



HAL
open science

Evaluation de la valeur pronostique de la dynamique du CA-125 chez les femmes atteintes d'un cancer des ovaires avancé nouvellement diagnostiqué sur une méta-analyse

Eleni Karamouza

► To cite this version:

Eleni Karamouza. Evaluation de la valeur pronostique de la dynamique du CA-125 chez les femmes atteintes d'un cancer des ovaires avancé nouvellement diagnostiqué sur une méta-analyse. Cancer. Université Paris-Saclay, 2024. Français. NNT : 2024UPASR032 . tel-04903792

HAL Id: tel-04903792

<https://theses.hal.science/tel-04903792v1>

Submitted on 21 Jan 2025

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.

Evaluation de la valeur pronostique de la dynamique du CA-125 chez les femmes atteintes d'un cancer des ovaires avancé nouvellement diagnostiqué sur une méta- analyse

*Evaluation of the prognostic value of CA-125 dynamics in women with newly
diagnosed advanced ovarian cancer in a meta-analysis*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 570, Santé Publique (EDSP)

Spécialité de doctorat : Sciences Médicales– Recherche Clinique

Graduate School : Santé Publique. Référent Université de Versailles Saint-Quentin-en-
Yvelines

Thèse préparée dans l'unité de recherche **CESP (Université Paris-Saclay, UVSQ,
Inserm)**,

sous la direction de **Xavier PAOLETTI** PU-PH, Directeur de recherche
& la co-direction de **Cécile PROUST-LIMA**, Directrice de recherche

Thèse soutenue à Villejuif, le 19 décembre 2024, par

Eleni KARAMOUZA

Composition du Jury

Membres du jury avec voix délibérative

Etienne AUDUREAU PU-PH, Université Paris Est Créteil	Président
Carine BELLERA CR, HDR, Institut Bergonié, Université Bordeaux	Rapporteur & Examinatrice
Emmanuel CHAMOREY PharmD, Professeur des universités, Centre Antoine Lacassagne, Université Côte d'Azur	Rapporteur & Examineur
Bernard ASSELAIN Professeur Emérite, Université Paris Cité	Examineur

Titre : Evaluation de la valeur pronostique de la dynamique du CA-125 chez les femmes atteintes d'un cancer des ovaires avancé nouvellement diagnostiqué sur une méta-analyse

Mots clés : cancer de l'ovaire, biomarqueur, CA-125, méta-analyse, modèles linéaires mixtes

Résumé : Le cancer de l'ovaire est le 18^{ième} cancer en termes d'incidence et le 14^{ième} en termes de mortalité. En raison d'un manque de symptômes spécifiques et d'un diagnostic tardif à un stade avancé, le pronostic de survie de ces patientes est mauvais. Après une prise en charge initiale, la majorité des patientes vont progresser. Les options thérapeutiques vont alors dépendre de la rapidité de la progression ; la mise en place d'un suivi simple et non intrusif est indispensable.

Malgré l'utilisation du marqueur sanguin, CA-125, depuis les années 80, les meilleurs moyens d'analyser son déclin et son rôle pronostique sont mal quantifiés.

En utilisant les données individuelles de la méta-analyse d'essais thérapeutiques randomisés réalisée au sein de l'intergroupe GCIg, cette thèse a pour objectif l'évaluation de différents résumés de la dynamique du CA-125 sur les premiers mois après le début du traitement pour prédire la progression et la survie des patientes.

Ces résumés comprennent des mesures observées et estimées obtenues par un modèle linéaire mixte (LMM) avec des splines. Leurs performances ont été évaluées en cross validation et des scores adaptés comme l'aire sous la courbe ROC (AUC) et le Brier Score.

Dans une première étude, nous avons montré qu'une surveillance de 3 mois avec le CA-125 permettait d'obtenir des prédictions individuelles de la survie globale jusqu'à 48 mois pour les patients recevant une chimiothérapie de première ligne. La deuxième étude nous a permis de conclure qu'une simple mesure du CA-125 à 3 mois donnait une assez bonne capacité prédictive pour détecter les patientes résistantes au platine, mais avec un risque de mauvaise classification qui conduirait néanmoins à une mauvaise orientation thérapeutique qui incite à la prudence.

Cette thèse contribue à établir un cadre pour l'utilisation du CA-125 afin de guider les cliniciens dans leur prise de décision pour le suivi des patientes atteintes d'un cancer de l'ovaire. Elle souligne également l'importance de mettre à jour la méta-analyse afin d'inclure les derniers développements en matière de traitement et des pratiques cliniques.

Title : Evaluation of the prognostic value of CA-125 dynamics in women with newly diagnosed advanced ovarian cancer in a meta-analysis

Keywords : ovarian cancer, biomarker, CA-125, méta-analysis, linear mixed model

Abstract : Ovarian cancer ranks 18th in terms of incidence and 14th in terms of mortality. Due to a lack of specific symptoms and late diagnosis at an advanced stage, the survival prognosis for these patients is poor. After initial management, the majority of patients will progress. Treatment options will then depend on the speed of progression, and simple, non-intrusive follow-up is essential.

Despite the use of the CA-125 blood marker, since the 1980s, the best means of analyzing its decline and prognostic role are poorly quantified.

Using individual patient data from the individual patient data meta-analysis initiated by the GCIg, this thesis aims to evaluate the different summaries of CA-125 dynamics over the first few months after the start of treatment to predict patient progression and survival.

These summaries include observed and estimated measurements obtained by a linear mixed model (LMM) with splines. Their performance was assessed by cross-validation and adapted scores such as the area under the ROC curve (AUC) and the Brier Score.

In our first study, we showed that 3-month monitoring with CA-125 provides individual predictions on overall survival up to 48 months for patients receiving first-line chemotherapy. The second study concludes that a simple CA-125 measurement at 3 months gave a fairly good predictive capacity for detecting platinum-resistant patients, but with a risk of misclassification that would nonetheless lead to a wrong therapeutic orientation, which prompts caution.

This thesis contributes to establishing a framework for the use of CA-125 to guide clinicians in their decision-making for the follow-up of ovarian cancer patients. It also highlights the importance of updating the meta-analysis to include the latest developments in treatment and clinical practice.

Στον Πάρη και την Ίριδα

AVANT-PROPOS

Cette thèse été préparée dans le Bureau de Biostatistique et d'Epidémiologie de Gustave Roussy, dirigé par le Dr Stefan Michiels, et au sein de l'équipe INSERM ONCOSTAT du Centre de recherche en Epidémiologie et Santé des Populations (CESP) dirigé par Dr Stefan Michiels.

Ce travail a été réalisé sous la direction du Dr Xavier Paoletti, directeur de la recherche clinique de l'Ensemble Hospitalier à l'Institut Curie, à St Cloud, et professeur de biostatistique à l'Université de Versailles St Quentin / Université de Paris Saclay et du Dr Cécile Proust-Lima directrice de recherche en biostatistique au centre de recherche Inserm Bordeaux Population Health et Université de Bordeaux.

Ce travail a été réalisé en parallèle de mon activité de biostatisticienne au sein du Bureau de Biostatistique et d'Epidémiologie de Gustave Roussy.

REMERCIEMENTS

Je remercie le Pr. Etienne Audureau d'avoir accepté de présider mon jury de thèse. Je remercie également grandement le Dr. Carine Bellera et le Pr. Emmanuel Chamorey d'avoir accepté d'être rapporteurs de ma thèse. Je les remercie pour tous leurs précieux commentaires et suggestions. Enfin, ma gratitude s'adresse également au Dr. Bernard Asselain d'avoir accepté d'examiner ce travail. Je vous exprime à tous les membres de mon jury ma grande gratitude d'avoir pris le temps d'évaluer ce travail.

J'adresse mes très chers remerciements au Pr. Xavier Paoletti d'avoir dirigé ce travail de thèse. Je te remercie de m'avoir permis de vivre cette expérience très enrichissante qui s'appelle méta-analyse MAOV et être part de ce projet de sa naissance jusqu'au travail de méthodologie statistique de cette thèse. Je te remercie pour ton excellence statistique, ta pédagogie, ta patience qui m'ont apporté tout le long de ces huit dernières années.

J'adresse mes très chers remerciements au Pr Cécile Proust-Lima d'avoir co-dirigé ce travail de thèse. Je te remercie pour ta disponibilité, le partage de ton expertise malgré la distance et ta bienveillance qui m'ont accompagné tout le long de ce travail. Les modèles linéaires mixtes ont été un jeu d'enfant à tes cotés. C'était un grand honneur de pouvoir travaillée et apprendre à tes côtés.

Je remercie le Gynecologic Cancer InterGroup (GCIG) d'avoir financé ce projet.

Je remercie le Dr Stefan Michiels de m'avoir accueilli au sein du bureau de Biostatistique et d'Epidémiologie de Gustave Roussy et pour tout le soutien tout le long de cette thèse.

Je remercie l'ensemble du Bureau de Biostatistique et d'Epidémiologie de Gustave Roussy, pour votre présence, vos sourires et votre bonne humeur.

Je tiens à remercier particulièrement Nadège Munier, Françoise Terrier et Christine Larue. Nadège, je te remercie pour ton soutien quotidien, tes bons conseils et ta disponibilité. A Françoise, je te remercie pour ta bienveillance, ta disponibilité et ta compréhension. Christine, je te remercie pour ta disponibilité et toujours présente pour m'encourager et aider dès qu'il y a un besoin.

Je tiens à remercier mes collègues, Nathalie, Matthieu, Gwenaël pour votre disponibilité, nos discussions existentielles et scientifiques tout le long de cette thèse.

Je remercie mes anciens collègues : mon ancien bureau 50 pour votre soutien malgré la distance et nos meetings de famille - par téléphone à présent -: Caroline, Maritina, Raissa, Caroline et Benjamin ainsi que Alicia, et Elvire pour les petites pauses gourmandes et leur soutien.

Je remercie mes collègues du bureau pour leur patience et leur soutien : Sarah-

Flora, Martina, Ikram.

Je remercie Guillemette pour nos discussions scientifiques et ton aide précieuse afin que je puisse prendre en main mon sujet de recherche.

Ένα απεραντο ευχαριστω απο καρδιας στην οικογένειά μου στην Ελλάδα. Στην μητέρα μου, στον πατέρα μου, στην αδερφή μου,την ανιψιά μου, τη γιαγιά μου. Ήταν και είναι πολύ σημαντικό στήριγμα με όποιο τρόπο μπορούσαν.

Un grand merci à ma belle-famille : mes beaux-parents, Céline, David & Ely pour leur soutien morale et toutes les heures de babysitting pour que ce projet puisse voir le jour.

Un mot spécial pour ma famille, mon cheri et mes deux petits cœurs. Amir je te remercie pour ton soutien quotidien, ta présence à mes côtés et d'avoir toujours trouvé les mots pour me remotiver à poursuivre mon rêve et cette aventure de thèse. A nos deux petites créations (du début et de la fin de cette thèse) Pâris et Iris : Σας αγαπώ.

ABRÉVIATIONS

CA-125 : Cancer Antigen 125

FIGO: International Federation of Gynecology and Obstetrics

GFIG: Gynécological Cancer Inter Group

HLMM : Modèle linéaire mixte Hiérarchique (Hierarchical Linear Mixed Model)

HR: Rapport des risques instantanés (Hazard Ratio)

HRD: déficit de recombinaison homologue (homologous recombination deficiency)

HRP : compétence en matière de recombinaison homologue (homologous recombination proficiency)

IARC: Agence internationale sur les recherches contre le cancer (International Agency for Research on Cancer)

KELIM : Constante K du taux d'élimination du CA-125

MADI : Méta-analyse des données individuelles

OR : Rapport des côtes (Odds Ratio)

OS: Survie Globale (Overall Survival)

PFS : Survie sans Progression (Progression Free Survival)

PSA : Prostate Cancer Antigen

RR : Risque Relatif

LISTE DES PRODUCTIONS SCIENTIFIQUES

Directement issu du travail de thèse :

Karamouza, E.; Glasspool, R.M.; Kelly, C.; Lewsley, L.-A.; Carty, K.; Kristensen, G.B.; Ethier, J.-L.; Kagimura, T.; Yanaihara, N.; Cecere, S.C.; et al. CA-125 Early Dynamics to Predict Overall Survival in Women with Newly Diagnosed Advanced Ovarian Cancer Based on Meta-Analysis Data. *Cancers* 2023, 15, 1823. <https://doi.org/10.3390/cancers15061823>

En lien avec la thèse :

Paoletti X, Lewsley LA, Daniele G, Cook A, Yanaihara N, Tinker A, Kristensen G, Ottevanger PB, Aravantinos G, Miller A, Boere IA, Fruscio R, Reyners AKL, Pujade-Lauraine E, Harkin A, Pignata S, Kagimura T, Welch S, Paul J, **Karamouza E**, Glasspool RM; Gynecologic Cancer InterGroup (GCIg) Meta-analysis Committee. Assessment of Progression-Free Survival as a Surrogate End Point of Overall Survival in First-Line Treatment of Ovarian Cancer: A Systematic Review and Meta-analysis. *JAMA Netw Open.* 2020 Jan 3;3(1):e1918939. doi: 10.1001/jamanetworkopen.2019.18939. PMID: 31922558; PMCID: PMC6991254.

Pauline Corbaux, Benoit You, Rosalind M. Glasspool, Nozomu Yanaihara, Anna V. Tinker, Kristina Lindemann, Isabelle L. Ray-Coquard, Mansoor R. Mirza, Fabien Subtil, Olivier Colombari, Julien Péron, **Eleni Karamouza**, Iain McNeish, Caroline Kelly, Tatsuo Kagimura, Stephen Welch, Liz-Anne Lewsley, Xavier Paoletti, Adrian Cook, Survival and modelled cancer antigen-125 ELIMination rate constant K score in ovarian cancer patients in first-line before poly(ADP-ribose) polymerase inhibitor era: A Gynaecologic Cancer Intergroup meta-analysis, *European Journal of Cancer*, Volume 191, 2023, 112966, ISSN 0959-8049, <https://doi.org/10.1016/j.ejca.2023.112966>.

Table des matières

1. INTRODUCTION	14
1.1 Contexte et objectifs de la thèse	14
1.2 Plan de la thèse	17
2. CONTEXTE CLINIQUE	18
2.1 Le cancer de l’ovaire : Epidémiologie & Evolution	18
2.1.1. Les ovaires.....	18
2.2 Les tumeurs ovariennes : Les tumeurs bénignes	18
2.3 Les tumeurs malignes.....	19
2.4 Facteurs de risque	19
2.5 Les différents types et stades du cancer de l’ovaire.....	20
2.5.1 Symptômes et diagnostic	22
2.5.2 Le biomarqueur CA-125 et son utilisation (Panorama de la surveillance)	23
2.6 Le dépistage, la prise en charge et le suivi des patientes.....	24
2.6.1 Le dépistage du cancer de l’ovaire	24
2.6.2 Prise en charge des patientes.....	24
2.6.3 Suivi des patientes	27
3. METHODES POUR LA VALIDATION DE FACTEURS PRONOSTIQUES LONGITUDINAUX	30
3.1 Analyse de données longitudinales.....	30
3.2 Modèles de régression pour données longitudinales	31
3.3 Modèles à effets aléatoires.....	31
3.4 Estimation des paramètres du modèle	33
3.5 Modélisation de trajectoires non linéaires	34
3.5.1 Définition des splines	35
3.5.2 Application des splines pour la modélisation de trajectoires	37
3.6 Estimation individuelle	37
3.7 La problématique des données manquantes.....	38
3.8 Modèle mécanistique pour le CA-125	40
4. ANALYSE DE SURVIE	44
4.1 Définitions.....	44
4.2 La fonction de vraisemblance	45
4.3 Estimation de la survie par Kaplan-Meier.....	45
4.4 Modèle à risques proportionnels	46

4.4.1	Modèle de Cox stratifié.....	47
4.4.2	Extensions du modèle de Cox.....	48
4.5	Analyse de landmark.....	49
4.6	Prédictions individuelles dans le cadre de l'approche landmark.....	50
4.7	La modélisation jointe.....	51
5.	META-ANALYSES	53
5.1.	Principes généraux.....	53
5.2	Méta-analyse d'effets d'intervention.....	54
5.2.1	<i>Revue systématique</i>	55
5.2.2	<i>Vérification</i> des données et standardisation.....	55
5.2.3	Méthodes d'analyse.....	56
5.3	Méta-analyse de modèles pronostiques.....	58
5.3.1	Contexte.....	59
5.3.2	Synthèse de modèles.....	59
5.3.3.	Qualité des publications de méta-analyses et recommandations.....	61
5.3.4	Mesure de qualité des études à inclure dans une méta-analyse.....	61
5.4	Construction de facteurs pronostiques sur données individuelles de méta-analyse.....	62
5.4.1	Méthodes statistiques pour la méta-analyse d'un biomarqueur longitudinal.....	62
5.4.2.	Modèle longitudinal.....	62
5.5.	Modèle de survie.....	64
5.6.	Estimation des paramètres pour un nouveau patient d'une nouvelle étude.....	65
5.6.1	Prédiction individuelle dans le cadre de la méta-analyse.....	65
5.6.2	: Prédiction pour un patient futur et recalibration.....	66
5.7	La méta-analyse MAOV.....	69
5.7.1	Méthodes.....	69
5.7.2	Résultats.....	70
6.	MESURES DE LA CAPACITE PRONOSTIQUE	70
6.1	Evaluation de la calibration et la discrimination.....	70
6.1.1	L'aire sous la courbe ROC.....	70
6.1.2	Le score de Brier.....	75
6.2	Validation interne et validation externe.....	76
6.3	Validation dans le cadre d'une méta-analyse.....	78

7. EVALUATION DE LA TRAJECTOIRE DU BIOMARQUEUR CA-125 LES PREMIERS MOIS APRES L'INITIATION AU TRAITEMENT POUR PREDIRE LA SURVIE GLOBALE DES FEMMES ATTEINTES D'UN CANCER DE L'OVAIRE AVANCE NOUVELLEMENT DIAGNOSTIQUE, SUR LA BASE DE DONNEES DE META-ANALYSE.	80
7.1 Introduction.....	80
7.2 Méthodes.....	82
7.2.1 Population d'analyse	82
7.2.2 Critère principal.....	83
7.2.3 Méthodes statistiques.....	83
7.2.4 Modèle linéaire mixte hiérarchique.....	84
7.2.5 Modélisation de la trajectoire avec des splines	85
7.2.6 Évaluation de la performance prédictive des résumés du CA-125	85
7.3 Les résultats	87
7.3.1 Sélection et caractéristiques des études.....	87
7.3.2 Critère principal et cadre temporel de Landmark	92
7.3.3. Temps de Landmark à 3 mois	93
7.3.4 Temps de Landmark à 6 mois	96
7.3.5. Performance de la synthèse du CA-125 pour les sous-groupes de patients en fonction des caractéristiques.....	102
7.3.6 Analyse de sensibilité et valeurs prédictives de la valeur normale du CA-125	106
7.4 Discussion	107
7.5. Conclusions.....	109
8. LE ROLE DU CA-125 DANS LA DETECTION DE LA RESISTANCE A LA CHIMIOOTHERAPIE A BASE DE PLATINE DANS LE CANCER DE L'OVAIRE AVANCE : META-ANALYSE DES DONNEES INDIVIDUELLES DES PATIENTS DU GCIG.	110
8.1 Introduction.....	110
8.2 Matériel et méthodes.....	111
8.2.1 Méta-analyse sur données individuelles des patients	111
8.2.2 Mesures du CA-125	112
8.2.3 Résultats	112
8.2.4 Analyse statistique.....	112
8.2.5 Calcul du résumé du CA-125.....	112
8.2.6 Association des résumés du CA-125 avec la progression.....	113
8.2.7 Évaluation des capacités prédictives des résumés du CA-125	114
8.2.8 Quantification de la performance.....	115

8.3	Résultats	115
8.3.1	Description de la population.....	115
8.3.2	Performances prédictives des résumés du CA-125	123
8.3.3	Performance des résumés du CA-125 combinée avec le stade FIGO et la maladie résiduelle.....	126
8.4	Discussion	130
9.	EVALUATION DU SCORE KELIM SUR LA META-ANALYSE MAOV	132
9.1	Introduction.....	132
9.2.1.	Estimation de la cinétique individuelle KELIM	132
9.2.2.	Valeur pronostique du KELIM	133
9.3	Résultats.....	133
9.3.1.	Sélection des essais et des patients	133
9.3.2.	Valeur pronostique	134
9.4.	Conclusion	135
10.	DISCUSSION	136
10.1	Discussion générale.....	136
10.2	Conclusion	142
11.	ANNEXES.....	143
12.	BIBLIOGRAPHIE.....	157

1. INTRODUCTION

1.1 Contexte et objectifs de la thèse

Dans son dernier rapport fondé sur les registres de cancers de 184 pays, l'agence internationale sur les recherches contre le cancer (IARC) rapporte qu'en 2022, le cancer de l'ovaire est le 18^{ème} cancer en terme d'incidence (9^{ème} rang des cancers féminins), mais aussi le 14^{ème} en terme de mortalité avec plus de 200000 décès.[1] En France, Santé Publique France estime pour l'année 2018 une incidence de 5193 nouveaux cas en notant une diminution ces dernières années pour des raisons inconnues. Le cancer de l'ovaire est associé à une survie nette standardisée sur l'âge à 5 ans dramatique de 43% même si depuis 2005, on observe une amélioration du taux de mortalité à 5 et 10 ans. (B. Trétarre et al 2020) Le diagnostic est compliqué par le manque de symptômes spécifiques à un stade précoce ; de nombreux diagnostics sont posés à un stade avancé avec un pronostic sombre.

