré-emploi par exemple sur un essai confirmatoire prospectif.

Bibliographie

- Blangero, Y., Rabilloud, M., Ecochard, R., and Subtil, F. (2020a). A bayesian method to estimate the optimal threshold of a marker used to select patients' treatment. *Statistical Methods in Medical Research*, 29(1) :29–43. 122
- Blangero, Y., Rabilloud, M., Laurent-Puig, P., Le Malicot, K., Lepage, C., Ecochard, R., Taieb, J., and Subtil, F. (2020b). The area between roc curves, a non-parametric method to evaluate a biomarker for patient treatment selection. *Biometrical Journal*. 122
- Buuren, S. v. (2007). Worm plot to diagnose fit in quantile regression. *Statistical Modelling*, 7(4) :363–376. 123
- Huang, Y., Laber, E. B., and Janes, H. (2015). Characterizing expected benefits of biomarkers in treatment selection. *Biostatistics*, 16(2) :383–399. 123
- Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482) :637–649. 122
- Rossi, L., Laas, E., Mallon, P., Vincent-Salomon, A., Guinebretiere, J., Lerebours, F., Rouzier, R., Pierga, J., and Rey, F. (2015). Prognostic impact of discrepant ki67 and mitotic index on hormone receptor-positive, her2-negative breast carcinoma. *British journal of cancer*, 113(7) :996–1002. 122

Annexe A

Annexes

A.1 Code R pour simulation de données chapitre 2

```
rm(list=ls())

# {{{ Parametres des simulations
# Parametres
n <- 1000 # taille de l'échantillon
mu1 <- -4.8 # moyenne du marqueur Y2
sd1 <- 1.8 # écart type du marqueur Y2
myseed <- 123456 # générateur

# coefficients du modèle logistique avec covariable Y2, trt
beta0 <- -1.24 # L'intercepte du modèle logistique
beta1 <- -0.29 # l'effet du traitement
beta2 <- 0.6 # l'effet du marqueur Y2
beta3 <- -1.5 # l'effet de l'interaction entre le traitement et le marqueur Y2
# }}}

# {{{ Calcul des probas p1, p2, p3, p4 pour l'étape 1 de l'algorithme

# {{{ verifications

### La probabilité marginale  $P(D(0)=1)=\int P(D(1)=1|Y_2=y_2)f(y_2)dy_2$ 
# qui doit être égale à 0.24
F1 <- function(x){(exp(beta0+beta2*x)/(1+exp(beta0+beta2*x)))*dnorm(x)}
```

```

# fonction pour calculé l'intégrale
Pa <- integrate(F1,lower=-20,upper=20)$value
Pa  # OK : 0.24

### La probabilité marginale  $P(D(1)=1)=\int_{-20}^{20} P(D(1)=1|Y_2=y_2)f(y_2)dy_2$ 
## qui doit etre egale a 0.21

F2 <-function(x){(exp(beta0+beta1+(beta2+beta3)*x)/
(1+exp(beta0+beta1+(beta2+beta3)*x)))*dnorm(x)}
Pb <- integrate(F2,lower=-20,upper=20)$value
Pb  # OK : 0.21
# }}}

# {{{ calculs des p1,...,p4

## La probabilité marginale  $P(D(0)=0,D(1)=0)=\int_{-20}^{20} P(D(0)=0,D(1)=0|Y_2=y_2)f(y_2)dy_2$ 

F5 <-function(x){(1-(exp(beta0+beta2*x)/(1+exp(beta0+beta2*x))))*
(1-(exp(beta0+beta1+(beta2+beta3)*x)/(1+exp(beta0+beta1+(beta2+beta3)*x))))*dnorm(x)}

P1 <- integrate(F5,lower=-20,upper=20)$value
P1  # 0.59

## La probabilité marginale  $P(D(0)=0,D(1)=1)=\int_{-20}^{20} P(D(0)=0,D(1)=1|Y_2=y_2)f(y_2)dy_2$ 
F3 <-function(x){(1-(exp(beta0+beta2*x)/(1+exp(beta0+beta2*x))))*
(exp(beta0+beta1+(beta2+beta3)*x)/(1+exp(beta0+beta1+(beta2+beta3)*x)))*dnorm(x)}
P2 <- integrate(F3,lower=-20,upper=20)$value
P2  # 0.17

### La probabilité marginale  $P(D(0)=1,D(1)=0)=\int_{-20}^{20} P(D(0)=1,D(1)=0|Y_2=y_2)f(y_2)dy_2$ 
F4 <-function(x){(exp(beta0+beta2*x)/(1+exp(beta0+beta2*x)))*
(1-(exp(beta0+beta1+(beta2+beta3)*x)/
(1+exp(beta0+beta1+(beta2+beta3)*x))))*dnorm(x)}