La prise en charge thérapeutique repose sur la chirurgie et des traitements systémiques par chimiothérapie et par thérapie ciblée. Le faible taux de guérison nécessite un suivi attentif de l'évolution de la maladie et plusieurs lignes de traitement sont fréquemment administrées. Au niveau international, on constate une certaine homogénéité dans l'algorithme de choix thérapeutiques fondé sur des recommandations portées par le gynécologic cancer inter group (GFIG). L'élément principal pour orienter le traitement de cet algorithme est le pronostic des patientes mesuré par l'émergence précoce de résistances aux traitements induisant une progression rapide de la maladie. La progression est documentée par scanner ; une augmentation des lésions tumorales (métastases), un envahissement du péritoine sont autant d'éléments caractérisant cette progression.

Au début des années 80, il a été observé que la protéine antigène tumorale 125 (CA-125) était produite par les cellules cancéreuses de l'ovaire. Celle-ci est présente dans le sang et peut être mesurée facilement par simple titration. Cette facilité l'a amenée à être utilisée pour le suivi des patientes atteintes d'un cancer de l'ovaire.

Une diminution de la concentration du CA-125 dans le sang est classiquement observée en réponse au traitement alors qu'une augmentation indique un échappement au traitement qui serait précurseur d'une progression mesurée radiologiquement. Le suivi des patientes comprend donc un examen clinique, des dosages réguliers du biomarqueur et/ ou des scanners afin de détecter une progression de la maladie ; ce sont donc des données longitudinales quantitatives d'un biomarqueur qui sont utilisées pour suivre les patientes.

Le GFIG a proposé des critères précis pour définir une progression ou une réponse au traitement en se basant sur les mesures du CA-125. Toutefois leur utilisation reste controversée et la valeur pronostique du CA-125 mal quantifiée. On observe ainsi une variation importante d'usage entre les pays. Plusieurs études ont essayé d'étudier son rôle pour la surveillance des patientes [3], [4] mais la meilleure façon d'analyser la décroissance du CA-125, les délais entre la progression du CA-125 et la progression radiologique, la reproductibilité entre les grandes études est insuffisamment documentée. En effet, l'évaluation de la valeur pronostique de l'évolution précoce du CA-125 en fonction du temps jusqu'à un événement (décès ou

progression) soulève de nombreux défis statistiques. La trajectoire du CA-125 peut être résumée avec plusieurs méthodes comme par exemple : la valeur du CA-125 au départ, la valeur du CA-125 à 3 ou 6 mois après le début du suivi ou encore la décroissance à 3 ou 6 mois. En outre, le CA-125 est un biomarqueur sujet à la variation biologique comme des erreurs de mesure, et par conséquent, les valeurs observées brutes ne sont potentiellement pas optimales pour évaluer sa valeur pronostique. Si bien qu'à la question du meilleur résumé s'ajoute la question de l'utilisation d'un modèle pour calculer ce résumé.

Nous nous intéressons ici au rôle du CA-125 chez les femmes avec un cancer de l'ovaire avancé nouvellement diagnostiqué, à partir de données de méta-analyses collectées dans le cadre d'essais thérapeutiques de traitements systémiques.

L'évaluation de la capacité prédictive d'un marqueur quantitatif longitudinal repose sur trois étapes : (i) la modélisation de la trajectoire et l'extraction d'une statistique résumée de celle-ci, (ii) la construction d'un modèle d'association entre ce résumé et le critère clinique de survie ou de survie sans progression et (iii) la quantification de la capacité prédictive de ce résumé.

Plusieurs auteurs ont proposé des méthodes de modélisation qui permettent de gérer ce type des données, caractérisées par la dispersion et le bruit [5], [6] et des méthodes pour gérer les données manquantes [7]. En cancer, ces travaux ont été largement appliqués au 'Prostate Cancer Antigen' (PSA) pour le suivi et la prise en charge de patients atteints d'un cancer de la prostate. Récemment, le taux d'élimination du CA-125 (KELIM) [8] basé sur des mesures longitudinales du CA-125, ainsi que sur les paramètres pharmacocinétiques et pharmacodynamiques du traitement, a permis d'éclairer d'un jour nouveau ce biomarqueur.

L'association entre la trajectoire et le pronostic est classiquement abordée par deux grandes approches statistiques qui permettent de contrôler les biais de sélection liés à l'étude d'une variable dépendante du temps sur un temps d'événement ; la modélisation jointe, et l'analyse de *landmark*, qui ont été développées avec succès dans autres types de cancer.[9]. La modélisation jointe permet de modéliser la dynamique de l'antigène spécifique de la prostate (PSA) au fil du temps conjointement avec le processus de survie [6]. Les deux processus sont étudiés sur la même fenêtre de temps en prenant en compte leur inter-dépendance concomitante.

Lorsqu'on s'intéresse à la prédiction d'un délai avant événement (progression ou décès) sur la base d'un biomarqueur, où le délai diffère entre le processus longitudinal (quelques mois) et le processus de survie (quelques années), une approche dite de *landmark* peut être plus appropriée. Son principe est de fixer le temps de *landmark* au moment de la prédiction afin de mesurer l'exposition avant ce temps de *landmark* et d'étudier son association avec le pronostic après le temps de *landmark*. Les résumés des mesures répétées du biomarqueur jusqu'au temps de *landmark* sont évalués pour prédire une progression ou un décès à un temps ultérieur du temps de *landmark*. Décaler la date d'origine du suivi des patients pour l'événement permet de retrouver une séquence temporelle en accord avec les conditions de validité des analyses de survie.

Lors du développement d'outils de prédiction, l'évaluation des performances prédictives doit être effectuée avec soin. La valeur prédictive du marqueur permettant de prédire le risque de

décès doit reposer sur des mesures adaptées aux données de survie comme par exemple le score de Brier et l'aire sous la courbe ROC (AUC) dépendante du temps [10]–[12]. Les performances prédictives peuvent être trop optimistes lorsqu'elles sont évaluées sur les mêmes données que celles utilisées pour entraîner le modèle. Pour corriger ce biais, des techniques de validation externe ou de validation croisée peuvent être utilisées [13].

L'application de ces approches sur des données de méta-analyse où le pronostic des patients de différentes études peut varier du fait de critères d'inclusion différents, de prise en charge spécifique au pays reste marginale. Récemment Riley et collègues ont entamé une réflexion sur la validation de critères pronostiques à partir de données de méta-analyses [14]–[16].

Ce travail de thèse vise à évaluer les performances de la dynamique précoce du CA-125 et à comparer des délais d'observation de la trajectoire pour prédire le décès ou la progression chez les patientes nouvellement diagnostiquées avec un cancer de l'ovaire avancé. Nous avons utilisé la méta-analyse des données individuelles (MADI) de patients conduits sous l'égide du groupe de méta-analyse du GCIIG. Le contexte méta-analytique permet, avec un échantillon large et diversifié, une validation croisée de la valeur pronostique et prédictive d'un biomarqueur, mais cela soulève un certain nombre de questions méthodologiques auxquelles nous nous sommes attelés.

1.2 Plan de la thèse

Le Chapitre 2 présente le contexte clinique du cancer de l'ovaire. Il décrit les caractéristiques et les particularités de cette maladie, dès les premiers symptômes jusqu'à la prise en charge et le suivi des patientes. Le principal cadre statistique dans le cas des mesures répétées d'un biomarqueur ainsi que la modélisation et le résumé de ce type des données sont décrits au chapitre 3. Le chapitre 4 présente les méthodes de survie et leurs particularités dans le cadre d'une analyse de facteurs pronostiques. Le chapitre 5 se concentre sur l'analyse des données longitudinales en survie dans le cadre d'une méta-analyse. Le chapitre 6 présente les méthodes d'évaluation de la capacité pronostique d'un biomarqueur en survie et dans le contexte méta-analytique.

Les chapitres 7 et 8 présentent l'application de ces méthodes à la méta-analyse sur les données individuelles appelée MAOV et initiée par le GCIIG (Gynecologic Cancer Intergroup). Nous avons étudié différentes méthodes de résumées des mesures longitudinales et évalué leur capacité pronostique dans le futur. Plus précisément, le chapitre 7 présente notre premier travail qui concerne l'évaluation de la trajectoire du biomarqueur CA-125 chez des femmes atteintes d'un cancer de l'ovaire avancé nouvellement diagnostiqué. Dans ce travail nous avons évalué la trajectoire du biomarqueur CA-125 pendant les premiers mois de suivi après l'initiation du traitement pour prédire la survie globale. Le chapitre 8 présente notre deuxième étude qui porte sur l'évaluation du rôle du CA-125 pour détecter la résistance des patientes à la chimiothérapie en base platine.

Au chapitre 9 nous présentons une approche mécanistique – appelé score KELIM- appliquée sur la méta-analyse MAOV. Ce travail a été effectué en collaboration avec l'équipe de recherche du Centre pour l'Innovation en Cancérologie de Lyon (CICLY).

Au chapitre 10, nous discutons les résultats et les perspectives de ce travail de recherche.

2. CONTEXTE CLINIQUE

2.1 Le cancer de l'ovaire : Epidémiologie & Evolution

Le cancer de l'ovaire était, en 2022, le 18^{ème} cancer en termes d'incidence dans le monde. La réunion de registres et données de 184 pays par l'IARC, GLOBOCAN, a estimé le poids de cette maladie à 324 000 nouveaux cas et 206000 décès à cette date. [1]. L'incidence serait plus élevée dans les pays développés avec un taux de 7.5 sur 100.000 cas et à contrario avec un taux inférieur à 5 pour 100.000 en Afrique subsaharienne. Le cancer de l'ovaire se situait au huitième rang parmi les causes de décès par cancer féminin avec un risque de décès qui est deux fois plus élevé dans les pays développés que dans les pays les moins développés.

En France, on compte environ 5 200 nouveaux cas par an et c'est la cinquième cause de mortalité par cancer chez la femme [17].

Entre 2010 et 2015, le taux de survie à cinq ans, s'élève à 45% avec un âge médian au diagnostic à 70 ans.

En France, santé publique France rapporte un taux d'incidence et un taux de décès qui sont en légère amélioration ces dernières années. Plus précisément, entre 2010 et 2023, l'incidence du cancer de l'ovaire diminue à un taux de 1.2% par an. Le taux de décès à 5 ans a quant à lui diminué de 1.7% par an pendant cette période. [17]

2.1.1. Les ovaires

Les ovaires sont les organes qui permettent la reproduction chez la femme. Elles sont situées à la continuité des trompes de Fallope, des deux côtés de l'utérus. Elles contiennent les follicules qui sont les cellules des ovules. Les ovaires ont deux principales fonctionnalités : Premièrement, ils sont responsables de la production et maturation des ovules avec l'aide de l'hormone FSH (Hormone Folliculo-Stimulante) ; deuxièmement, ils participent au système endocrinien avec des fonctions de synthèse d'hormones principalement des œstrogènes et de la progestérone ainsi que d'autres hormones qui sont responsables de la puberté, la transformation de la silhouette, la maturité des ovocytes. Avec l'avancement de l'âge, les ovaires produisent de moins à moins d'hormones, jusqu'à l'arrêt de leur fonctionnement qui correspond à la ménopause.

2.2 Les tumeurs ovariennes : Les tumeurs bénignes

L'apparition des kystes sur les ovaires est un phénomène relativement fréquent. Ces kystes sont heureusement des tumeurs bénignes qui ne se propagent pas et n'entraînent généralement pas de maladie grave. En France, cela concerne plus de 45000 femmes (arcagy.org). 10% des femmes en âge de procréer vont présenter le syndrome des ovaires

polykystiques (SOPK) [18]. Le SOPK est une pathologie hormonale parmi les plus fréquentes chez les femmes. Malgré son nom, il ne s'agit pas de kystes mais plutôt d'un dysfonctionnement de la production d'hormones.

2.3 Les tumeurs malignes

Le cancer de l'ovaire survient lorsque des cellules cancéreuses commencent à se multiplier et induire la création de kystes malins. Ces cellules ont tendance de se propager sur les cellules et organes voisins et d'autres parties du corps en entraînant des métastases.

L'origine du cancer de l'ovaire n'est pour le moment pas définie. Deux théories sont le plus souvent avancées : l'origine ovarienne ou l'origine par les trompes de Fallope. Dans le cas ovarien, on suppose que suite à l'ovulation et au détachement de l'ovule, il y a un besoin de création des cellules pour la cicatrisation de l'épithélium. La concentration d'anomalies génétiques au moment de la cicatrisation peuvent conduire à des tumeurs malignes. Il est également possible d'avoir une inflammation au moment de la cicatrisation et avec une accumulation des facteurs de risque, cela peut amener à une création des cellules cancéreuses. L'ovulation après la ménopause peut potentiellement conduire à un dysfonctionnement hormonal et une création des cellules cancéreuses.

Le cancer de l'ovaire séreux de haut grade est le plus fréquent et il présente des caractéristiques communes avec le cancer de trompes de Fallope et le cancer du péritoine. En effet, la théorie d'une possible communication par la case abdominale avec la possibilité d'échange de cellules cancéreuses est une piste étudiée par les spécialistes. Ce comportement peut potentiellement expliquer la rapidité de l'apparition des métastases suite à la découverte du cancer de l'ovaire.

Un lien entre les patientes avec un cancer du sein et une mutation BRCA2 et le cancer de l'ovaire est également possible. De Weerdt propose une ablation des trompes de Fallope et des ovaires sous certaines conditions d'âge et la présence de mutations.[19]

Enfin la cinétique de la maladie reste peu connue et il y a un fort besoin de meilleure compréhension afin d'améliorer les pratiques de prévention, du suivi et du traitement.

2.4 Facteurs de risque

Il existe des facteurs qui augmentent les chances de développer un cancer des ovaires. La présence d'un facteur de risque augmente les possibilités d'avoir un cancer mais ceci ne se réalise pas obligatoirement.

Le principal facteur de risque est dû à l'historique familial et la présence de gènes héréditaires [20]. Les gènes BRCA1 et BRCA2 augmentent le risque du cancer des ovaires. La grande majorité des cancers de l'ovaire et des cancers du sein est dû à ces deux mutations. [21]–[23]. Ces gènes aident à la réparation de l'ADN en présence des lésions. Lorsque cette fonctionnalité est

impossible on parle des mutations. Le syndrome de Lynch est une maladie génétique et héréditaire dans laquelle les patients ont une ou plusieurs mutations génétiques connues qui altèrent la réparation des mésappariements de l'ADN et est responsable d'une augmentation du risque de développer certains cancers. Les femmes porteuses du syndrome de Lynch ont un risque à vie de 12 à 15% de présenter un cancer de l'ovaire. [24], [25]. Les antécédents familiaux au premier degré de cancer du sein et de cancer des ovaires constituent un facteur de risque avec un risque relatif de 1.13[95% IC 1.02-1.26] pour le cancer ovarien de type séreux et 1.47 [95% IC 1.15-1.87] d'histologie endométrioïde [26].

Comme la plupart des cancers, l'âge joue également un rôle défavorable, puisque à partir de 45 ans, il est constaté une augmentation importante du risque [26]. Celui-ci est plus important chez les femmes de plus de 70 ans [27]. La précocité des règles, l'âge à la ménopause et l'utilisation d'hormonothérapie augmentent aussi le risque de l'apparition d'un cancer ovarien. [28] Plus précisément, les femmes qui reçoivent un supplément hormonal pendant cinq ans autour de l'âge de 50 ans, ont environ un cancer de l'ovaire supplémentaire par tranche de 1 000 utilisatrices. [29] Le risque diminue avec le temps une fois le traitement hormonal arrêté.

La présence des maladies chroniques pelviennes comme l'endométriose augmentent le risque de l'apparition d'un cancer ovarien et il est conseillé de surveiller davantage ces patientes. [30]

Kim et al 2021, suggèrent que la prise de poids augmente également le risque d'apparition d'un cancer des ovaires[31]. L'amiante a été identifié comme responsable pour l'apparition de plusieurs cancers dont le cancer de l'ovaire. [32]

La ligature des trompes, le nombre de grossesses et l'allaitement ont un effet protecteur vis à vis de l'apparition d'un cancer des ovaires. En effet, l'absence d'ovulation joue un rôle favorable aux personnes à risque. [33]. La contraception orale peut également jouer un rôle protecteur. En effet, une prise de contraceptif même de façon discontinue pendant au moins 5 ans diminue le risque de l'apparition d'un cancer de l'ovaire même trente ans après l'arrêt. [29]

2.5 Les différents types et stades du cancer de l'ovaire

Le cancer de l'ovaire est une maladie caractérisée par une grande hétérogénéité. Plusieurs efforts sont effectués pour caractériser ces sous-types afin de pouvoir identifier des stratégies thérapeutiques spécifiques à chaque situation. [34], [35]

Le cancer de l'ovaire est caractérisé par deux grandes catégories dépendantes de la localisation des cellules cancéreuses au départ.

- Adénocarcinome épithélial de l'ovaire (AEO) se développe sur la couche extérieure de l'ovaire et correspond à 90% des cas.

- Tumeurs non épithéliales : Il s'agit de tumeurs germinales, d'adénocarcinomes à cellules claires ou mucineux de l'ovaire et autres qui correspondent à 10% des cas et sont des tumeurs malignes rares.

La détermination du stade de la maladie est une étape indispensable qui permet de déterminer l'avancement de la maladie et sa nature. Son objectif est de fournir aux cliniciens du monde entier un outil avec des règles précises et détaillées. La standardisation des pratiques permet aussi de pouvoir comparer les données et les résultats.

La stadification a pour objectif de définir l'état de l'avancement de la maladie :

- (i) Fournir un plan thérapeutique en identifiant le traitement le plus adéquat par rapport à la situation de la patiente.
- (ii) Prédire l'évolution de la maladie avec la possibilité de l'intensification du suivi des patientes les plus sensibles afin de pouvoir modifier le parcours de soins.
- (iii) Effectuer un travail prospectif avec l'enregistrement des réponses au traitement des caractéristiques spécifiques afin de permettre la comparaison des différents traitements et de faciliter la décision du traitement des nouveaux patients avec les mêmes caractéristiques.

Afin de pouvoir déterminer l'étendue de la maladie, il est indispensable d'effectuer certains examens : un examen clinique, biopsie, imagerie et des analyses sanguines. Il existe deux systèmes de classification équivalents : la classification T.N.M défini par AJCC (*American Joint Committee on Cancer*), la plus ancienne et la classification FIGO (*International Federation of Gynecology and Obstetrics*).[36] Les deux classifications se sont basées sur les trois facteurs suivants : la taille de la tumeur représentée par 'T', la propagation aux ganglions lymphatiques voisins 'N'(nodes), la présence de métastases vers des sites distants 'M'. L'ampleur de la propagation de la maladie est déterminée par quatre stades et/ou les combinaisons TNM.

Le tableau 1 situé en annexes présente en détails la classification FIGO et la correspondance TNM mise à jour en 2018 . [37], [38] Ces informations sont également disponibles sur le site internet du groupe. (<https://www.figo.org/fr/news/stadification-pour-les-tumeurs-malignes-ovftp>) .

Il existe également une classification reposant sur l'histopathologie et subdivisée en trois grades, montrant à quel niveau les cellules cancéreuses sont anormales au microscope.

- Grade I : Cellules bien différenciées : <5 % de composante solide
- Grade II : Cellules moyennement différenciées : 6 à 50 % de composante solide
- Grade III : Cellules peu différenciées : >50 % de composante solide

Le type histologique permet de caractériser le cancer de ovaires. On distingue 5 types Séreux : bas grade et haut grade, Mucineux, Endométrioïde, Cellules claires, Indifférenciés. Plusieurs autres facteurs participent potentiellement à la caractérisation du cancer de l'ovaire comme par exemple le facteur de croissance HER2, l'index KI-67 et le gène p 53. Pour le moment leur utilisation est peu répandue et leur définition reste compliquée à appliquer de façon standardisée. [39]–[41]

Kossai 2018 ont détaillé les principales caractéristiques pour les cinq types histologiques les plus fréquents (annexes table 2)[34], et Lheureux et collègues 2019 ont détaillé les caractéristiques des différents types du cancer de l'ovaire épithélial et le parcours de la thérapie des patientes [35].

Le résiduel de la maladie suite à une intervention chirurgicale d'exérèse représente un important facteur pronostique qui caractérise la qualité de l'exérèse. Une chirurgie avec des résidus microscopiques non visibles est le critère de qualité recommandé.[42] Les lésions résiduelles post-opératoires sont distinguées principalement par trois types de maladie résiduelle : pas de résidu de la maladie, résidu $< 1\text{cm}$, résidu $\geq 1\text{cm}$.

2.5.1 Symptômes et diagnostic

Le cancer de l'ovaire est considéré comme 'tuer en silence' puisque les symptômes en stade précoce sont des maux plutôt ordinaires et de faible intensité. Par conséquent le diagnostic est fait à un stade avancé avec de ce fait un mauvais pronostic pour la survie de ces patientes.

Plusieurs études avec des nombres de patientes variés ont essayé de donner des indications pour identifier les symptômes de cette maladie [43]–[45]. Goff 2007 ont présenté un questionnaire basé sur 20 symptômes associés à un cancer de l'ovaire.[46] Ce questionnaire, basé sur la sévérité et la fréquence de ces symptômes permet d'identifier les femmes à risque d'une tumeur maligne et déclenche des examens supplémentaires. En effet, 94% de patientes ont eu des symptômes l'année qui précède le diagnostic du cancer.

Le symptôme le plus fréquent est des maux abdominaux à priori de faible intensité et durée pour les stades précoces. Pour les stades avancés les maux abdominaux ont une intensité de plus en plus forte, ainsi qu'une durée et une fréquence plus importantes.

Plusieurs symptômes peuvent être liés au cancer de l'ovaire comme par exemple : l'augmentation de la taille abdominale, la sensation des ballonnements, maux au bas du ventre ou abdominaux, fatigue, problèmes urinaires et envie urgente d'uriner, perte de poids, perte de l'appétit, pertes vaginales anormales et/ou saignements, troubles intestinaux, troubles respiratoires, dégradation de l'état général.

Pour la détection du cancer de l'ovaire, il n'existe pas d'examen standardisé comme par exemple la mammographie utilisée pour le cancer du sein ou le 'PAP' test reposant sur un frottis vaginal pour la détection du cancer du col de l'utérus. Les antécédents familiaux, l'historique personnel et la présence de facteurs de risque conduisent à la recherche de cancer

de l'ovaire. Une échographie ou une radiographie peuvent orienter le diagnostic et déterminer la nécessité d'examen complémentaires et la continuité du suivi.

2.5.2 Le biomarqueur CA-125 et son utilisation (Panorama de la surveillance)

Le biomarqueur Cancer Antigen 125 (CA-125) a été identifié par DR Robert C. Bast Jr en 1983 [47]. C'est la première protéine sanguine identifiée pour surveiller le parcours des patientes atteintes d'un cancer de l'ovaire. Dans son laboratoire ont été réalisées des études sur les problématiques du cancer de l'ovaire comme le diagnostic tardif et la résistance des patientes au traitement proposé. Leurs études figurent parmi les premières qui ont prouvé une corrélation entre le CA-125 et une régression, stabilité ou progression d'un cancer épithélial de l'ovaire.[48]

Le CA-125 est présent à la surface des tumeurs épithéliales de l'ovaire et il est produit en grande quantité par la tumeur de l'ovaire. Il est utile pour le suivi des patientes puisqu'un taux supérieur à la normale est le signe de la présence d'un cancer ovarien. Pour la plupart des études, le seuil inférieur à 35 UI/ml est considéré comme normal. Traditionnellement, le CA-125 est supérieur à 35 U/mL pour 50% des patientes avec un cancer de l'ovaire de stade I. [48]

Néanmoins, le taux du CA-125 peut également être élevé en présence d'autres types de cancer comme par exemple : le cancer du pancréas, de l'utérus, de la trompe de Fallope, du poumon, de l'estomac, du sein. Un taux élevé de CA-125 est également possible dans des situations non cancéreuses comme par exemple : les kystes ovariens le fibrome, la grossesse, l'inflammation du pancréas, de la plèvre, du foie.

L'utilisation du CA-125 pour décider d'un changement de traitement reste controversée puisqu'il est observé que son augmentation peut varier suivant le type et le stade du cancer. Par exemple, une faible augmentation de la concentration du CA-125 est observée pour des stades précoces d'un cancer des ovaires et d'histologie mucineuse alors qu'une augmentation plus importante est observée pour des stades III et IV d'un cancer sévère de haut grade.[47]. Kurman and Shih ont proposé un modèle dualistique qui propose une classification des différents types de cancer ovarien en deux grands groupes mais semble un peu simplifié et un besoin de validation est nécessaire[49]. Fuston et al. avec une étude qui a inclus plus de 50.000 patientes, proposent l'utilisation du CA-125, particulièrement pour les patients de plus de 50 ans, non seulement pour le suivi d'un cancer de l'ovaire mais également d'autres types de cancer.[50]

Le CA-125 est le biomarqueur le plus prometteur et joue un rôle très important au diagnostic du cancer de l'ovaire. Sa présence dans le sang lui donne plusieurs avantages pour son utilisation pour le suivi des patientes du fait de la facilité et la rapidité de prélèvement à moindre coût.

Il est important de noter que chaque laboratoire d'analyses peut disposer d'une technologie différente pour le calcul de sa concentration dans le sang, il est donc important que le suivi

soit assuré par la même structure afin de pouvoir mieux détecter les fluctuations du biomarqueur durant le suivi ; c'est ce qui est fait en recherche clinique.

2.6 Le dépistage, la prise en charge et le suivi des patientes

2.6.1 Le dépistage du cancer de l'ovaire

Des grandes études de dépistage ont été menées afin de pouvoir établir des règles standards mais malheureusement sans grand succès. Aux Etats-Unis, l'étude PLCO (Prostate, Lung, Colorectal and Ovarian Cancer) a inclus 78.216 femmes dans deux bras randomisés entre 1993 et 2001[51]. La moitié des patientes ont reçu un suivi par le CA-125 pendant 6 ans et une échographie transvaginale pendant 6 ans et l'autre moitié une surveillance standard (ces patientes n'ont reçu aucune intervention, uniquement leurs soins médicaux habituels) avec un maximum de suivi de 13 ans. Malgré le suivi plus important, aucune différence a été détectée concernant le décès dû à un cancer de l'ovaire.

Au Japon, entre 1985 et 1999, plus de 82.000 femmes en post ménopause ont été randomisées dans deux bras. Cette étude est l'une des premières études prospectives randomisées sur le dépistage du cancer de l'ovaire. Le bras interventionnel avec échographie pelvienne et suivi par le CA-125 et le bras avec surveillance standard (ces patientes n'ont reçu aucune intervention, uniquement leurs soins médicaux habituels). Dans le bras control, uniquement 32 femmes ont développé un cancer de l'ovaire. La détection d'un cancer de l'ovaire de stade I était plus fréquent au bras avec surveillance 63% vs 38% pour le bras control sans différence significative. [52]

Au Royaume-Uni, entre 2001 et 2005, l'étude UKCTOCS a randomisée 202.562 femmes en post ménopause dans trois bras avec un ratio 1 :1 :2. Le premier bras comprend un dosage du CA-125 et une échographie endovaginale en cas d'augmentation du biomarqueur, le deuxième bras comprend juste une échographie endovaginale et le troisième bras aucune surveillance. L'incidence de la maladie de stade I ou II était 39.2% (95% CI 16.1 à 66.9) plus élevée dans le premier groupe que dans le groupe sans surveillance, tandis que l'incidence de la maladie de stade III ou IV était 10.2% (-21.3 à 2.4) plus faible. Aucune différence significative dans la survie des patients avec un cancer de l'ovaire n'a été détectée parmi les deux bras de surveillance. [53]

Actuellement, il n'existe aucune recommandation par des organismes internationaux ou nationaux concernant le dépistage d'un cancer de l'ovaire. Uniquement, pour les femmes qui présentent des facteurs de haut risque et d'antécédents familiaux, on peut proposer un suivi spécifique incluant des tests génétiques.

2.6.2 Prise en charge des patientes

Une fois que le diagnostic est établi, plusieurs types de traitements peuvent être proposés aux patientes. La prise en charge du cancer de l'ovaire est dépendante principalement du stade, du

type histologique, de l'extension du cancer, de l'état de la patiente et de la présence de mutations génétiques de la réparation de l'ADN.

Deux grandes catégories de traitements sont pratiquées :

- Les traitements locaux, sans impacter le reste du corps : la chirurgie, la radiothérapie
- Les traitement systémiques : la chimiothérapie, l'hormonothérapie, les thérapies ciblées, l'immunothérapie

Il s'agit de traitements oraux ou intraveineux qui peuvent atteindre les cellules cancéreuses peu importe la localisation.

2.3.2.1 Les traitements locaux

- La chirurgie

La chirurgie est le traitement principal pour le cancer de l'ovaire. L'éligibilité du patient à une chirurgie dépend fortement de l'étendue de la maladie et du stade. L'objectif du chirurgien est de retirer si possible la totalité de la tumeur, ou avoir un résidu tumoral de moins de 1cm. Celle-ci est considérée comme une exérèse complète qui permet au patient d'avoir un meilleur pronostic.

Tanyi et al, ont récemment proposé l'injection de la molécule pafolacianine. Cette dernière a la capacité de s'attacher aux cellules cancéreuses et lors d'une intervention chirurgicale, un système d'imagerie spécifique donne une couleur fluorescente aux cellules cancéreuses. Cette méthode permet l'identification de lésions supplémentaires qui seraient autrement non identifiables. Elle donne des résultats encourageants. [54]

- La radiothérapie

C'est un traitement plutôt agressif avec des rayons X qui peuvent toucher directement les cellules cancéreuses. Cette méthode est rarement utilisée pour le traitement d'un cancer de l'ovaire. Elle peut être proposée dans des cas spécifiques quand l'étendue du cancer est très importante avec des métastases distantes.

2.3.2.2 Les traitement systémiques : la chimiothérapie

Ce type de traitement permet d'atteindre les cellules cancéreuses partout dans le corps humain même distantes ainsi qu'après la chirurgie pour atteindre les micro cellules potentiellement encore existantes. Il s'agit d'un traitement oral ou intraveineux.

La chimiothérapie la plus couramment utilisée comprend des molécules à base de platine : carboplatin ou cisplatine et des taxanes : paclitaxel, docetaxel. Chaque cycle de traitement dure 21 jours et le traitement classique comprend 3 ou 4 cycles pour le cas d'un stade précoce. [55]

A chaque fin de cycle une réévaluation de l'état de la patiente est effectuée avec une évaluation clinique, un scanner et un dosage du CA-125. Des nouvelles lignes de traitement sont proposées ou de la chirurgie si cela n'était pas possible initialement ou encore une deuxième chirurgie. Si l'avancement du cancer est maîtrisé, la patiente peut être traitée avec le même type de chimiothérapie plus tard si besoin.

Si une progression de la maladie est observée 6 mois après la fin du traitement de première ligne, d'autres traitements peuvent être proposés. Cette situation est définie comme une résistance au platine. Dans ce cas, une deuxième ligne de traitement est proposée avec les molécules suivantes : Doxorubicine : qui a montré son efficacité et une bonne tolérance dans les cas de résistance au platine, ainsi que les Topotécan, Ifosfamide, Altrétamide, Etoposide

Les effets indésirables de la chimiothérapie sont dépendants de la dose reçue et de la durée du traitement. Les effets les plus communs sont les suivants : nausées, vomissements, perte de cheveux et de l'appétit, neutropénie, anémie, fatigue.

- L'hormonothérapie

Ce type de traitement est utilisé pour des cas spécifiques de bas grade. Leur objectif est de diminuer le niveau des œstrogènes chez les femmes après la ménopause afin d'éviter la stimulation des cellules cancéreuses par celui-ci.

- Les thérapies ciblées

Les thérapies ciblées sont basées sur des protéines qui imitent le fonctionnement des anticorps qui eux sont fabriqués par les lymphocytes B. Ce sont des molécules fabriquées en laboratoire qui ont des mécanismes spécifiques qui visent à déstructurer le fonctionnement de cellules cancéreuses. Ces molécules se fixent directement sur la cellule cible afin de neutraliser son fonctionnement. Chaque type de thérapie ciblée a son propre mécanisme d'action qui vise à affaiblir le processus de développement de la tumeur en changeant la façon dont il interagit avec les cellules voisines.

Des nouvelles thérapies ciblées ont récemment montré des résultats prometteurs avec une amélioration de l'efficacité et de la précision du traitement [35]. Il s'agit des anti-angiogéniques (anti-VEGF) et des inhibiteurs de PARP (PARPi) :

- Les anti-VEGF : Le Bevacizumab (Avastin) est un inhibiteur d'angiogenèse. Son mécanisme vise à détruire les vaisseaux sanguins, créés par les cellules cancéreuses afin de s'alimenter. Il se fixe sur la protéine VEGF. Ceci permet le ralentissement du développement des cellules cancéreuses.

- Les inhibiteurs de la poly-ADP-ribose-polymérase (PARPi) : Des résultats prometteurs des nouvelles thérapies ont transformé la gestion du traitement dans le cancer de l'ovaire et ceci est dû aux inhibiteurs PARP (Foo, George, and Banerjee 2020; Grimm, Cropet, and Ray-Coquard 2020; Vergote et al. 2021; Mirza et al. 2022) Ces derniers permettent une réparation de l'ADN dans le cas des mutations BRCA1 et BRCA2.

En 2020, ASCO (American Society of Clinical Oncology) a publié un guide avec les bonnes pratiques thérapeutiques incluant les PARPi. Ce guide propose l'utilisation des traitements comme : Niraparib, Olaparib et Rucaparib en fonction du stade FIGO, des mutations BRCA et de la ligne de traitement. [61]

Le traitement de PARPi est conseillé pour les patients qui n'ont pas progressé 6 mois après un traitement à base de platine et qui n'ont pas reçu précédemment un autre traitement PARPi. W.P Tew et al 2020 déconseillent l'utilisation des traitement PARPi en combinaison avec la chimiothérapie, d'autres agents ciblés ou des agents d'immuno- oncologie dans un contexte récurrent en dehors du contexte d'un essai clinique.

- L'immunothérapie

L'objectif de ce type de traitement est d'activer le système immunitaire afin de reconnaître les cellules cancéreuses pour participer à leur destruction. Les tumeurs profitent de plusieurs processus immunologiques en utilisant les points de contrôle immunitaire tels que les lymphocytes T et les antigènes à leur avantage. [62]

Le récepteur du point de contrôle immunitaire PD-1 (PD-L1 et PD-L) peut conduire à une réponse immunitaire anti tumorale efficace avec l'activation d'un mécanisme de blocage de l'interaction avec les cellules tumorales. [63], [64]

Le Pembrolizumab est un anticorps monoclonal qui cible la protéine PD-1, qui est présente sur les lymphocytes de type T et évite à ces cellules d'attaquer les cellules immunitaires du corps.[65] Ceci permet de ralentir la croissance des tumeurs et renforce la réponse immunitaire de la patiente. Il présente une efficacité et bonne tolérance pour les stades avancés qui ne répondent pas à la première ligne de traitement.

Toutefois, le traitement standard constitué d'une chirurgie suivie par une chimiothérapie ne permet que très rarement la guérison. Des résistances au traitement initial apparaissent rapidement. L'hétérogénéité de la maladie explique en partie cette inefficacité.

2.6.3 Suivi des patientes

Le suivi des patientes est donc un élément clé, puisque 70% des patientes vont présenter une récurrence dans les deux ans et pour la plupart des cas avec une diffusion au péritoine aggravant le pronostic.

Le GCI (Gynecologic Cancer Intergroup) a proposé un cadre précis pour l'évaluation de la réponse au traitement selon la situation clinique (patiente nouvellement diagnostiquée ou à la récurrence). Plus précisément, en première ligne chez des patientes opérées, Rustin et al 2011, ont recommandé l'utilisation du CA-125 pour la définition de la progression dans le cas d'une première ligne de traitement. Son utilisation n'est en revanche pas conseillée pour la réponse au traitement [66]. Pour les études en rechute, ils ont recommandé l'utilisation du CA-125 pour

définir la réponse au traitement et la progression. Enfin, pour les études de thérapies administrées en maintenance, l'absence de données suffisantes les conduit à attendre des études supplémentaires pour validation.

La réponse au traitement, d'après le GCIg, est reflétée par une diminution de la concentration du CA-125 d'au moins de 50% par rapport à sa concentration à l'initiation du traitement. La réponse est confirmée si elle est maintenue au moins pendant 28 jours. La notion de données longitudinale est donc centrale avec les aspects de délais, de dynamique de réduction et de seuil de décision. Le GCIg a également souligné que les patients ne peuvent pas être évalués par le CA-125 s'ils ont reçu des anticorps anti-souris ou ont reçu une intervention chirurgicale au niveau de leur péritoine ou leur plèvre. [67]

La définition de la progression, d'après le GCIg, dépend du niveau initial et de la trajectoire du CA-125 :

- Lorsque le CA-125 pendant le traitement est passé sous les >35ng/ml, considéré comme limite supérieure normale, une progression est déclarée au moment où le CA-125 remonte à deux fois la limite supérieure normale
- Pour celles qui ne se normalisent jamais, elles doivent présenter des signes de CA-125 supérieur ou égal à deux fois sa valeur au nadir.
- Lorsque le CA-125 est dans la plage normale avant traitement, une progression est déclarée au moment où le CA-125 remonte à deux fois la limite supérieure normale.

Pour tous les cas, une progression doit être confirmée par des mesures à deux reprises à au moins une semaine d'intervalle.

Deux stratégies de suivi des femmes sont donc utilisées :

- i) Le GCIg propose l'utilisation des scanners tout au long du suivi des patientes avec en plus à chaque visite un dosage du CA-125. [66] (critère standard)
- ii) Le NCCN (National Comprehensive Cancer Network) a proposé en 2013 un suivi clinique tous les trois mois avec un dosage du CA-125 optionnel (dépendant au choix du patient) et d'autres marqueurs tumoraux qui deviennent obligatoires avec des signes cliniques et dans le cas d'une augmentation du biomarqueur confirmée ou d'une suspicion suite à l'examen clinique, un scanner doit être déclenché.[68]

La surveillance des patientes reste encore aujourd'hui un sujet controversé puisque les deux méthodes ont des points positifs et négatifs. La première méthode est beaucoup plus précise puisque à chaque suivi, un scanner est effectué. Mais ceci peut entraîner à un grand nombre de rayons X que le patient reçoit régulièrement ainsi qu'un coût de suivi important. [69], [70]

Dans le deuxième cas, le suivi avec un dosage régulier du CA-125 est laissé au jugement du clinicien. C'est une option moins invasive et plus simple à mettre en place.

Une étude randomisée a donc été proposée pour comparer deux prises en charge sur la survie globale : chez les patientes avec une « progression » du CA-125, démarrer un traitement de seconde ligne immédiatement versus lorsque des symptômes cliniques apparaissent [69]. L'absence de différence sur les courbes de survie a posé la question de l'intérêt du suivi par CA-125. Les patientes avec suivi régulier, ont eu une qualité de vie détériorée mais sans avoir une différence de survie entre les deux bras.[66]

En France, les recommandations de l'INCA en 2019 pour les patientes avec résection complète sont un suivi avec le CA-125, si celui-ci était initialement élevé et le déclenchement d'un scanner suite à une augmentation du CA-125. C'est une position intermédiaire pour concilier la limitation des irradiations aux rayons X, et les performances meilleures des scanners pour le diagnostic de la récurrence. Toutefois, comme nous le verrons sur les données de la méta-analyse MAOV, cette stratégie de suivi est appliquée uniquement par certains pays. D'autres préfèrent le suivi par scanners.