```

```

P3 <- integrate(F4,lower=-20,upper=20)$value
P3 # 0.20

## La probabilité marginale  $P(D(0)=1,D(1)=1)=\int_{-\infty}^{\infty} P(D(0)=1,D(1)=1|Y_2=y_2)f(y_2)dy_2$ 

F6 <-function(x){(exp(beta0+beta2*x)/(1+exp(beta0+beta2*x)))*
  (exp(beta0+beta1+(beta2+beta3)*x)/(1+exp(beta0+beta1+(beta2+beta3)*x)))*dnorm(x)}

P4 <- integrate(F6,lower=-20,upper=20)$value
P4 # 0.04

# verifions que la somme vaut 1
P1 + P2 + P3 + P4 # OK
# }}}

# }}}

# {{{ Simulation (une base)

set.seed(myseed) # graine pour des simulations reproductibles

# On suit les etapes de l'algo du document de Bassirou

# {{{ Etape 1
d <- t(rmultinom(n,size=1, p=c(P1,P2,P3,P4)))
head(t(d))

rownames(d) <- paste("subject",1:n)
colnames(d) <- c("D(0)=D(1)=0","D(0)=0 & D(1)=1","D(0)=1 & D(1)=0","D(0)=1 & D(1)=1")
head(d)

D1 <- d[, "D(0)=0 & D(1)=1"] + d[, "D(0)=1 & D(1)=1"]
D0 <- d[, "D(0)=1 & D(1)=0"] + d[, "D(0)=1 & D(1)=1"]

```

```

# Verification (avec n grand !)
data.frame(observe=c(mean(d[,1]),mean(d[,2]),mean(d[,3]),mean(d[,4])),
               theorique=c(P1,P2,P3,P4))
mean(D1)
mean(D0)

# }}}

# {{{ Etape 2

##### Calcul de la densité conditionnelle f(Y/D(0)=d0,D(1)=d1) #####

densitey1 <- function(x,d0,d1){
  ProbTreat <- (exp(beta0+beta1+(beta2+beta3)*x)/(1+exp(beta0+beta1+(beta2+beta3)*x)))
  # Prob(D(1)=1 | Y)
  ProbNotTreat <- (exp(beta0+beta2*x)/(1+exp(beta0+beta2*x)))
  # Prob(D(0)=1 | Y)

  if(d0==0 && d1==0) {
    return((1-ProbNotTreat)*
            (1-ProbTreat)*dnorm(x)/(P1))
  }
  if(d0==0 && d1==1) {
    return((1-ProbNotTreat)*
            (ProbTreat)*dnorm(x)/(P2))
  }
  if(d0==1 && d1==0) {
    return((ProbNotTreat)*
            (1-ProbTreat)*dnorm(x)/(P3))
  }
  if(d0==1 && d1==1){
    return((ProbNotTreat)*
            (ProbTreat)*dnorm(x)/(P4))
  }
}

```

```

# }}}

# }}}

##### Fonction permettant de calculer la fonction de répartition
# {{{ conditional cdf
### Fonction permettant de calculer la fonction de répartition conditionnelle
CDf <- function(y,d0,d1){
  integrate(function (x){densitey1(x,d0,d1)}
    ,lower=-10,upper=y,abs.tol=0.000001)$value
}
# }}}

## Vérification
CDf(2,0,1)
CDf(200,0,1)
x <- seq(-5,2,length.out=5000)
y <- sapply(x,function(y)CDf(y,0,1))
y1 <- sapply(x,function(y){densitey1(y,0,1)})
plot(x,y)
plot(x,y1)

##### Inversion de la fonction de répartition #####

U <- runif(n) # uniforme sur (0,1)
Y <- rep(NA,n)
for(i in 1:n) {
  Y[i] <- uniroot((function(x)(CDf(x,d0=D0[i],d1=D1[i])-U[i]))
    ,lower=-100,upper=100)$root # renvoie  $F^{-1}(U)$ 
}
## Y <- sapply(U,function(u){uniroot((function(x)(CDf(x,0,1)-u))
  ,lower=-100,upper=100)$root})

```

```

# }}}

# {{{ Etape 3
Treat <- rbinom(n,size=1,prob=0.5)
# génère le traitement avec P(trt=1)=0.5
# }}}

# {{{ Etape 4
D <- D1*Treat + (1-Treat)*D0
# }}}

# {{{ create data frame
DataSimul <- data.frame(D=D,D1=D1,D0=D0,Y=Y,Treat=Treat)
head(DataSimul)

# }}}