3. METHODES POUR LA VALIDATION DE FACTEURS PRONOSTIQUES LONGITUDINAUX

3.1 Analyse de données longitudinales

Dans le contexte de suivi régulier des patientes, plusieurs mesures du biomarqueur sont collectées dans le temps. Avec le développement de marqueurs sériques sanguins, faiblement invasifs pour les patients, ce type de données répétées ou longitudinales est de plus en plus collecté en médecine, en recherche clinique et en biologie. La difficulté de l'analyse de ce type des données porte sur le fait qu'un individu dispose de plusieurs mesures qui ne peuvent pas être considérées comme indépendantes les unes des autres et qu'il existe inévitablement des mesures manquantes. En plus, au vu de leur trajectoire souvent non linéaire, décrire leur trajectoire peut s'avérer complexe.

La figure 3.1 représente les trajectoires individuelles du log des mesures de CA-125 chez des femmes atteintes d'un cancer des ovaires avancé. Elles se caractérisent par une variabilité importante, tant entre les patientes, qu'entre les mesures d'une même patiente. La courbe bleue décrit une moyenne empirique lissée au cours du temps, qui a une allure non linéaire.

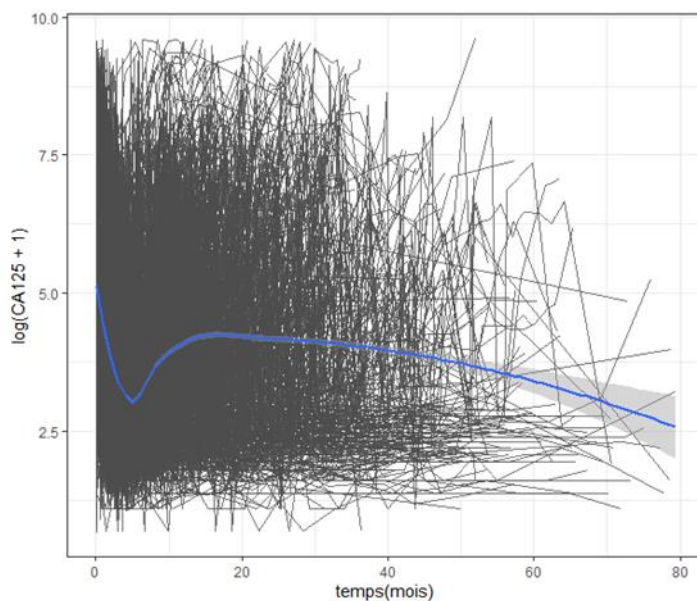


Figure 3.1: Trajectoires individuelles du biomarqueur CA-125- .

Il existe des méthodes d'analyse simples des mesures longitudinales observées : (i) la valeur un temps donné qui ne prend pas en compte la dynamique, (ii) la vitesse de décroissance observée entre deux temps différents. Cependant, le fait de simplifier ces trajectoires par une seule valeur à un temps ou la décroissance entre deux temps peuvent conduire à une simplification excessive et une interprétation inadéquate des données, puisqu'elles ne prennent pas en compte les possibles erreurs de mesures, la gestion des données manquantes à un temps

donné, le taux d'attrition causé par les patients qui sortent du suivi au cours du temps et la corrélation entre les mesures au sein d'un individu.

L'objectif des études portant sur le recueil des données longitudinales est souvent de répondre à des questions de recherche concernant le lien entre la trajectoire des données longitudinales et l'incidence d'évènements tels que le décès, la progression, la réponse au traitement et la toxicité. Plusieurs publications utilisent des méthodes erronées pour la comparaison entre deux groupes des patients conduisant à des estimations biaisées. Certaines méthodologies ne prennent pas en compte la totalité d'informations sur l'étude qui implique une perte d'efficacité statistique [71], [72]. Par exemple, il est commun que des comparaisons brutes soit effectuées entre des patients avec une durée de traitement inégale ou entre des patients qui ont eu l'évènement d'intérêt rapidement après l'initiation du traitement. En effet, les réponders doivent vivre suffisamment longtemps pour que la réponse soit observée, ce qui n'est pas le cas pour les non-réponders. Par conséquent, il est indispensable d'utiliser des méthodes prenant en compte la nature des données répétées au cours du temps qui eux sont associées à un risque d'évènement au cours du temps observé via des temps de mesure censurés.

L'objectif de notre étude est de synthétiser la trajectoire du biomarqueur obtenu par les mesures répétées au cours du temps avec des méthodes adaptées et d'évaluer leur pertinence pour prédire l'évolution de la maladie en termes de survie globale et de progression.

L'analyse de ce type des variables longitudinales continues s'est fait le plus souvent par l'utilisation des modèles linéaires mixtes (MLM). Ces derniers permettent la modélisation des trajectoires des patientes au cours du temps en prenant en compte la corrélation entre les différentes mesures.

3.2 Modèles de régression pour données longitudinales

On indexera les n patients par $i=1,\dots,n$, et les mesures répétées par $j=1,\dots,n_i$ pour le patient i , avec la somme des $n_i = N$, le nombre total de mesures.

Soit Y_{ij} la variable à expliquer (par exemple la valeur du biomarqueur CA-125) pour le patient i au temps j . Cette variable est supposée normalement distribuée. On utilisera la notation vectorielle $Y_i \in R$ pour représenter l'ensemble des mesures de i et $X_i \in R^p$ la matrice des p -variables explicatives ou covariables de dimension $n_i \times p$.

3.3 Modèles à effets aléatoires

Les modèles linéaires mixtes, une extension des modèles linéaires simples, sont utilisés pour gérer la dépendance sérielle entre les données.

La caractéristique principale de la modélisation linéaire mixte est de modéliser explicitement la trajectoire de chaque individu. Cette trajectoire peut être représentée par une fonction linéaire ou des fonctions non linéaires du temps. De plus, le modèle peut être basé sur paramétrisation linéaire ou non linéaire. Dans le cas présent, nous nous concentrons sur les modèles linéaires en les paramètres.

Cela permet d'expliquer la réponse Y_i non seulement par profil de variables explicatives et l'hétérogénéité entre les sujets mais aussi par la corrélation intra-patient ou de façon équivalente l'hétérogénéité entre les patients. Cela permet aussi de faire des prédictions au niveau du sujet, et pas uniquement pour un groupe de sujets.

Laird & Ware en 1982 ont introduit les modèles mixtes permettant de prendre en compte la nature répétées des données et par conséquent la corrélation au sein d'un individu et intra-individus[73]. Il existe plusieurs types de modèles mixtes qui sont dépendants des caractéristiques de la variable de réponse Y . Les modèles linéaires mixtes (et une partie des modèles non linéaires mixtes) font l'hypothèse que Y suit une loi normale alors que les modèles linéaires mixtes généralisés supposent que Y suit une loi de la famille exponentielle, possiblement non normale.

Le modèle pour un patient s'écrit

$$Y_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})b_i + \varepsilon_{ij} \quad (3.1)$$

où la mesure $Y_i(t_{ij})$ au temps t_{ij} dépend de deux vecteurs de covariables $X_i(t_{ij})$ et $Z_i(t_{ij})$ (de dimension p et q respectivement) associés aux effets fixes β , et aux effets aléatoires b_i supposés suivre une loi normale centrée de matrice de variance covariance D . Pour le patient i , X_i est de dimension $n_i \times p$ et le Z_i est de dimension $n_i \times q$. L'erreur résiduelle indépendante Gaussienne $\varepsilon_{ij} \sim N(0, \sigma^2)$ est supposée indépendante du vecteur des effets aléatoires b_i [74].

Les effets fixes beta représentent l'effet moyen de la variable X au niveau de la population. Les effets aléatoires sont spécifiques à chaque individu et représentent la déviation individuelle à la trajectoire moyenne de Y pour chaque élément de Z .

La variance de Y_i est définie par :

$$var(Y_i) = V_i = Z_i B Z_i' + \Sigma_i \quad (3.2)$$

Où Σ_i est la matrice de covariance de l'erreur résiduelle ; l'équation (3.2) fait bien apparaitre les deux sources de variabilité : la variabilité entre les individus avec $Z_i B Z_i'$ et la variabilité entre les mesures répétées d'un même individu avec Σ_i . La forme de la matrice de variance covariance dépend de l'hypothèse sur l'erreur résiduelle. Par exemple, avec l'hypothèse de l'indépendance entre les erreurs résiduelles, la matrice a une forme diagonale $\Sigma_i = \sigma^2 I_{n_i}$; on suppose alors que toute la corrélation parmi les mesures d'un même patient est capturée par la déviation du patient à la moyenne de l'échantillon : une fois prise en compte la « susceptibilité individuelle », la mesure d'un patient à un instant t_{ij} est indépendante de sa mesure au temps $t_{ij'}$ avec $j \neq j'$. Alternativement, la présence de corrélation résiduelle entre les erreurs pourra être modélisée en choisissant une matrice Σ_i adaptée, conduisant à une forme plus complexe de dimension

$n_i \times n_i$. On notera, que dans le cas d'une régression linéaire, on passe naturellement de l'estimation des variances B et Σ_i à la variance du modèle marginal $var(Y_i)$, ce qu'on appelle aussi marginalisation du modèle.

Cela fait aussi apparaître la question de la capacité à distinguer ces deux sources de variabilité à partir d'un jeu de données. En l'absence d'un grand nombre de répétitions de mesures pour chaque sujet, les structures complexes de variance-covariances deviennent non identifiables, ce qui peut être particulièrement problématique pour effectuer des prédictions.

Les modèles de prédiction clinique sont utilisés lorsque on s'intéresse à prédire le risque d'évènement à la présence d'une maladie. La meilleure méthode pour valider ces modèles sont des grandes bases des données qui permettent une validation externe. Les méta-analyses sur données individuelles font partie de ces bases des données permettant naturellement cette validation externe. Cependant, ces grands ensembles de données peuvent également apporter des défis méthodologiques supplémentaires. La présence d'hétérogénéité entre les études doit être prise en compte à la construction du modèle. En effet, les différentes stratégies de traitement, l'appartenance d'un patient à un protocole, le pays et la région font partie des paramètres à examiner pour vérifier la présence de cette hétérogénéité. Pour gérer l'hétérogénéité, on peut ajuster ou stratifier sur la ou les variables qui créent l'hétérogénéité. Dans ce travail, nous proposons une approche alternative qui est une analyse en sous-groupe chez les patients du bras contrôle, rendue possible par la grande taille de l'effectif et le grand nombre d'événements observés dans une méta-analyse. Il doit permettre d'étudier la question de la généralisation du modèle. Cela correspond à la recommandation de Riley et al dans leur publication de 2016 qui proposent dans un cadre de méta-analyse sur données individuelles d'examiner l'effet traitement et l'analyse en sous-groupes par rapport à ce facteur. [75]

3.4 Estimation des paramètres du modèle

L'estimation d'un modèle à effets aléatoire nécessite de construire la vraisemblance des observations. Cette vraisemblance dépend d'effets aléatoires non observés distribués selon une loi normale centrée et de leur variance qu'il faut estimer. Dans le cas d'un modèle linéaire en les paramètres et les effets aléatoires, on utilise la réécriture du modèle linéaire comme un modèle marginal en intégrant la partie liée aux variabilités individuelles (effets aléatoires) dans l'erreur (Laird and Ware 1982). On retrouve ainsi l'équation (3.3).

Notons $\vartheta = (\beta, \text{Vec}(D), \sigma)$ l'ensemble des paramètres du modèle, la fonction de log-vraisemblance d'un modèle linéaire mixte s'écrit :

$$l_N(\vartheta) = -\frac{1}{2} \sum_{i=1}^N \{n_i \log(2\pi) + \log(|V_i|) + (Y_i - X_i\beta)^\top V_i^{-1} (Y_i - X_i\beta)\} \quad (3.3)$$

Où $|V_i|$ est le déterminant de la matrice V_i

La maximisation de la fonction est obtenue par des algorithmes d'optimisation itératifs de type Newton-Raphson ou *Expectation Maximisation* (EM)

3.5 Modélisation de trajectoires non linéaires

L'objectif de notre étude est d'utiliser la dynamique d'un biomarqueur (Figure 3.5.1) sous l'effet d'un traitement pour faire la prédiction à un temps futur. La non linéarité de cette trajectoire doit être prise en compte dans le modèle longitudinal.

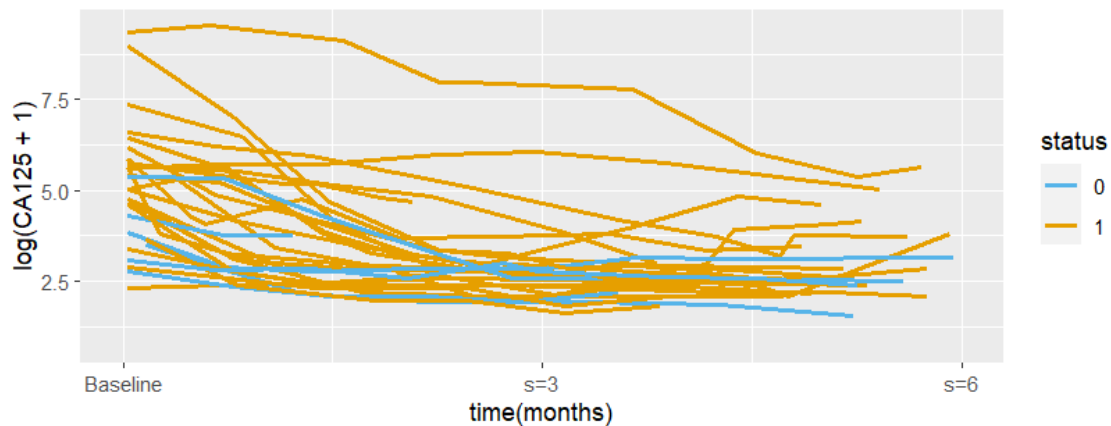


Figure 3.5.1: La trajectoire du CA-125 pour des patientes (avec ou sans évènement) d'une durée de 6 mois

Si les modèles non linéaires permettent de proposer des formes non linéaires particulières, elles se heurtent à deux difficultés principales : (i) le choix d'une forme non linéaire adaptée qui repose sur la connaissance d'un mécanisme sous-jacent d'évolution du biomarqueur au cours du temps en fonction des covariables, (ii) l'estimation des paramètres qui peut rapidement devenir délicate, nécessiter des méthodes sophistiquées assorties d'hypothèses parfois difficile à effectuer, et d'un ensemble de mesures suffisantes pour identifier le modèle adéquat. Plus précisément, avec un modèle non linéaire, la distribution du marqueur ne peut plus en général être écrite de façon marginale. Par conséquent, la vraisemblance se définit alors conditionnellement aux effets aléatoires et nécessite une intégration numérique sur la distribution des effets aléatoires

Une approche de modélisation intermédiaire permet de modéliser des trajectoires non linéaires de façon souple tout en restant dans le cadre des modèles linéaires mixtes. Cette approche repose sur l'usage des fonctions de *splines* ou des polynômes fractionnaires, ou bien n'importe quel ensemble de fonctions qui approche la trajectoire des données. Ces fonctions consistent en un changement de variable par intervalles du support de la variable temps pour approcher des formes non linéaires.

3.5.1 Définition des splines

Les *splines* sont des fonctions polynômes $f(X)$, de relation bijective, définies par morceaux pour apporter une flexibilité d'ajustement aux données continues ; appliquées à la variable temps, elles permettent un large éventail de modélisation des trajectoires tout en restant dans le cadre des modèles linéaires. Perperoglou et al ont publié une documentation détaillée afin d'expliquer ces fonctions et de faciliter leur utilisation avec leur implémentation dans R. [76]

Plus précisément, les *splines* sont des polynômes de degré d définis sur un sous-ensemble de la fenêtre du temps qui forment un espace vectoriel V . Les différents morceaux sont définis à partir de k nœuds, conduisant à la fonction suivante :

$$f(X) = \sum_{k=1}^{K+d+1} w_k W_k(X) \quad (3.4)$$

où W_k est l'ensemble de fonctions de base définissant l'espace V , et w_k sont les coefficients de *splines* associés ; k nœuds donnent donc $k+1$ polynômes qui se traduisent par $k+d+1$ fonctions associées à $k+d+1$ paramètres. L'avantage de la modélisation des *splines* est que l'estimation de la fonction f se réduit à l'estimation de la combinaison linéaire des coefficients w_k .

Les *splines* doivent respecter un ensemble de propriétés, comme le fait que les dérivées d'ordre $d-1$ soient elles-mêmes dérivables et que le nombre de nœuds soit fixe.

Les *splines* se définissent par le degré, le nombre de nœuds et par le type de fonction. Le nombre de nœuds ainsi que le degré du polynôme peut être augmenté pour apporter plus de flexibilité. Cependant, une augmentation du nombre de nœuds peut conduire à un sur-ajustement des données et par conséquent une augmentation de la variance tandis qu'une diminution du nombre de nœuds peut conduire à une fonction moins flexible. Il est de pratique commune de se restreindre au degré $d=3$; on parlera alors de *splines* cubiques. Des simulations ont montré que l'erreur de mesure de la variable (ici un biomarqueur biologique sérique) est souvent supérieure au gain d'ajustement obtenu avec des degrés plus élevés [76]. Les mesures d'ajustement aux données, en termes de vraisemblance, montrent une amélioration marginale qui ne compense pas la pénalisation des paramètres supplémentaires.

Le type de fonction est souvent exprimé à travers la base vectorielle W_k . Parmi les types les plus populaires, on a :

- Polynômes fractionnaires tronqués : les effets non linéaires sont exprimés dans une base de polynômes de degré d définie, entre 2 nœuds consécutifs, par :

$$W_1(x) = 1, W_2(x) = x, W_3(x) = x^2, \dots, W_{d+1}(x) = x^d, \\ W_{d+2}(x) = (x - \tau_1)^d, \dots, W_{K+d+1}(x) = (x - \tau_K)^d.$$

où τ_k dénotent les nœuds internes.

Cette base conduit à des corrélations entre les coefficients, et on lui préfère une base de type B.

- Les B-splines : Si on note ξ_k la position des différents nœuds :

$$\xi_1 \leq \dots \leq \xi_d \leq \xi_{d+1} := a < \xi_{d+2} := \tau_1 < \dots < \xi_{d+K+1} := \tau_K < \xi_{d+K+2} := b \leq \dots \leq \xi_{2d+K+2}$$

où a et b sont les nœuds externes, τ_k : nœuds internes ξ_i : nœuds additionnels arbitraires. Une stratégie courante consiste à mettre ces derniers à égalité avec les nœuds limitrophes. Si les nœuds internes et les nœuds externes sont choisis pour être équidistants, la formule des B-splines a la forme suivante :

$$W_k^d(x) = \frac{x - \xi_k}{\xi_{k+d} - \xi_k} W_k^{d-1}(x) - \frac{\xi_{k+d+1} - x}{\xi_{k+d+1} - \xi_{k+1}} W_{k+1}^{d-1}(x), k = 1, \dots, K, +d + 1$$

Les fonctions de base de B-splines ont l'avantage d'être non nuls dans l'intervalle entre $d+2$ nœuds et nuls ailleurs.

- Les *splines* naturelles : Ces *splines* sont caractérisées par la linéarité dans les deux intervalles extrêmes $(-\infty, a]$, leur dérivée étant nulle. Une *spline* naturelle de k nœuds a $k+1$ degrés de liberté.
- *Splines* 'Cardinal' : ce sont des *splines* naturelles définies par leur valeur aux nœuds τ_1, \dots, τ_k et vérifiant

$$W_k(\tau_k) = 1 \text{ et } W_k(\tau_j) = 0, \tau_j \neq \tau_k.$$

Par conséquent, le coefficient de chaque nœud est égal à la fonction f au nœud τ_k . Cette caractéristique facilite leur interprétation.

- Les P-splines : Ces *splines* pénalisées doivent aider à choisir le nombre et l'emplacement des nœuds pour maximiser l'adéquation aux données sous contrainte de limiter les valeurs des dérivées secondes de f ; cela conduit à lisser la fonction de façon plus ou moins importante selon le paramètre de pénalisation. Il s'agit d'une extension des B-splines avec une pénalisation des coefficients

En règle générale, il est conseillé de choisir le nombre des nœuds et des bornes avec parcimonie. De nombreux packages sont disponibles dans le logiciel R, par exemple : mgcv, gamlss, splines, splines2, splinectomeR. La majorité de ces packages sont bien documentés et permettent à l'utilisateur de choisir les valeurs des paramètres des fonctions splines. Il est important de faire attention aux valeurs par défaut d'un ensemble de splines qui peuvent être un très bon choix pour dériver un modèle pour un objectif donné et un mauvais choix pour un autre objectif. L'utilisateur doit connaître le rôle des paramètres clés pour sélectionner une spline appropriée. Il est essentiel de poursuivre l'orientation et la comparaison des méthodes de splines, et l'équipe de Perperoglou et al a l'intention d'approfondir cette question dans ses travaux futurs. [76]

3.5.2 Application des splines pour la modélisation de trajectoires

Appliqué à la modélisation de l'évolution d'un biomarqueur au cours du temps, la fonction *spline* est appliquée à la variable temps. La forme de la trajectoire d'un biomarqueur au cours du temps, comme on peut voir à la figure 3.5.1, est caractérisée par une diminution rapide attendue sous traitement puis une possible stagnation des valeurs avant ré-augmentation plutôt lente. Plus généralement, ces données ont une forme non linéaire. Dans notre cas nous avons choisi les *splines* cubiques naturelles à cause de la linéarité aux bords des nœuds. L'expression de l'équation (3.1) devient :

$$Y_{ij}(t_{ij}) = f_1(t_{ij})\beta + f_2(t_{ij})b_i + \varepsilon_{ij} \quad (3.5)$$

Où f_1 et f_2 sont deux bases de splines. La fonction multivariée f_1 est associée aux effets fixes et donne la forme d'évolution au niveau de la population et f_2 est la fonction multivariée associée aux effets aléatoires qui capte la déviation individuelle par rapport à la trajectoire moyenne.

Les plus souvent ces fonctions sont les mêmes mais ceci n'est pas nécessaire. Ici, les deux fonctions sont des bases de splines cubiques avec X nœuds.

Les fonctions splines peuvent être mises en interaction avec des variables explicatives indépendantes du temps, par exemple pour évaluer la trajectoire moyenne selon des caractéristiques au diagnostic.

Avec des splines naturelles, le nombre de fonctions splines dans la base est $K+1$ pour K le nombre de nœuds. Au niveau des effets aléatoires, cela introduit donc un vecteur de $K+1$ effets aléatoires avec une matrice D de variance-covariance des effets aléatoires de dimension $(K+1, K+1)$.

Avec l'utilisation de splines naturelles, le nombre de nœuds et leur positionnement sont le plus souvent sélectionnés sur la base de la connaissance de la trajectoire attendue, ainsi que sur l'ajustement aux données au travers de critères d'information de type AIC ou BIC est réalisée. Même si ces choix reposent sur les attendus de la trajectoire, les *splines* restent un outil avant tout exploratoire qui évite de faire des hypothèses fortes et conduit en général à une bonne adéquation du modèle aux données observées. La solidité de la modélisation le contrôle du risque de sur-apprentissage et la construction d'un modèle parcimonieux reposent alors sur la taille de l'échantillon, le nombre de mesures par patients, le nombre des paramètres et la représentativité de l'échantillon quant à la population cible.

3.6 Estimation individuelle

Les modèles linéaires mixtes permettent d'obtenir des estimations individuelles. Pour ce faire, il est nécessaire d'estimer la valeur de l'effet aléatoire pour un sujet donné. En utilisant le théorème de Bayes, on construit un estimateur empirique bayésien des effets aléatoires conditionnellement aux observations qui se calcule comme suit à partir de la distribution bivariée des effets aléatoires et de la variable dépendant au point d'estimation (*):

$$\hat{b}_i(\hat{\theta}) = E(b_i | Y_i; \hat{\theta}) = D Z_i^T \hat{V}_i^{-1} (Y_i - X_i \hat{\beta}) \quad (3.6)$$

A partir de l'estimation de l'effet aléatoire pour le patient i , on peut déduire une prédiction de la valeur de $Y_i(t_{ij})$ en n'importe quel temps :

Donne l'équation : $E(Y | \hat{b}_i) = X_i \hat{\beta} + Z_i \hat{b}_i$

Ainsi que d'autres résumés comme la pente à l'origine, la pente à un temps donné pour un patient etc. On l'appelle prédiction, mais c'est une estimation à partir des données observées ; elle peut être extrapolée pour des valeurs du temps non mesurées.

Ces prédictions, issues de l'approche par modèle mixte, ont trois caractéristiques qui nous seront utiles dans ce travail :

- Elles prennent en compte l'ensemble des mesures réalisées pour un patient, ainsi que les mesures de l'ensemble des patients ;
- Elles peuvent être calculées pour chacun des patients, indépendamment du nombre et du temps des mesures, ce qui s'apparente à une gestion des données manquantes sous certaines hypothèses (cf. 3.9); même si une seule mesure au début du suivi est disponible, il est possible de reconstruire une prédiction qui utilisera l'effet aléatoire du patient et les mesures de population ;
- En distinguant les variations inter-patients des erreurs résiduelles (ou erreurs de mesure), on obtient des prédictions dont on a retiré cette erreur de mesure ; sous l'hypothèse d'un modèle bien spécifié, la prédiction représente une meilleure quantification que la valeur observée lorsqu'elle est bruitée.

Ce dernier point est naturellement controversé ; s'il est admis qu'il existe une fluctuation liée à la mesure elle-même qui se retrouverait si on répétait la quantification du biomarqueur au même temps chez le même patient, on ne peut prétendre connaître le vrai modèle ; la notion de modèle bien spécifié reste de l'ordre de l'hypothèse.

3.7 La problématique des données manquantes

En recherche biomédicale sur patients, la présence des données manquantes est un phénomène fréquent. Si les essais thérapeutiques offrent un cadre protocolisé qui tend à réduire le risque de données manquantes, elles restent inévitables. Dans un travail séminal, trois types des données manquantes ont été définies par Rubin [77] :

- Missing Completely At Random (MCAR) : La probabilité d'avoir une mesure manquante est indépendante des mesures non observées mais elle est conditionnelle à des raisons extérieures, aléatoires et sans rapport avec son état de santé. Par exemple, il peut être dû à un simple oubli ou un déménagement.
- Missing At Random (MAR) : La probabilité d'avoir une mesure manquante est indépendante des mesures non observées mais elle est conditionnelle aux covariables et/ou mesures observées du patient. Par exemple, un patient qui a eu une détérioration de ses mesures durant les derniers suivis, est susceptible de ne pas venir au prochain rendez-vous de suivi.
- Missing Not At Random (MNAR) : La probabilité d'avoir une mesure manquante est dépendante des mesures non observées en plus possiblement des covariables et des observations du marqueur. Par exemple, les personnes dont l'état de santé s'est détérioré considérablement depuis la dernière visite risquent d'être absentes lors de la visite de suivi suivante.

Les différents mécanismes de données manquantes sont pris en compte par des méthodes de modélisation adaptées, afin de réduire le risque de biais sur les estimations des effets moyens comme de leur variance.

En pratique, on parle de données manquantes ignorables pour les cas MCAR et MAR puisque les observations seules suffisent à prédire la survenue de donnée manquante (en fait c'est le processus de données manquantes qui est ignorable). Par opposition, le cas MNAR est qualifié de non ignorable et requiert de modéliser explicitement le processus de données manquantes, alors même qu'il ne peut être caractérisé à partir des données.

Les approches reposant sur la vraisemblance sont robustes aux données manquantes MAR. Sous cette hypothèse, les estimations sont sans biais lorsque le modèle de régression inclut les variables associées au processus de données manquantes puisque le rapport de vraisemblance ignore le processus qui provoque l'absence de données.[77]. Ainsi, les estimations des modèles linéaires mixtes sont robustes aux données manquantes MCAR et MAR.

Dans le cadre des données longitudinales, une distinction supplémentaire a été réalisée :

- Les données manquantes ont une structure monotone si, à la première occurrence de donnée manquante, plus aucune mesure n'est effectuée (toutes les mesures postérieures sont manquantes). Cela correspond par exemple à des patients perdus de vue ou qui sortent d'étude pour une raison ou une autre, ce que nous avons appelé « attrition ». Par exemple, les valeurs manquantes dans le vecteur de réponses, y_i , sont monotones, si chaque fois que y_{ij} est manquant, $y_{ij'}$ l'est aussi pour tout $j' \geq j$.
- Les données manquantes ont une structure non monotone si certaines valeurs du sujet sont observées à nouveau après l'occurrence d'une valeur manquante.

Ibrahim et Molenbergs, dans un article de revue [78] dans TEST, 18(1). p. 1-43., ont décrit les étapes de l'algorithme d'estimation par la vraisemblance (ou la vraisemblance restreinte) d'un

modèle à effets aléatoires pour obtenir une statistique suffisante pour l'estimation de β , de D et de Σ sous l'hypothèse MAR. Dans le cadre MNAR, trois grandes approches statistiques ont été proposées pour modéliser conjointement la distribution des données observées et des données manquantes : les modèles de mélange, les modèles de sélection et les modèles à effets aléatoires partagés. Tous reposent sur une factorisation de la vraisemblance et nécessitent une hypothèse forte sur la nature exacte du processus de données manquantes.

Les données de biomarqueurs sur lesquelles nous avons travaillé ont été collectées dans un protocole de recherche d'une étude randomisée ; les temps de mesure étaient prévus à l'avance et correspondaient au suivi des patientes sous traitement ; l'absence de données était la conséquence probable de prélèvements non exploités ou inexploitable (échec du dosage) ou non fait (oubli ou visite décalée). La structure des données manquantes est principalement de type non monotone pour les patientes sans événement clinique conduisant à sortir d'étude ; nous avons de plus fait l'hypothèse de données manquantes au hasard et donc ignorable.

3.8 Modèle mécanistique pour le CA-125

Les modèles mixtes sont dits descriptifs puisqu'ils décrivent la forme de la trajectoire sans hypothèse sous-jacente sur le mécanisme en jeu. Au contraire, les modèles mécanistiques permettent de décrire le mécanisme biologique de la cinétique du biomarqueur. [79] Ceci nécessite une bonne connaissance du processus biologique sous-jacent de la maladie ainsi que des facteurs biologiques associées. En santé, de nombreux modèles mécanistiques sont exprimés en fonction de vitesse de production/ destruction/ transformation d'éléments avec l'aide d'équations différentielles. Dans le contexte de biomarqueurs, la construction des modèles mécanistiques nécessite de séparer les informations sur la production du biomarqueur, son élimination et sa réponse sous l'action d'un traitement. Les modèles de pharmacocinétique de population proposent de modéliser les vitesses de production, d'élimination du biomarqueur et son altération par des traitements et présentent donc des similitudes intéressantes avec notre contexte. [80]

Récemment, l'utilisation du CA-125 ELIMINATION Rate Constant K (KELIM) [8], qui est basé sur des mesures longitudinales du CA-125 a rencontré un succès important dans la communauté de la cancérologie gynécologique. Ducoulombier de l'équipe Lyonnaise a proposé le schéma suivant (Figure 3.8.1) pour la production et l'élimination du CA-125 après l'administration d'un traitement systémique [81]. On retrouve ici un modèle à trois compartiments (les 2 premiers étant virtuels) issu des recherches en pharmacocinétique des populations ; le premier compartiment, dit central et noté A1, est celui où est injectée la chimiothérapie ; avec une cinétique K , le produit passe dans le compartiment de transit, noté A2 qui correspond au délai d'action du traitement ; cet effet, noté EFFECT dans le schéma suivant se traduit par une production et une inhibition du CA-125 dans le 3^{ème} compartiment, noté A3. La production de CA-125 se fait à un taux K_{prod} , tandis que son élimination résultant de l'action du traitement se fait au taux noté KELIM.

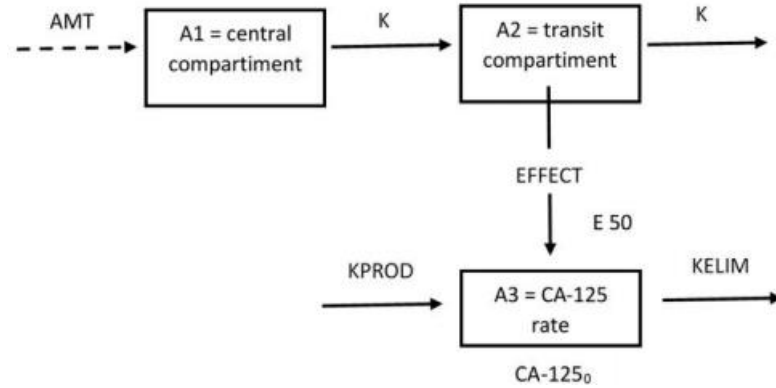


Figure 3.8.1: Description du modèle semi-mécanistique

Cette figure tirée de Ducoulombier et al. [81] où AMT représente l'administration du traitement médicamenteux, K est la cinétique d'action du traitement, K_{prod} est la vitesse de production du CA-125 et $CA-125_0$ correspond à la valeur à la mise sous traitement. E50 est la concentration du traitement conduisant à une inhibition de 50% du CA-125.

En faisant un certain nombre d'hypothèses, chacune de ces composantes peut être exprimée par des relations directes et indirectes avec le biomarqueur [82] ; l'expression de ces différents éléments résulte en un système d'équations différentielles par rapport au temps t et à des modèles non linéaires :

$$\begin{cases} \frac{dA_1}{dt} = -KA_1 \\ \frac{dA_2}{dt} = KA_1 - KA_2 \\ \frac{dA_3}{dt} = K_{prod} \left(1 - \frac{A_2}{E50 + A_2} \right) - KELIM \times A_3 \end{cases} \quad (3.7)$$

Les conditions initiales sont supposées connues à l'initiation du traitement avec aucun présence de traitement dans les deux compartiments virtuels et le niveau de CA-125 qui est celui mesuré chez le patient.

Dans le cas de vitesses linéaires, des solutions arithmétiques peuvent être calculées ; dans les cas moins simples, des solutions peuvent être approchées numériquement. Le système ci-dessus faisant intervenir le taux d'élimination du marqueur, qui est une stratégie rationnelle pour comprendre les composantes des changements temporels du CA-125 est obtenu par l'équation suivante :

$$\frac{d \text{CA-125}}{dt} = K_{prod} \left(1 - \frac{A_2}{E50 + A_2} \right) - KELIM \times \text{CA-125}$$

Les auteurs proposent d'estimer la trajectoire individuelle à partir d'un minimum de 3 mesures sur les 100 premiers jours.

En 2013, la phase III CALYPSO, chez les patients avec un cancer de l'ovaire en situation de rechute, a montré que le devenir selon la valeur KELIM mesurée à 100 jours (soit 3 mois) était significativement différent justifiant des vérifications supplémentaires [82]. Depuis, le rôle

pronostique du KELIM mesuré à 100 jours a été testé dans plusieurs essais et a suscité un intérêt important dans de multiples contextes comme celui des maladies récurrentes, des traitements néoadjuvants, différents traitements de première intention après la chirurgie). Par exemple, la réanalyse des données de patientes incluses dans l'étude CHIVA, une phase II randomisée, a montré que le KELIM est associé à la sensibilité des patientes à la chimiothérapie néoadjuvante de première ligne.[83] Dans une revue Pr You a identifié plusieurs études identifie le KELIM et les résultats de la chirurgie de cancer primitif comme des facteurs liés à la chimio sensibilité et propose de prendre en compte ces résultats pour améliorer la prise en charge des patients peu chimiosensibles lors du développement des nouvelles études [84] L'essai clinique de phase III 'VELIA', chez les patients nouvellement diagnostiqués avec un cancer de l'ovaire de haut grade, a réalisé une analyse exploratoire du KELIM en tant que facteur pronostique en combinaison avec le statut chirurgical et le statut BRCA [85]. Cette étude a montré que les patientes avec une valeur du KELIM favorable et une mutation BRCA ont un bénéfice plus important du veliparib en comparaison avec le bras control (placebo) pour la PFS. Lauby et al proposent l'utilisation du KELIM pour l'orientation des patientes en première ligne de traitement et l'identification des patients sensibles à la chimiothérapie [86].

La constante KELIM a été également étudié dans le contexte du cancer de la prostate pour modéliser le biomarqueur PSA, chez les patients avec une maladie localisée précédemment traitée par prostatectomie radicale primaire ou radiothérapie et présentant une augmentation isolée du PSA. [87]. L'étude a montré que le KELIM a une valeur pronostique significative associée à la PFS. La médiane de la PFS était de 10,1 mois chez les patients avec KELIM défavorable contre 15,1 mois chez les patients avec KELIM favorable ($p < 0,001$). La combinaison du KELIM avec le statut PARP a été analysé aux études ARIEL 2 et STUDY 10. Colombaro et al 2023 ont montré que cette combinaison est un indicateur pragmatique de l'efficacité du rucaparib chez les patientes atteintes d'un carcinome ovarien récurrent.[88] Cette étude a montré que les patients atteints d'un cancer de type BRCA- négatif et ayant un KELIM-PARP favorable ont bénéficié d'une longue PFS avec le rucaparib. Chez les patientes atteintes d'une maladie résistante au platine, KELIM-PARP a été associé à une réponse radiologique ultérieure (odds-ratio = 2,80, 95% CI 1,82-4,72).

Dans le cas du KELIM, malgré sa complexité et les limites de cette méthode au vu du nombre d'hypothèses indispensables pour les modèles mécanistiques, son association avec l'état des patientes (survie, progression) a été démontré. Sa capacité discriminante mesurée par l'AUC n'a pas été suffisamment étudiée d'où l'intérêt de le faire sur des données de méta-analyses.

Depuis février 2023, Pr You conseille l'utilisation du KELIM pour la prise en charge du cancer de l'ovaire en tant qu'indicateur de la chimio sensibilité.

Son équipe propose des outils, pour le calcul du score KELIM, accessibles sous forme d'application connectée, permettant d'ajouter à chaque cycle de traitement la valeur du CA-125 ainsi que les caractéristiques des patientes. Le résultat fourni par l'application aide à la décision du clinicien sur la suite de la thérapie.

En effet, les applications sont accessibles via les liens suivants :

Pour les patientes sous chimiothérapie à base de platine en 1^{er} ligne thérapeutique pour les 100 premières jours de traitement :

- <https://www.biomarker-kinetics.org/CA-125>

et pour les patientes en traitement neo-adjuvant (avant chirurgie) :

- <https://www.biomarker-kinetics.org/CA-125-neo>

4. ANALYSE DE SURVIE

4.1 Définitions

L'analyse de données de survie étudie le temps qui s'est écoulé jusqu'à la survenue d'un évènement d'intérêt. En cancérologie, ce dernier est principalement le décès ou la rechute mais il peut également s'appliquer à d'autres évènements comme par exemple la survenue d'une toxicité. Les modèles de régression de survie permettent d'estimer des fonctions du temps de survie (durée de survie, risque instantané,) et de comparer la durée de survie entre différents groupes de patients.

Les définitions du temps sont distinguées comme suit :

- Date d'origine ou date de début de suivi : Lors d'un essai clinique la date d'origine correspond à la date de l'enregistrement ou de randomisation dans l'étude.
- Date de l'évènement d'intérêt : correspond à la date ou le patient a eu l'évènement d'intérêt (progression, décès, toxicité, ou tout autre évènement qu'on souhaite étudier)
- Date de dernières nouvelles : correspond à la date de l'évènement d'intérêt (s'il a eu lieu) ou du dernier suivi du patient.
- Date de point : correspond à la date d'intérêt à partir de laquelle nous ne tiendrons pas en compte les éventuelles nouvelles données du patient.

Les délais sont calculés comme la différence entre la date d'origine ou du début du suivi et la date de l'évènement qu'on souhaite étudier (parmi les trois autres). Par exemple la date de premier évènement est le minimum parmi les trois dates.

Les données de survie sont caractérisées par la présence d'observations incomplètes concernant l'évènement d'intérêt. Plus précisément, il s'agit des phénomènes de la censure et la troncature qui sont déterminés par le type de l'information incomplète.