```

Annexe B

Annexes

B.1 Code R pour l'extension de données réels

```
set.seed(1234)

## {{{ technicalities
library(survival)
library(timereg)
library(prodlim)

d <- JBR.10_data
## }}}

## {{{ sample size
n <- 1000
timepoints <- c(2,6) # give in increasing order !!
## }}}

## {{{ estimate the model from the data
estima <- comp.risk(Event(OS_time,OS_status)~Treat+const(Marquer)+
                    const(Treat*Marquer),
                    data=d,
                    cause=1,
                    times=timepoints,
                    resample.iid=1,
                    n.sim=1000,
```

```

        model="logistic")

## }}}

## {{{ generate biomarker
Y <- sample(x=d$Marquer, size=n, replace=TRUE)
# This is equivalent to generate the values by inverting the empirical
#cumulative distribution function.
# Indeed, in both case, each new value is drwan among values of
  #d$Marquer, with probability 1*/nrow(d)
## }}}

## {{{ generate treatment
treat <- rbinom(n,size=1,prob=0.5)
## }}}

## {{{ generate time to event

## {{{ first generate the risk given marker and treatment at each time point
RiskGivenYTreat <- function(x){
  predict(estima,newdata = data.frame(Marquer=x[1],Treat=x[2]),times=timepoints)$P1
}
AllRisks <- t(apply(cbind(Y,treat),1,RiskGivenYTreat))
colnames(AllRisks) <- paste0("Risk at t=",timepoints)
head(AllRisks)
## }}}

## {{{ Second, create a function which linearly extrapolates between
## the time points,to get the risk for any time

ExtrapolatedRisk <- function(t,allRisk,timepoints){
  if(t<timepoints[1]){
    out <- t*allRisk[1]/timepoints[1]
  }else{
    i <- sum(t>=timepoints)
    if(i<length(timepoints)){

```

```

    a <- allRisk[i]
    b <- (allRisk[i+1]-allRisk[i])/(timepoints[i+1]-timepoints[i])
    if(b==0){b <- 0.001}
# we need to force the risk to be increasing, to avoid computational issue
    out <- a + b*(t-timepoints[i])
  }else{
    a <- allRisk[i-1]
    b <- (allRisk[i]-allRisk[i-1])/(timepoints[i]-timepoints[i-1])
    out <- a + b*(t-timepoints[i-1])
  }
}
min(out,1)
}

# example of the curve of predicted risk given marker and treatment
whichSubject <- 30
ttt <- seq(from=0,to=30,by=0.01)
yyy <- sapply(ttt,function(x){ExtrapolatedRisk(x,AllRisks[whichSubject,],timepoints)})
plot(ttt,yyy,type="l",ylab="predicted risk", xlab="time", axes=FALSE)
points(timepoints,AllRisks[whichSubject,], col="red", pch=16)
axis(1)
axis(2,las=1)
## }}}

## {{{ Generate time to event
Temps <- rep(NA,n)
U <- runif(n)
for(i in 1:n){
  Temps[i] <- uniroot(f=function(t)(ExtrapolatedRisk(t,AllRisks[i,],timepoints)-U[i]),
                    lower=0,
                    upper=1000,
                    tol=0.0001)$root
}

# note: actually we do not really need to use uniroot in that case,
#because we can find a simple formula

```

```

# for the root of the equation (as it is piecewise linear).
## }}}

## }}}

## {{{ generate censoring time
AllTimeCens <- unique(sort(d$OS_time[d$OS_status==0]))
FitCens <- prodlim(Hist(OS_time,OS_status)~1,data=GSE14814.randomise,reverse=TRUE)
survCens <- predict(FitCens,times=AllTimeCens)
Jumps <- -diff(c(1,survCens,0))
cbind(c(AllTimeCens,max(d$OS_time)+1),Jumps)
length(AllTimeCens)
length(Jumps)
sum(Jumps)
MyTcens <- sample(x=c(AllTimeCens,max(d$OS_time)+1),size=n,replace=TRUE,prob=Jumps)
# the idea is simply that inverting the estimated empirical
# cumulative distribution function (ecdf) is equivalent
# to draw censored time, among those observed in the real dataset,
# with probability equal to the jump of the Kaplan-Meier
# estimator of the censoring distribution.
# because Kaplan-Meier does not always reach 0,
# I added the time max(d$OS_time)+1 (beyond the last
# observed time in the data set), which is drawn with prob 1-sum(jumps).

## Just to Check that it works
plot(FitCens,confint=FALSE)
for(i in 1:10){
  Tcens <- sample(x=c(AllTimeCens,max(d$OS_time)+1),size=n,replace=TRUE,prob=Jumps)
  di <- data.frame(time=Tcens,status=rep(0,n))
  plot(prodlim(Hist(time,status)~1,data=di,reverse=TRUE),add=TRUE,col="red")
}
plot(FitCens,add=TRUE,confint=FALSE)