La censure à droite est la plus habituelle à ce type des données et correspond aux patients dont l'évènement d'intérêt n'a pas été observé avant la date de point ou la date de dernière nouvelle. Pour ces patients la seule information utilisée est le fait qu'il est vivant sans évènement jusqu'à sa dernière visite. Par la suite nous allons appeler ce temps C_i .

La censure à gauche, correspond aux patients qui ont déjà eu l'évènement d'intérêt avant le début de la période d'observation. La censure par intervalle, ou l'évènement d'intérêt est survenue entre deux dates mais nous ne connaissons pas la date exacte de ce dernier. Il existe aussi la troncature à gauche, dans ce cas nous ne connaissons pas les personnes qui ont eu l'évènement avant. Dans les deux cas c'est difficile à avoir l'information puisque $T < T_0$

Dans la suite, nous parlons de survie comme si l'évènement d'intérêt était le décès mais que cela s'applique aussi à d'autres.

La durée de survie (T) est une variable aléatoire positive ou nulle et continue. Sa distribution est caractérisée par les fonctions suivantes :

1. La durée de survie est définie par la probabilité de ne pas avoir l'évènement jusqu'à un temps donné t , fixé.

$$S(t) = P(T > t), t \geq 0$$

2. La fonction de répartition correspond à la probabilité de décéder avant un temps t :

$$F(t) = 1 - S(t) = P(T \leq t), t \geq 0$$

3. La densité de probabilité f qui représente la probabilité de décès dans un petit instant après t . On suppose que F accepte une première dérivée au point t :

$$f(t) = \frac{\partial F(t)}{\partial t} = \lim_{h \rightarrow 0} \frac{P(t \leq T < t + h)}{h}, t \geq 0$$

4. Le risque instantané λ , correspond au risque de décès dans un petit intervalle de temps h après t , conditionnellement au fait d'être vivant au temps t :

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T < t + h | T \geq t)}{h} = \frac{f(t)}{S(t)} = -\ln(S(t))'$$

5. La fonction de risque cumulé Λ , correspond à l'intégrale du risque instantané :

$$\Lambda(t) = \int_0^t \lambda(u) du = -\ln(S(t))$$

Par conséquent,

$$S(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(u) du\right)$$

4.2 La fonction de vraisemblance

Nous considérons N le nombre total d'individus, le patient $i \in \{1, 2, \dots, N\}$ a un temps de survie T_i^* et un temps de censure C_i . Son temps de survie observé est $T_i = \min(T_i^*, C_i)$ et l'indicatrice de son statut correspondant à l'évènement d'intérêt $\delta_i = 1_{\{T_i^* \leq C_i\}}$, où $\delta_i = 1$, si l'évènement a eu lieu au temps T_i et $\delta_i = 0$, s'il a été censuré à droite. On suppose que la censure est non informative, c'est-à-dire que la distribution de la censure C_i est indépendante de la distribution du temps avant évènement T_i^* . La vraisemblance d'un modèle de paramètres θ est écrite :

$$L = \prod_{i=1}^N f(t_i | \theta)^{\delta_i} S(t_i | L(\theta))^{1-\delta_i}$$

4.3 Estimation de la survie par Kaplan-Meier

La fonction de survie peut être estimée sous différentes lois paramétriques (loi lognormale, Weibull,..) ou non-paramétriques. L'estimateur non paramétrique de Kaplan Meier [89] est celui qui est habituellement utilisé à cause de sa capacité de gérer les données avec censure à droite.

On suppose R_i le nombre d'individus à risque et $T_{(i)}, i = 1, \dots, n$ les différents temps d'évènements et le nombre d'évènements d_i en $T_{(i)}$. La fonction en escalier qui permet d'estimer la fonction de survie est :

$$S_{KM}^{\wedge}(t) = \prod_{i; T_{(i)} \leq t} \left(1 - \frac{d_i}{R_i}\right)$$

Afin de calculer d'intervalle de confiance, la variance est estimée avec l'aide de l'estimateur de Greenwood

$$Var(S_{KM}^{\wedge}(t)) = (S_{KM}^{\wedge}(t))^2 \sum_{i; T_i \leq t} \frac{d_i}{R_i(R_i - d_i)}$$

4.4 Modèle à risques proportionnels

Les modèles à risques proportionnels sont traditionnellement utilisés pour étudier l'effet des covariables (par exemple les caractéristiques des patients : poids, âge, ...) sur le risque de survenue d'un évènement (progression, décès, toxicité, etc..). La fonction de risque à un temps t est défini comme suit :

$$\lambda_i(t | W_i) = \lambda_0(t) \exp(W_i^T \gamma)$$

Où W_i est le vecteur des covariables et γ le vecteur des coefficients correspondants et λ_0 est la fonction du risque de base. L'exponentiel d'un élément du coefficient γ correspond au rapport de risque instantané d'évènement pour une augmentation de 1 de la covariable correspondante γ . Une hypothèse majeure du modèle à risques proportionnels est donc, comme son nom l'indique, la proportionnalité du risque instantané selon le niveau de la variable explicative : le rapport de risque instantané pour deux niveaux de la variable explicative est indépendant du temps.

Pour un modèle paramétrique, $\lambda_0(t)$ peut être spécifié par une loi de Weibull une loi lognormale ou autre, et l'estimation se fait par le maximum de la vraisemblance définie en section.4.2 selon le vecteur de paramètres theta incluant ceux du risque de base et γ .

Cependant, les estimateurs paramétriques de la fonction de survie reposent sur des hypothèses de distribution des temps avant évènement. Ainsi le risque instantané est supposé monotone avec le modèle Weibull. Lorsque ce type d'hypothèses ne peut être réalisé et que les tailles d'échantillon sont suffisantes pour estimer des formes de fonction de risque plus complexes, des estimations non paramétriques sont préférées.

L'estimateur non-paramétrique de Breslow est défini comme suit $\hat{\Lambda}(t) = -\log(S_{KM}^{\wedge})$

L'estimateur d'Aalen est généralement préféré. Il est défini sur tous le temps d'événement j par :

$$\hat{\Lambda}(t) = \sum_j \frac{d_j}{R_j}$$

qui donne l'estimateur de la survie dit de Fleming-Harrington.

Il est toutefois nécessaire d'avoir un suivi suffisamment long sur l'échantillon d'estimation car, au contraire de l'estimation paramétrique, l'estimation non paramétrique ne permet pas d'extrapoler au-delà de la fenêtre de temps observée. L'estimateur d'Aalen permet de caractériser le risque à un temps t observé, avec une incertitude qui augmente avec la diminution de l'effectif à risque.

Le modèle à risque proportionnel devient alors semi-paramétrique avec $\lambda_0(t)$ qui n'est pas spécifiée. L'estimation du modèle semi-paramétrique est faite avec la maximisation de la vraisemblance partielle $L(\vartheta, W)$ qui traite $\lambda_0(t)$ comme un paramètre de nuisance.

$$L(\vartheta, W) = \prod_{i=1}^N \frac{\exp(W_i^T \gamma)}{\sum_{l: T_l \geq t_i} \exp(W_l^T \gamma)}$$

Ce modèle semi-paramétrique a été introduit en 1972 par Sir David Cox. Il est largement utilisé pour l'analyse des données de survie et permet d'analyser l'association entre les covariables et le risque d'événement. Ce modèle fait l'hypothèse de (i) la proportionnalité des risques pour chacune des covariables et (ii) la log-linéarité de l'effet des covariables continues. S'il permet d'estimer l'association entre des covariables et le temps avant événement, il ne fournit aucune estimation directe du temps avant événement. Pour effectuer des prédictions, la relation $S(t) = \exp(-\Lambda(t))$ est appliquée au modèle semi-paramétrique ce qui donne $S(t) = S_0(t) \exp(W_i^T \gamma)$. Le risque cumulé $\Lambda_0(t)$ doit alors être estimé.

4.4.1 Modèle de Cox stratifié

On peut construire un modèle de Cox stratifié sur une variable permettant d'avoir un risque de base différencié selon la strate de la variable.[90] La stratification permet de prendre en compte une forme différente du risque de base pour chaque strate ; il est donc particulièrement adapté lorsque les risques ne sont pas proportionnels au cours du temps pour une variable dont on ne cherche pas à quantifier l'effet. L'effet relatif des autres covariables est alors supposé similaire quelle que soit la strate. L'interaction entre une covariable et la variable de stratification ne peut être estimée dans le cadre de ce modèle. Une vraisemblance partielle est estimée pour chaque strate dont le produit donne la vraisemblance totale. Le risque de base spécifique de chaque strate n'est donc pas estimé, à moins d'utiliser un estimateur non paramétrique comme introduit dans le paragraphe précédent. Si on indice par k les différentes strates, alors le modèle stratifié devient :

$$\lambda_{ik}(t|W_i) = \lambda_{0,k}(t) \exp(W_i^T \gamma)$$

Par exemple, dans le cas d'une méta-analyse, un modèle de Cox stratifié permet de prendre en compte l'appartenance à un essai k , et modélise un risque de base spécifique à l'essai. Cela permet de traduire le fait que les populations ou les prises en charge peuvent être un peu différentes. Ce modèle suppose ensuite que les risques sont proportionnels pour les individus au sein d'un essai, mais pas forcément entre deux essais.

4.4.2 Extensions du modèle de Cox

Une extension importante du modèle de Cox consiste à inclure dans le modèle des covariables $W_i(t)$ qui dépendent du temps.

La fonction de vraisemblance partielle s'écrit alors simplement comme le produit pour chaque patient à chaque temps d'événement $t_{(j)}$ de la probabilité de faire l'événement conditionnellement aux valeurs des covariables mesurées au temps $t_{(j)}$ parmi les patients à risque au temps $t_{(j)}$; cet ensemble à risque est noté $R_{(j)}$. On note D le nombre d'événements, en supposant l'absence d'ex-aequo, à savoir que chaque événement se produit à un temps différent. On remarquera que (j) représente les temps d'événement

$$L(\vartheta, W) = \prod_{j=1}^D \frac{\exp(W_i^T(t_{(j)})\gamma)}{\sum_{R_{(j)}} \exp(W_i^T(t_{(j)})\gamma)}$$

Cette formule fait apparaître une difficulté majeure : la valeur des covariables doit être connue à chaque temps d'événement $t_{(j)}$ pour l'ensemble des patients encore à risque. Si cette information peut être disponible lorsque les covariables prennent des valeurs prédéfinies à des temps connus (par exemple, la date de greffe, la date de naissance d'un enfant ou de prise d'un traitement, elle est le plus souvent inaccessible lorsqu'elle requiert une mesure biologique. Ainsi, il n'y aura pas de mesure du CA-125 chez toutes les patientes, à tous les temps d'événement. On se retrouve alors avec des données manquantes.

Des stratégies d'imputation peuvent être réalisées. Ainsi, la valeur mesurée au temps précédant $t_{(j)}$ est souvent utilisée comme si la mesure biologique ne variait pas entre 2 temps de mesures. Cette hypothèse est forte et probablement complètement irréaliste lorsque les mesures sont espacées ou très variables d'une occurrence à une autre. On parle de « last observation carried forward » (LOCF). Cette technique est privilégiée à l'interpolation linéaire qui induirait d'utiliser des valeurs postérieures à $t_{(j)}$ et ainsi de conditionner sur le « futur ».

D'autre part, ces variables dépendantes du temps doivent en théorie être exogènes, à savoir prendre des valeurs indépendamment du suivi du sujet, de la distribution d'événement et de

censure. Par exemple, des mesures répétées de pollution atmosphérique constituent une variable exogène dans l'étude de l'association entre pollution et risque de développer une pathologie. Au contraire, un marqueur biologique mesuré chez la malade et utilisé pour sa surveillance, est une variable endogène. Cela induit un risque de biais d'estimation de l'association via un modèle de Cox, notamment si la dégradation clinique du patient déclenche la recherche d'une progression et la réalisation de mesures du biomarqueur. Cette approche suppose aussi que la variable explicative est mesurée sans erreur.

Enfin, l'écriture du modèle montre que c'est la valeur courante de la covariable dépendante du temps qui est directement associée au risque instantané d'événement. En d'autres termes une valeur élevée ou basse au temps $t(j)$ augmente ou diminue le risque d'événement à ce même temps. L'introduction d'un paramètre de décalage l de type $W_i(t-l)$ permet de modéliser une association décalée avec le risque instantané. C'est une hypothèse forte mais tout dépend à la question qu'on étudie et à laquelle on souhaite y répondre.). Ainsi, si on s'intéresse à l'impact d'une diminution au cours du temps du biomarqueur, appelé parfois cinétique, il devient préférable de se tourner vers d'autres types de association.

En résumé, les modèles de Cox avec covariables dépendantes du temps ne sont pas la méthode à privilégier pour étudier des cinétiques au cours du temps de biomarqueurs mesurées de façon espacées chez des patients dans le cadre d'un suivi clinique. Dans le cas de l'analyse d'un biomarqueur en recherche clinique, où l'évolution du biomarqueur est fortement dépendante de la survenue d'un événement, on s'orientera vers d'autres approches.

4.5 Analyse de landmark

A l'occasion d'une analyse erronée de l'espérance de vie des lauréats du prix Nobel comparée à la population générale,[91] a rappelé qu'on pouvait éviter le biais d'immortalité introduit par l'utilisation de covariables variant au cours du temps et mesurées au cours du suivi, par l'analyse de landmark ; celle-ci consiste juste à avancer la date d'origine du suivi des patients à un temps postérieur (appelé temps de *landmark*) et à analyser uniquement les sujets indemnes d'événement à ce temps. L'exposition est alors mesurée pendant la période avant le temps de landmark, si bien que le groupe d'exposition est bien défini avant le début du suivi, et ce sans avoir à recourir à des modélisations par variables dépendantes du temps ; en outre la variable d'exposition peut être n'importe quel résumé des observations pré-landmark comme la pente estimée sur des données répétées, le nadir etc., ce qui apporte une grande flexibilité pour l'estimation de la cinétique de variables. Les statisticiens Van Houwelingen 2007 et Putter 2017 ont approfondi les analyses de landmark et étudié leurs propriétés [92], [93].

Soit le *temps de landmark*, s . Toutes les données du biomarqueur à partir de la randomisation jusqu'au *temps de landmark* sont utilisées pour prédire le risque de progression ou décès d'un instant s à un instant $s + w$, désignant *l'horizon de prédiction*. Une illustration est fournie sur la figure 4.5.1.

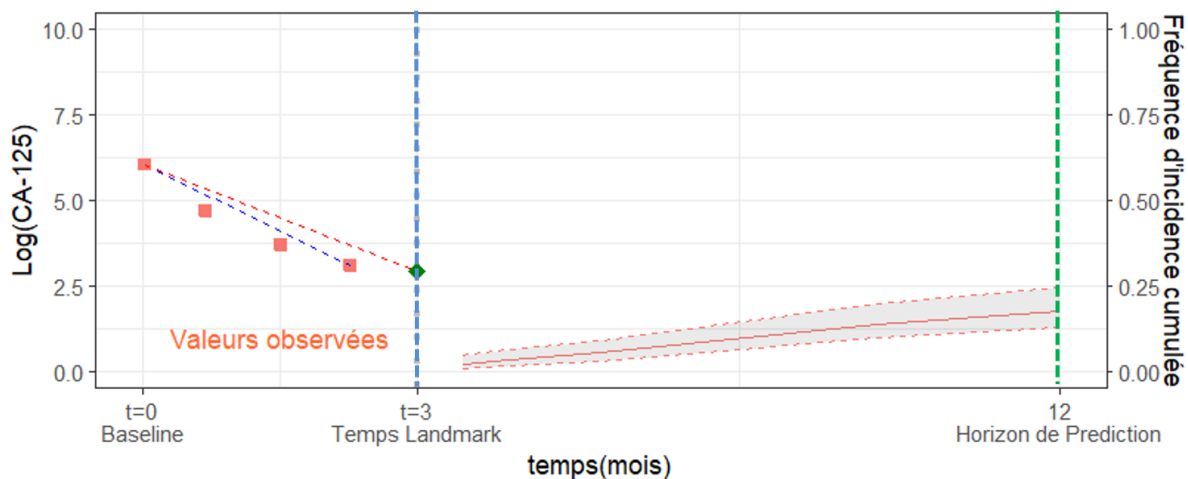


Figure 4.5.1: Illustration de l'approche 'landmark' pour un horizon de prédiction future pour 2 patientes

Avec cette approche le modèle de survie se constitue de tous les patientes encore à risque jusqu'au temps s .

Cette approche présente toutefois des limites qui peuvent restreindre son utilisation. Tout d'abord, les sujets qui font l'événement d'intérêt avant la date de landmark sont exclus des analyses. Outre la perte de puissance que cela occasionne, l'exclusion secondaire de patients peut induire des biais sur l'estimation de l'association entre l'exposition et l'événement si l'exposition a un rôle immédiat et transitoire sur l'événement. Il est donc recommandé d'utiliser un temps de landmark plutôt court après le début du traitement. Même si tout dépend également de l'objectif de l'étude et la question qu'on souhaite y répondre. De plus, les éventuelles modifications de l'exposition après le temps de landmark ne sont pas prises en compte dans l'analyse. La pertinence du choix du temps de landmark (puisque plusieurs temps landmark sont possibles) doit donc être étudiée dans le contexte clinique afin d'obtenir des résultats avisés. Par exemple, dans le cas du cancer colorectal où la réponse au traitement est attendue rapidement après le début du traitement, Buyse et al (2000) ont analysé l'effet pronostique de la réponse tumorale sur la survie par la méthode de landmark afin d'éliminer des analyses l'échantillonnage biaisé en termes de longueur. Cette méthode consiste à ignorer les réponses qui surviennent après un temps repère arbitraire et les décès survenus avant ce temps. [94] Inversement, cette méthode n'est pas appropriée pour les maladies chroniques où la réponse au traitement est plus longue.

4.6 Prédiction individuelles dans le cadre de l'approche landmark

Considérons les N patients à risque au temps s . Pour un individu i , ($i=1, \dots, N$), ayant les mesures longitudinales du biomarqueur Y_i :

$$Y_i(s) = \{y_{ij}(t_{ij}), 0 \leq t_{ij} \leq s, j = 1, \dots, n_i^s\}.$$

Cet individu a n_i mesures répétées jusqu'au temps s . X_i représente le vecteur des co-variables fixes, comme par exemple les caractéristiques des patientes, mesurées à la randomisation, T_i le temps jusqu'à l'évènement (progression ou décès) et nous avons c variables indicatrices $\mathbb{1}$ correspondant aux différents essais

Notre objectif est d'estimer la probabilité conditionnelle d'avoir un évènement ou décès avant l'horizon de temps $s + w$, sachant que l'individu est vivant à l'instant s pour un individu appartenant à une étude.

$$P_i(s + w | s) = P(s < T_i < s + w | T_i > s, Y_i(s), X_i)$$

Cette probabilité peut être estimée avec l'aide des modèles paramétriques et semi-paramétriques et elle a la forme suivante :

$$P_i(s + w | s) = P(s < T_i < s + w | T_i > s, Y(s)_i, X_i, \hat{\gamma})$$

Putter et al. en 2017, à la suite de Houwelingen en 2007, ont proposé d'analyser des temps de Landmark différents et prédéfinis avec un « super-modèle » stratifié sur le temps de landmark [92, 93].

Soit $Y_i(s)$ le vecteur des covariables au temps de landmark s et le modèle de Cox devient :

$$\lambda(t | s, Y_i(s)) = \lambda_0(t | s) \exp(\hat{\gamma}(t | s)^T Y_i(s))$$

Avec $s < t < s + w$ où $s + w$ note l'horizon de prédiction. Xu & O'Quigley ont montré que même si $Y(s)$ varie avec le temps, ce modèle fournit des prédictions asymptotiquement sans biais si on censure administrativement en $s + w$ et si la prédiction est uniquement appliquée à l'horizon $s + w$. [95]

H. Putter étend cela en proposant d'appliquer ce modèle pour chaque temps s séparément. L'estimation du paramètre $\hat{\gamma}$ dépend alors de « s » ; une paramétrisation simple de la relation entre $\hat{\gamma}$ et s est proposée. Une estimation du modèle complet en stratifiant sur s et en corrigeant la variance par l'estimateur sandwich proposé par Lin & Wei conduit à une forme de prédiction dynamique [96].

La question qu'on se pose est comment intégrer les données répétées dans l'approche landmark. Plus précisément, la problématique est de choisir quelles covariables nous allons introduire dans le modèle. Plusieurs choix sont possibles comme par exemple la dernière mesure avant le temps « s », d'autres observations avant le temps landmark (par exemple la mesure à l'initiation du traitement, l'estimation de la valeur du marqueur au temps « s »).

Dans ce travail de thèse nous avons utilisé des estimations obtenues par un modèle linéaire mixte en utilisant les données longitudinales de l'initiation du traitement jusqu'au temps de landmark « s ».

4.7 La modélisation jointe

Le principe de la modélisation jointe est basé sur l'association de deux processus : celui des mesures longitudinales par exemple d'un biomarqueur, avec un modèle de survie qui sont estimés conjointement pour prendre en compte les interrelations entre les deux. Plus précisément, il est composé de deux sous-modèles : un modèle mixte pour les mesures répétées vu dans la section 3.3 et un modèle de survie à risques proportionnels ou non vu dans la section 4.4, pour modéliser la distribution du temps jusqu'à l'événement.

$$\begin{cases} Y_i(t_{ij}) = X_i^T(t_{ij})\beta + Z_i^T(t_{ij})b_i + \varepsilon_{ij} \\ \lambda_i(t) = \lambda_0(t)\exp(X_i^E T \gamma + W_i(t|b_i)^T \eta) \end{cases}$$

Où $\lambda_0(t)$ est le risque de base qui pour la plupart du temps est paramétrique. X_i^E est le vecteur des facteurs pronostiques associé au vecteur des coefficients γ . b_i est le vecteur des effets aléatoires

Ces deux sous-modèles sont liés du fait de l'inclusion dans le modèle de survie d'une fonction (possiblement multivariée) du processus longitudinal. Plusieurs options sont privilégiées dans la littérature : $W_i(t|b_i) = Y_i^*(t)$ Correspond au niveau du marqueur sans erreur ou à la pente du marqueur au temps courant où $W_i(t|b_i) = \partial Y_i^*(t)/\partial t$. Mais d'autres associations plus complexes sont aussi possibles comme des effets cumulés [97].

De multiples alternatives ont été proposées comme par exemple, l'utilisation d'approches bayésiennes [98], [99], des modèles longitudinaux non linéaires [100], la présence de plusieurs processus longitudinaux [101], la présence de événements concurrents [102]. Récemment, un modèle marginal a été proposé, tenant en compte la corrélation entre les réponses observées de forme continue et binaire [103].

Le vaste panorama des différentes applications de la modélisation jointe est présenté avec plusieurs publications de revues [104]–[107]

La modélisation jointe permet l'analyse simultanée de la totalité d'information collectée tenant en compte l'association entre le processus longitudinal et le risque d'évènement. C'est un outil puissant qui permet de résumer avec des nombreuses possibilités le processus longitudinal. Grâce à sa grande flexibilité permet d'obtenir des prédictions individuelles dynamiques sous condition d'être correctement spécifié. En effet, à un instant t le résumé du processus longitudinal est conditionnellement mis à jour et donc l'information de par exemple 100 premières jours tend à peser moins. Le temps et la complexité des calculs sont dépendants de la complexité du modèle et de la taille de l'échantillon et doivent être pris en compte. Sachant la complexité des structures d'évolution et de dépendance, leur modélisation est souvent simplifiée. En plus, les logiciels ne peuvent pas prendre forcément en compte des structures multiniveaux avec des effets aléatoires emboîtés.

5. META-ANALYSES

5.1. Principes généraux

Le gold standard pour évaluer l'efficacité, la tolérance ou encore la sécurité d'un nouveau médicament ou de nouvelles pratiques thérapeutiques par rapport aux pratiques standards, est l'essai clinique randomisé. Il est fréquent que plusieurs essais cliniques comparent les mêmes thérapies avec des tailles de population variables et des résultats pour la plupart relativement similaires. Les méta-analyses permettent de synthétiser les résultats de ces essais cliniques afin de répondre à la question initiale avec un niveau de preuve élevé [108]

En effet, il est souvent possible que les essais cliniques n'arrivent pas à donner une réponse claire à la question posée ou que différents essais étudiant la même intervention conduisent à des conclusions contradictoires, du fait d'échantillons insuffisants et d'un manque de puissance ou de simples fluctuations d'échantillonnage. Les méta-analyses permettent de consolider la réponse à une question clinique en incluant toutes les études qui traitent du même sujet résultant en une grande puissance statistique et une plus grande représentativité des échantillons.

Une méta-analyse consiste en une estimation unique d'une quantité à partir de l'ensemble de données collectées sur la question d'intérêt. Les méta-analyses reposent donc sur une hypothèse d'homogénéité, à savoir qu'à travers les essais, les résultats sont considérés comme les réalisations d'une même variable aléatoire, justifiant l'utilisation d'une statistique unique.

La méta-analyse des différentes études peut se faire à partir de deux grands types de sources de données :

(a) les méta-analyses sur données publiées (MADP).

Elles sont basées sur les statistiques résumées extraites des publications, c'est-à-dire des résultats publiés dans l'article rapportant l'étude, ou fournies par le promoteur. Les limites de celles-ci sont que (i) les résultats peuvent être obtenus par des méthodes statistiques hétérogènes entre études, ou sur des critères ou des populations différents, (ii) les éléments nécessaires pour recalculer la statistique de test peuvent être manquants (notamment dans les cas de comparateurs multiples), (iii) la vérification des données devient impossible (iv) le biais de publication peut être élevé du fait que les études négatives ne sont pas nécessairement publiées ou publiées plus tardivement que les études positives.

(b) les méta-analyses sur données individuelles (MADI)

Elles reposent sur la collecte des données individuelles de chaque participant aux essais fournis par chaque promoteur et mises en commun. Cette procédure d'assemblage des données peut être chronophage. Les points forts des méta-analyses sur données individuelles portent sur l'augmentation du suivi par rapport à ce qui a été publié, la diminution du biais de publication si les essais publiés et non publiés sont récoltés, la possibilité de vérifier la qualité de l'étude et de réaliser une vaste gamme d'analyse statistiques sur des critères d'évaluation et des populations bien standardisées. C'est

surtout sur l'exploration de sous-groupes que les MADP offrent des atouts considérables puisqu'on peut considérer des groupes définis par les caractéristiques des patients et pas seulement par les caractéristiques des études. En particulier, les MADP permettent de tester de l'interaction entre les différentes caractéristiques des patients et le traitement pour identifier possibles effets prédictifs de la réponse au traitement et des groupes de patients ayant un bénéfice plus important. Leur limite majeure est la difficulté d'accéder aux données individuelles et d'obtenir l'exhaustivité. Les structures pérennes comme la méta-analyse d'Oxford avec une revue systématique des essais tous les cinq ans et une quasi exhaustivité des données au moment de la méta-analyse, sont un des éléments clefs pour récupérer les données.

Les limites des méta-analyses sur données publiées ont été largement détaillées. En comparant ces deux types de méta-analyses sur les mêmes essais Stewart et Parmar 1993 [109] ont conclu que les MADP ont tendance de surestimer l'effet du traitement. Les auteurs soulignent que ce comportement est dû au fait que les méta-analyses sur données individuelles peuvent inclure : les essais non publiés, la population en intention de traiter et avoir un suivi mis à jour. Nous nous concentrons sur les méta-analyses sur données individuelles.

5.2 Méta-analyse d'effets d'intervention

La première application de la méta-analyse fut l'estimation de l'effet traitement à partir de plusieurs essais randomisés contrôlés. Les agences réglementaires donnent à ce type de méta-analyse le niveau de preuve le plus élevé. [110], [111]

La création d'un protocole qui définit les critères d'inclusion et exclusion, la question scientifique et la méthodologie statistique est une étape incontournable quelle que soit la question de recherche (étude de l'effet d'une intervention ou étude de facteurs pronostiques).

Cette démarche a été formalisée dans les recommandations PRISMA '*Preferred Reporting Items for Systematic Reviews*' [112]. Il s'agit d'un guide de bonne conduite dans la réalisation d'une méta-analyse développé par des groupes collaboratifs d'experts multidisciplinaires. Le protocole peut être enregistré au registre international PROSPERO ('International prospective register of systematic reviews') (par exemple sous le numéro CRD42017068135 pour la méta-analyse sur le cancer des ovaires).

Pour l'étude d'une intervention, la méta-analyse repose sur l'exhaustivité de la collecte des données afin d'éviter un biais de publication qui résulte du fait que les résultats favorables sont plus facilement publiés ou rendus disponibles que les résultats défavorables.

L'enjeu est alors triple : (i) collecter l'ensemble des données produites dans le cadre d'essais randomisés répondant à la question, (ii) estimer l'effet de l'intervention en conservant la structure en grappe et le bénéfice de la randomisation, (iii) explorer les possibles sources d'hétérogénéité entre les différentes études.

5.2.1 Revue systématique

La plupart des méta-analyses sont des études rétrospectives. Par conséquent, il est indispensable de pouvoir identifier le plus grand nombre des essais traitant cette même question de recherche. Cette recherche consiste usuellement en une revue systématique, voire une méta-analyse sur données de la littérature. Pour cela, un algorithme de recherche de l'ensemble des études qui ont abordé une question est établi ; il permet d'interroger différentes bases de données, comme MEDLINE, Pubmed mais aussi des registres d'essais cliniques comme 'clinicaltrials.gov' ou autres registres de médicaments et les comptes-rendus des congrès afin d'identifier également des essais potentiellement non publiés pour atteindre à l'exhaustivité. Cette démarche permet de diminuer le biais de publication puisqu'il est observé que les essais positifs sont publiés plus vite et en anglais.

5.2.2 Vérification des données et standardisation

Suite à la collection et centralisation des données, l'étape suivante consiste à vérifier la qualité des données. Contrairement aux méta-analyses sur données publiées où la qualité de la méthodologie de l'étude est souvent confondue avec la qualité du rapport de l'étude, les données individuelles fournissent la possibilité de vérifier les éléments centraux que sont la procédure de randomisation, la population d'analyse, les critères d'évaluation. Tout d'abord, on s'attache à vérifier la qualité de randomisation : l'équilibre du rythme d'inclusion au cours du temps et celle des facteurs de stratification dans les groupes de randomisation. Ensuite, il est indispensable de contrôler la cohérence des variables recueillies (par exemple la cohérence chronologique des variables du temps, valeurs manquantes, valeurs extrêmes, etc...) et reproduire les résultats publiés afin d'éviter tout type d'erreur. La description de la base reçue et la reproduction des résultats sont envoyées au correspondant de l'essai pour validation. Si des incohérences sont identifiées, un ou plusieurs échanges sont nécessaires afin de comprendre l'origine du problème. Dans le cas où une étude a une qualité insuffisante, il est possible qu'elle soit exclue de l'analyse ou considérée uniquement en analyse de sensibilité.

Une fois la qualité des données validée, il est nécessaire de standardiser le codage et les unités des mesures à travers les études, par exemple uniformiser l'unité du temps en mois ou années. Dans cette étape, les variables propres à la méta-analyse sont créées. [113]

Les différentes étapes de la collecte -vérification jusqu'à l'analyse des données d'une méta-analyse sur données individuelles sont représentées à la figure 5.2.2.1

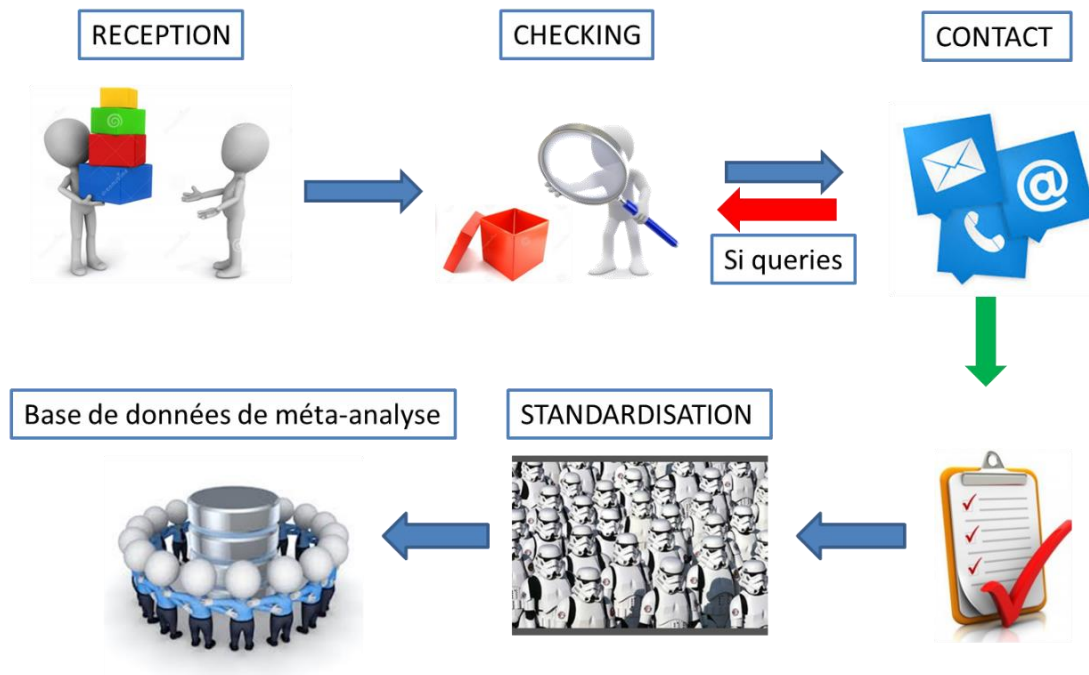


Figure 5.2.2.1: Circuit de la préparation de la base des données méta-analytique

5.2.3 Méthodes d'analyse

Deux grands modèles ont été proposés que nous introduisons ci-dessous : un modèle à effet fixe, aussi dit en deux étapes, et un modèle avec effets aléatoires.

Modèle en deux étapes :

Sous le modèle que les effets traitement pour chaque essai sont des variables aléatoires indépendantes et identiquement distribuées, alors les effets observés sont les réalisations d'une même loi de distribution sur la population d'intérêt dont on cherche à estimer l'effet moyen et son intervalle de confiance. Cet effet moyen est donc la moyenne des effets de chaque étude et sa variance est la moyenne des variances de chaque étude pondérée par sa précision ; la précision est classiquement quantifiée par l'inverse de la variance de l'effet traitement ; cette variance est proportionnelle à la taille de l'étude ou au nombre d'événements pour les analyses de données censurées. Ainsi, seuls les effets moyens dans chaque étude ainsi que leur écart type sont nécessaires pour obtenir une méta-analyse de l'effet traitement global. La première étape consiste donc à estimer l'effet dans chaque étude indépendamment des autres, et la seconde à les combiner pour estimer l'effet commun. La première étape pourra être ajustée sur les facteurs de stratification de chaque randomisation ou introduire des corrections selon le protocole choisi pour corriger d'éventuels biais systématiques.

Dans le cas sans ajustement où l'effet traitement est quantifié par le rapport des risques instantanés (sous l'hypothèse de proportionnalité des risques), Peto a proposé d'utiliser pour chaque étude c , l'estimation exprimée en fonction du nombre d'événements observés O_k et

attendus E_c dans le bras expérimental, en transformant le HR sur l'échelle logarithmique pour en normaliser la distribution.[114]

$$\ln(HR_c) = \frac{(O_c - E_c)}{V_c} \quad (5.1)$$

avec $V_c = Var(\ln(HR_c))$

Et l'estimation commune

$$\ln(\hat{HR}) = \frac{1}{C} \sum_c \ln(HR_c) \quad (5.2)$$

et

$$Var(\ln(\hat{HR})) = \frac{1}{C} \sum_c Var(\ln(HR_c)) \quad (5.3)$$

Les variations entre les effets traitements sont considérés comme des fluctuations d'échantillonnage. On peut donc tester si les variations des $\ln(HR_c)$ pour $c=1, \dots, C$ sont compatibles ou pas avec la distribution attendue sous cette hypothèse d'homogénéité des effets traitement. C'est le test d'hétérogénéité, appelé test de Cochran qui s'applique tant pour les rapport des risques instantanés (HR) que pour le Rapport des côtes (OR) ou le Risque Relatif (RR).

Comme tout test d'interaction, sa puissance dépend du degré d'hétérogénéité mais aussi du nombre d'essais dans la méta-analyse ; elle est souvent faible. C'est pour cela qu'Higgins a proposé une mesure de la quantité d'hétérogénéité, I^2 , directement dérivée de la statistique de test d'hétérogénéité de Cochran Q [115].

$$I^2 = \frac{Q - df}{Q} * 100\%$$

et définie sur $[0 ; 100]$. Où df dénote le nombre de degrés de liberté.

5.2.4 Modèle à effet aléatoire :

L'approche introduite par DerSimonian et Laird [73] est une variante de l'approche pondérée par l'inverse de la variance où l'hypothèse d'homogénéité des effets traitements est relâchée : les différentes études peuvent conduire à des estimations des effets traitements qui fluctuent au-delà des simples fluctuations d'échantillonnage même si toutes les estimations restent liées [116]. On obtient ainsi une méta-analyse à effets aléatoires.

Soit le modèle de survie à risques proportionnels pour la variable de traitement Z_c , dans l'essai c , (en supposant qu'il n'y a pas d'ajustement autre)

$$\log(h_c(t)) = \log(h_0(t)) + \beta_c Z_c$$

on considère que l'effet traitement β_c et sa variance notée v_c sont décrits par :

$$\hat{\beta}_c = \beta_c + \epsilon_c$$

avec $\epsilon_c \sim N(0, v_c)$

$$\beta_c = \beta + \eta_c$$

avec $\eta_c \sim N(0, \tau)$

β représente alors l'effet moyen (le log(HR)) sur l'ensemble des traitements. On peut réécrire le modèle en une seule étape en combinant les deux expressions.

Cette approche permet de traiter l'hétérogénéité qui ne peut pas être facilement expliquée par d'autres facteurs. Un modèle de méta-analyse à effets aléatoires suppose que les effets estimés dans les différentes études suivent une certaine distribution. Le modèle représente notre manque de connaissances sur les raisons pour lesquelles les effets réels ou apparents des traitements diffèrent. Le mode de la distribution aléatoire décrit la moyenne des effets, tandis que sa dispersion décrit le degré d'hétérogénéité. Le choix conventionnel de la distribution est une distribution normale pour les variables continues ou gamma pour les données de survie. Il est difficile d'établir la validité d'une hypothèse de distribution particulière, notamment lorsqu'il y a peu d'études et c'est une critique courante des méta-analyses à effets aléatoires. Distinguer la variation due à des différences de traitements ou de population des fluctuations d'échantillonnage est quasiment impossible sans introduire des hypothèses invérifiables à partir des données. En l'absence d'hétérogénéité, les deux approches fourniront les mêmes résultats. En cas d'hétérogénéité, la pondération respective de chacune des études changera fortement par rapport à ce qui est obtenu avec la méta-analyse à effet fixe : le modèle à effet aléatoire donne plus de poids aux études de petites tailles et dont l'effet est proche de l'effet moyen.

L'estimation commune et l'intervalle de confiance d'une méta-analyse à effets aléatoires correspondent au centre de la distribution des effets de l'intervention β_c , mais ne décrivent pas la dispersion de la distribution v_c . L'intervalle de confiance d'une méta-analyse à effets aléatoires décrit l'incertitude quant à la localisation de la moyenne β à travers les études. Il ne décrit pas le degré d'hétérogénéité entre les études. Par exemple, lorsqu'il y a de nombreuses études dans une méta-analyse, nous pouvons obtenir un intervalle de confiance très étroit autour de l'estimation de l'effet moyen des effets aléatoires, même s'il y a une grande hétérogénéité. Une solution à ce problème consiste à considérer un intervalle de prédiction. Un intervalle de prédiction à 95% autour de l'effet moyen HR estimé sur K études peut être calculé à partir du percentile de la distribution de Student t et de l'écart-type noté s.e comme :

$$\log(\hat{HR}) \pm t_{K-2} \sqrt{\tau^2 + s.e(\log(\hat{HR}))^2}$$

5.3 Méta-analyse de modèles pronostiques

5.3.1 Contexte

Tout naturellement, la méta-analyse s'est attachée à répondre à des questions de natures diverses, comme l'étude d'association, l'évaluation diagnostique, la quantification de la capacité pronostique de variables. La réalisation d'analyses jointes de plusieurs sources est ancienne. Ainsi l'initiative EPIC (*The European Prospective Investigation into Cancer and Nutrition*), lancée au début des années 90 sous l'égide de l'IARC consistait à réunir les données individuelles de différentes cohortes européennes (plus de 500 000 participants de 10 pays différents) qui exploraient les relations entre nutrition et état de santé en population générale. On peut aussi citer les méta-analyses de registres que ce soit au niveau national (FRANCIM) ou international.