```

```
## }}}
```

```
## {{{ Final simulated data
```

```
dataSimul <- data.frame(time=pmin(Temps,MyTcens),  
                        status=as.numeric(Temps<=MyTcens),  
                        treat=treat,  
                        Y=Y)
```

```
table(dataSimul$status)
```

```
table(d$OS_status)
```

```
head(dataSimul)
```

```
## }}}
```

```
## {{{ Check that estimates from simulated data match with those of the data set
```

```
## {{{ plot
```

```
plot(prodlim(Hist(time,status)~1,data=dataSimul),col="red",confint=FALSE)
```

```
plot(prodlim(Hist(OS_time,OS_status)~1,data=d),col="green", add=TRUE,confint=FALSE)
```

```
abline(v=timepoints)
```

```
points(timepoints,1-colMeans(AllRisks),cex=3,pch=16)
```

```
# Here the two Kaplan-Meier curve should (approximately) be equal
```

```
# at the times corresponding to timepoints
```

```
## }}}
```

```
## {{{ check values
```

```
estima1 <- comp.risk(Event(time,status)~treat+const(Y)+  
                    const(treat*Y),  
                    data=dataSimul,  
                    cause=1,  
                    times=timepoints,  
                    resample.iid=1,  
                    n.sim=5,  
                    model="logistic")
```

```
estima$cum
```

```
estimal$cum
```

```
## }}}
```

```
## }}}
```

Titre : Validation de biomarqueurs prédictifs de la réponse au traitement : extension des courbes de prédictivités à un critère de jugement censuré

Mots clés : Biomarqueurs, pronostique, prédictifs, calibration, cancer de l'ovaire, immunothérapie, régression quantile

Résumé : Il est fréquent en oncologie, de vouloir déterminer si seulement un sous groupe de patients va bénéficier ou pas d'un traitement. C'est un des paradigmes de la médecine personnalisée ou stratifiée. Les biomarqueurs prédictifs sont souvent utilisés pour sélectionner ces patients et la plupart de ces bio-marqueurs sont continus. Par exemple les signatures génomiques comme Oncotype-Dx. Il a été proposé dans la littérature une méthodologie d'évaluation d'un biomarqueur avec réponse binaire. L'objectif dans cette thèse est dans un premier travail d'étendre cette méthodologie avec un critère de jugement censuré et de déterminer à différents horizons de prédiction le seuil optimal du biomarqueur au-delà duquel le traitement sera attribué ou évité. Un modèle dont l'estimation de ces paramètres repose sur des pondérations par l'inverse des probabilités de censure est pro-

posé permettant ainsi d'obtenir un estimateur consistant. Une extension des courbes de prédictivité par bras de traitement dépendantes du temps est ainsi proposée. Dans un second travail, un test de l'hypothèse de calibration avec des données censurées à droite a été proposé. Ce test est valide au-delà 60% de taux de censure contrairement à ceux qui existe déjà dans la littérature et nous permettra d'étudier l'influence d'une mauvaise calibration sur la détermination du seuil. Un troisième travail s'intéresse à la détermination du seuil d'un nouveau biomarqueur pronostique pour le cancer de l'ovaire afin de classer les patients en haut ou bas risque de rechute. Enfin, dans un quatrième travail, nous proposons d'employer la régression quantile pour données censurées pour quantifier le bénéfice à long terme de l'immunothérapie.

Title : Validation of biomarkers predictive of response to treatment : extension of predictiveness curves with right-censored data. **Keywords :**

Biomarkers, prognostic, predictive, calibration, Ovarian cancer, immunotherapy, quantile regression

Abstract : It is common in oncology to want to determine whether or not only a subgroup of patients will benefit from a treatment. This is one of the paradigms of personalized or stratified medicine. Predictive biomarkers are often used to select these patients and most of these biomarkers are continuous. For example genomic signatures such as Oncotype-Dx. A methodology for evaluating a biomarker with binary response has been proposed in the literature. The objective of this thesis is in first work to extend this methodology with right-censored data and to determine at different prediction horizons the optimal threshold of the biomarker beyond which treatment will be attributed or avoided. A model whose estimates of these parameters are based on inverse censored probability weights is proposed to provide consistent estimators. An extension of the predictiveness curves will be carried out. In a second work, a test of the cali-

bration hypothesis with right-censored data has been proposed. This test will be valid beyond the 60% censoring rate contrary to those already existing in the literature and will allow us to study the influence of a bad calibration on the determination of the threshold. A third work focuses on the determination of the threshold of a new prognostic biomarker for ovarian cancer in order to classify patients at high or low risk of relapse. Finally, a fourth work consists in illustrating the relevance of the censored quantile regression for quantifying the long term benefit of immunotherapy in a reconstructed data set from a single randomized trial. The proposed methodology can be readily employed for individual patients data meta-analysis to summarize evidence of immunotherapy as quantified by the upper quantile of the survival distribution.

Université Paris-Saclay

Espace Technologique / Immeuble Discovery

Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France