Pour l'étude de facteurs pronostiques, ce n'est qu'assez récemment, qu'une réflexion méthodologique sur la réalisation de telles méta-analyses à partir de données individuelles a été initiée par des équipes hollandaises [117] et britanniques [118]. Alors que en 2001 Vickers et Altman pointaient déjà du doigt dans le BMJ [119] la médiocre qualité méthodologique des études de facteurs pronostiques ainsi que le manque de standardisation de leurs publications, et recommandait de lancer des méta-analyses sur données individuelles, il faut attendre le début des années 2010 pour que des recommandations soient publiées sous l'égide de la COCHRANE. Le groupe de travail PROGRESS (PROGnosis RESearch Strategy) a proposé un guide pour ce type de méta-analyses [120] en détaillant chaque étape. Il met le point sur la vérification du risque de biais des essais avec l'outil QUIPS et insiste sur la réalisation de différents méta-analyses dépendant du type des effets pronostiques (ajusté/non ajusté) et pour chaque type d'estimation (odds ratio/hazard ratio). Il reprend de nombreux éléments de TRIPOD '*Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis*' [121] qui avait contribué à standardiser la réalisation des études pronostiques ou diagnostiques.

En fait, on peut distinguer deux types de méta-analyses : (i) la synthèse (ou agrégation) de modèles de prédictions différents dans une méta-analyse (ii) l'estimation et validation d'un modèle pronostique à partir d'une méta-analyse. Ce travail s'inscrit exclusivement dans le second cas, mais j'introduis brièvement la synthèse de modèles différents car elle a donné lieu à des recommandations dont certaines sont applicables aux deux situations.

5.3.2 Synthèse de modèles

Ces approches concernent essentiellement la synthèse de différents modèles de prédiction, plutôt que la construction d'un modèle unique à partir de différents échantillons. Si M modèles de prédiction ont été proposés dans la littérature, deux grandes approches statistiques sont revues : (i) la moyenne des modèles, c'est-à-dire de calculer la prédiction obtenue de chacun des M modèles et en effectuer la moyenne pondérée par la taille de l'effectif utilisé ou (ii) par la valeur de l'information BIC (*Bayesian Information Criteria*), cette méta-régression consiste à prendre une combinaison linéaire des différents modèles [122].

Brièvement, Debray et al. notent $M_k, k = 1, \dots, K$ les K modèles pronostiques à combiner, chaque modèle reposant sur une combinaison linéaire LP_{ik} pour chaque sujet $i = 1, \dots, N$ et

une fonction notée $g(\cdot)$ (logistique, Poisson etc.) appliquée à l'échantillon de validation pour calculer un résultat prédit.

Le modèle de prédiction pour le patient i est écrit :

$$p_{ik}^{\wedge} = g^{-1}(LP_{ik}) \quad (5.4)$$

Une prédiction moyenne pondérée par la distribution *a posteriori* du modèle sera :

$$\bar{p}_i = \sum_{k=1}^K w_k p_{ik}^{\wedge} \quad (5.5)$$

Avec

$$w_k = \frac{\exp(-0.5BIC_k)}{\sum \exp(-0.5BIC_l)} \quad (5.6)$$

Un possible alternatif plus simple des poids sera de considérer $w_k = \frac{1}{K}$ pour chaque modèle.

Les prédictions \bar{p}_i sont utilisées pour développer le métamodèle. Pour ce faire, les auteurs proposent une analyse de régression linéaire dans laquelle les variables prédictives originales sont utilisées comme variables indépendantes et les prédictions moyennées comme variables dépendantes. Pour garantir la linéarité de la variable dépendante, la transformation logistique de cette dernière est indispensable et elle est donnée sous la forme $z_i = \text{logit}(\bar{p}_i)$. Les variables prédictives des modèles de la littérature pour lesquelles $w \geq 0.0001$, seront incluses dans le modèle. L'analyse de régression linéaire est alors la suivante :

$$z_i = \beta_0 + \sum_m^{M_{AVG}} \beta_m x_{im} + \varepsilon_i \quad (5.7)$$

Où $\varepsilon_i \sim N(0, \sigma^2)$, M_{AVG} est le nombre total de prédicteurs, β_0 est l'intercept du modèle et β_m les effets des prédicteurs.

En résumé, le calcul de la moyenne des modèles repose sur l'estimation de chaque modèle de la littérature à partir de l'échantillon de validation. Ensuite, les K paramètres additionnelles sont calculées pour obtenir les pondérations du modèle. Enfin, la synthèse est réalisée par régression linéaire dont les paramètres sont à nouveau estimés avec $1 + M_{AVG}$ paramètres à estimer à partir de l'échantillon de validation.

On remarquera que ces approches correspondent aux approches des « *super-learner* » utilisées en apprentissage statistique, sans toutefois la composante de la cross validation, nécessaire lorsqu'on combine apprentissage et validation.

La question de l'hétérogénéité suite à la présence de différences entre les études comme par exemple aux critères d'inclusion, le centre ou le pays peuvent induire des difficultés à l'interprétation des résultats. Une stratégie d'analyse avec des méthodes statistiques appropriées permettent de prendre en compte ces particularités. L'évaluation de l'hétérogénéité doit être quantifiée. Le test de Cochran et la statistique I^2 [123] permettent d'estimer cette quantité. Les différences entre les critères des essais et la prise en charge des patients peuvent entraîner des différences du risque de base des patients et donc l'effet l'essai doit être considéré à l'analyse. [114], [124]

5.3.3. Qualité des publications de méta-analyses et recommandations

Ahmed et al. [125] effectue une revue de littérature sur la construction et validation d'un modèle pronostique à partir d'une méta-analyses de plusieurs jeux de données. Sur 15 articles, ils trouvent que la plupart développent leur modèle en utilisant les données patientes de toutes les études disponibles et effectuent une validation interne sur le même ensemble de données. Dix des 15 articles ne tiennent pas compte des différences entre les études en ce qui concerne le risque de base, ce qui limite potentiellement l'applicabilité et la performance de leur modèle dans certaines populations. Seuls deux articles avaient utilisé une validation externe (sur des données différentes). Ils proposent un tableau de recommandations dont nous citons quelques éléments traduits. (cf. annexes table 3) [125]

Les méta-analyses sur données individuelles font partie de la base fondamentale pour le développement de la modélisation des études prédictives et pronostiques. En effet, elles permettent la construction de modèles de prédiction robustes et puissants, et la validation directe via la base des données méta-analytique. Cette dernière, disposant de plusieurs populations, a l'avantage de permettre la généralisation des résultats avec des méthodes statistiques adaptées.

Dans le cas de l'évaluation de l'utilité d'un biomarqueur mesurée de façon répétée dans le temps, le nombre de modèles disponibles n'est souvent pas très grand et beaucoup ont été construits de façon ad-hoc avec des défauts statistiques (prendre le nadir qui conduit à un risque de biais d'immortalité, absence de gestion des données manquantes etc.) qui les rendent inutilisables. C'est donc moins la recherche d'un modèle combiné que la construction et la validation d'un modèle pronostique du CA-125 que nous allons développer.

5.3.4 Mesure de qualité des études à inclure dans une méta-analyse

Il est important d'inclure dans la méta-analyse des études de bonne qualité et identifier la présence de possible biais. Mais l'évaluation de la qualité des études à inclure est également plus délicate qu'avec les méta-analyses d'essais randomisés d'intervention ; en effet alors que

la bonne réalisation du tirage au sort et un suivi similaire dans les deux groupes de patients sont des éléments décisifs de qualité d'une étude interventionnelle, la qualité des études observationnelles de facteurs pronostiques est beaucoup plus « multifactorielle » et difficile à évaluer. Plusieurs groupes ont proposé des lignes directrices de bonnes pratiques couvrant dès la construction du protocole, définition de la méthodologie jusqu'à la présentation des résultats. [121], [126]

Une nouvelle version du TRIPOD*centrée sur les méta-analyses des modèles prédictifs et pronostiques, a été récemment proposé [121] (c.f annexes table 4). Cette dernière permet de comparer des études ayant utilisé différentes méthodes prédictives ou pronostiques pour une même pathologie ; permettant de résumer et comparer les performances prédictives de deux ou plusieurs modèles de prédiction.

5.4 Construction de facteurs pronostiques sur données individuelles de méta-analyse

5.4.1 Méthodes statistiques pour la méta-analyse d'un biomarqueur longitudinal

La combinaison des statistiques résumées par la méthode de Peto de l'inverse de la variance généralisée n'est plus adaptée pour une association estimée d'une étude observationnelle qui repose sur la construction d'un modèle de prédiction qui peut dépendre de l'étude ; en particulier un modèle longitudinal de biomarqueur fait intervenir un schéma expérimental qui diffèrera en termes de nombre et timing des mesures, de variables d'ajustements, voire de technique de mesure. On cherche donc à construire un modèle pronostique utilisant toutes les données de toutes les études qui permette d'identifier des paramètres de populations, des paramètres spécifiques de l'étude et des paramètres spécifiques du patient afin qu'il soit largement applicable. On retrouve la notion d'apprentissage et de validation introduite précédemment. Comme vu en section 3.1.3, la construction du score pronostique se divise en deux étapes qui chacune doivent être étendues à la méta-analyse : la construction du modèle longitudinal et la construction du modèle pronostique du temps d'événement. Abo-Zaid et al [127] ont bien montré par une étude de simulation l'impact de ne pas prendre en compte l'effet grappe induit par les méta-analyses de facteurs pronostiques, trop souvent ignoré.

5.4.2. Modèle longitudinal

Les modèles à effets aléatoires offrent un cadre adapté à la prise en compte de différents niveaux de corrélation. On distingue en effet, l'effet patient pour prendre en compte la répétition des mesures du biomarqueur chez chaque patient, de l'effet étude pour prendre en compte l'hétérogénéité entre les centres venant de différents facteurs non mesurés liés aux critères d'éligibilité des différents essais, aux possibles variations de prise en charge des patients liées aux structures de soin, aux possibles effets temporels avec des études réalisées sur près de 20 ans etc.

Indiquons par c l'essai, le modèle linéaire pour un patient de l'étude ' c ' s'écrit comme suit :

$$Y_{ic} = \begin{pmatrix} y_{i1c} \\ y_{i2c} \\ y_{i3c} \\ \vdots \\ y_{in_i c} \end{pmatrix} = X_i^T \beta + Z_i^T b_{ic} + \mathbb{1}_{n_i} u_c + \varepsilon_i \quad (5.8)$$

Où X_i et Z_i sont des vecteurs des covariables possiblement temp-dépendants associés aux vecteurs des effets fixes β et au vecteur des effets aléatoires $b_{ic} \sim N(0, D)$, respectivement. La matrice X est de dimension $n_i \times p$ et le vecteur Z est de dimension $n_i \times q$. Le terme ε_{ic} correspond à l'erreur des mesures qui suit une distribution normale $N(0, \sigma^2 I_{n_i})$. Le vecteur u_c correspond aux effets aléatoires de l'essai et suit une loi normale : $N_q(0, \sigma_u^2)$.

Dans ce modèle hiérarchique, il existe deux grandes structures d'effets aléatoires : les effets emboîtés ou croisés. Dans le cas d'une méta-analyse où chaque patient appartient à une étude unique, la situation est celle des d'effets patients emboîtés dans l'effet étude, ce qui implique que la variance inter-patient est supposée la même conditionnellement à l'appartenance à une étude donnée.

Pour notre travail, nous supposons en outre que les effets aléatoires au niveau patient sont indépendants parmi chaque patient. Mais il y a la possibilité de modéliser une corrélation non nulle qui correspondrait par exemple à l'hypothèse que la pente pour un patient dépend de la valeur initiale de départ du calcul de la pente.

En groupant les effets aléatoires dans un seul vecteur, et en considérant l'équation matricielle du modèle mixte (i.e., pour toutes les valeurs répétées d'un individu simultanément), nous obtenons l'équation suivante :

$$Y_{ic} = X_{ic} \beta + Z_{ic}^* d_{ic} + \varepsilon_{ic} \quad (5.9)$$

Où, en supposant une évolution linéaire non ajustée sur des variables explicatives,

$$Z_{ic}^* = \begin{pmatrix} 1 & F_1(t_{i1}) & F_p(t_{i1}) & 1 \\ 1 & F_1(t_{in_i}) & F_p(t_{in_i}) & 1 \end{pmatrix}$$

&

$$d_{ic} = \begin{pmatrix} b_{ic} \\ u_c \end{pmatrix} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} D & 0 \\ 0 & \sigma_c^2 \end{pmatrix}\right), \text{ noté } N(0, B)$$

Où F est la fonction multivariée du temps

La distribution marginale s'écrit avec la forme suivante :

$$Y_{ic} \sim N(X_{ic}\beta, V_{ic} = Z_{ic}^* B Z_{ic}^{*'} + \sigma_c^2 I_{n_i})$$

à partir de laquelle on peut dériver la log-vraisemblance. On considère $\vartheta = (\text{Vect}(B), \sigma_c^2, \beta^\top)^\top$

$$l_N(\vartheta) = -\frac{1}{2} \sum_{i=1}^N \{n_i \log(2\pi) + \log(|V_{ic}|) + (Y_{ic} - X_{ic}\beta)^\top V_{ic}^{-1}(d)(Y_{ic} - X_{ic}\beta)\} \quad (5.10)$$

La vraisemblance, qui a une structure similaire à celle d'un modèle linéaire mixte classique, peut être maximisée par un algorithme itératif pour obtenir les estimateurs du maximum de vraisemblance.

5.5. Modèle de survie

Pour prendre en compte la structure en grappe, une approche par modèle à fragilité [128] est possible ; une distribution gamma pour le paramètre de fragilité est usuellement proposée qui permet de prendre en compte la dispersion des fonctions de risque instantané sur \mathbb{R}^{+*} . C'est l'approche retenue par Debray et al. [117] qui spécifient leur modèle comme suit :

$$h(t|X_i) = g(a, t) \exp(\delta X_i) \quad (5.11)$$

où $g(\cdot)$ est une fonction de paramètres a connue. La distribution exponentielle est un exemple simple qui suppose un risque constant λ au cours du temps. Un effet aléatoire, ou fragilité, permet de prendre en compte la variation du risque de base entre les études, ce qui donne:

$$h(t|X_{ic}) = \zeta_c \lambda \exp(\delta X_{ic}) \quad (5.12)$$

avec $\zeta_c \sim \Gamma(1, \theta_0)$ qui est l'effet étude

Le paramètre de variance θ_0 peut être interprété comme une mesure de l'hétérogénéité du risque de base entre les études. Lorsque θ_0 est faible, les valeurs de ζ_c sont étroitement concentrées autour du 1 et les effets de l'étude sont faibles. Si θ_0 est grand, les valeurs de ζ_c sont plus dispersées, ce qui induit une plus grande hétérogénéité dans les risques de base $\zeta_c \lambda$ spécifiques à l'étude. Ce modèle s'appuie aussi sur l'hypothèse que le risque d'événement est proportionnel en tout temps entre les centres. Le risque de base peut aussi être modélisé par d'autres fonctions comme par exemple des splines.

L'hypothèse de distribution de la fragilité est délicate à valider sur les données du fait du petit nombre d'études. O'Quigley and Stare ont montré que dans ce cadre d'un petit nombre de grappes chacune avec une large taille d'échantillon, le modèle de survie stratifié apportait la même efficacité statistique pour l'estimation des paramètres qu'un modèle à effet aléatoire. [129]

$$h(t) = h_{oc}(t)exp(\delta X) \quad (5.13)$$

C'est l'approche que nous retenons. Le risque de base $h_{oc}(t)$ est alors spécifique de chaque étude et admet des formes libres et possiblement différentes. Le risque de base de chaque étude est estimé de façon non-paramétrique, ce qui peut être fait avec précision lorsqu'on travaille sur des méta-analyses d'études de grandes tailles et avec un grand nombre d'événements.

Qu'on utilise l'approche à fragilité ou l'approche stratifiée, le modèle d'association est supposé le même à travers toutes les études. On suppose donc qu'il n'y a pas d'hétérogénéité de l'association à travers les études. Cette hypothèse a une implication directe sur la notion de validation et validation croisée ; l'appartenance d'un patient à une étude renseigne sur le risque d'événement et donc participe à la calibration du facteur pronostique ; en revanche étant connue l'étude, le rapport de risque entre deux patients est le même pour tous les patients.

5.6. Estimation des paramètres pour un nouveau patient d'une nouvelle étude.

Dans ce travail, avec une approche basée sur du landmark et un résumé d'information du CA125 basé sur des modèles mixtes, la prédiction de la probabilité d'événement pour une nouvelle patiente nécessite la prédiction des effets aléatoires au niveau de la patiente et de l'étude.

Pour calculer l'effet aléatoire de l'étude nous avons besoin des données de toutes les patientes appartenant à cette étude.

Par la suite nous intéresserons aux deux questions.

→ Comment estimer cet effet essai pour une nouvelle patiente appartenant à une nouvelle étude ?

→ Dans le cadre d'un modèle pronostic, nous avons besoin de valider ce dernier dans des données d'une nouvelle étude (validation externe). Par conséquent, nous nous questionnons sur quelle sera dans ce cas l'effet étude, qui est indispensable pour la méthode de modélisation que nous avons utilisée ?

5.6.1 Prédiction individuelle dans le cadre de la méta-analyse

Afin d'obtenir les prédictions individuelles, il faut estimer les valeurs des effets aléatoires pour chaque patient ainsi que pour chaque étude. Dans le cadre d'effets aléatoires emboîtés, on généralise aisément l'estimateur 'Best Linear Unbiased Predictor' (BLUP) tiré de l'espérance des effets aléatoires conditionnellement aux observations au cas d'un modèle à 3 niveaux (essais,

patient, mesure de biomarqueur); pour rappel, c dénote l'étude incluse dans la méta-analyse, i indexe le patient et on le calcule comme suit à partir de la distribution marginale (*):

$$\hat{d}_{ic} = E(d_{ic} | y_{ic}) = BZ_{ic}^{\top} V_{ic}^{-1} (y_{ic} - X_{ic} \beta) \quad (5.14)$$

Les prédictions estimées avec l'aide de l'équation (3.5) par le modèle linéaire mixte sont les suivantes :

(i) la valeur de $\widehat{Y}_i(t_l)$ estimée au temps t_l vaut :

$$\widehat{Y}_{ic}(t_l) = \widehat{\beta}_c F(t_l) + \widehat{b}_{ic} F(t_l) + \widehat{u}_c \quad (5.15)$$

(ii) La pente de la trajectoire de Y_i estimée à l'origine t_0 :

$$\frac{\partial \widehat{Y}_{ic}}{\partial t}(t_0) = \widehat{\beta}_c \frac{\partial F(t_0)}{\partial t} + \widehat{b}_{ic} \frac{\partial F(t_0)}{\partial t} \quad (5.16)$$

(iii) la pente de la trajectoire de Y_i estimée au temps t_l :

$$\frac{\partial \widehat{Y}_{ic}}{\partial t}(t_l) = \widehat{\beta}_c \frac{\partial F(t_l)}{\partial t} + \widehat{b}_{ic} \frac{\partial F(t_l)}{\partial t} \quad (5.17)$$

où F Correspond à la fonction 3.4 du chapitre 3.5.1

5.6.2 : Prédiction pour un patient futur et recalibration

La distribution marginale des prédictions fait intervenir l'effet étude (ou grappe). Si cette distribution permet d'évaluer les performances prédictives en comparant valeurs observées et prédites sur une méta-analyse (cf. prochaine section), la question de son applicabilité pour des patients futurs impose d'obtenir une estimation de l'effet étude. Comment estimer cet effet étude qui contribue à l'intercept de la trajectoire lorsqu'on cherche à faire une prédiction pour un nouveau patient. Cette question fait donc écho à la question de la (re)calibration d'un modèle pronostique. Bien entendu, elle ne se pose qu'en présence d'une certaine hétérogénéité entre les études tant pour l'intercept du modèle longitudinal que pour le risque de base. C'est toutefois une situation courante.

On distinguera deux cas : le patient fait partie d'une nouvelle étude et les données de tous les participants à cette étude (ou cohorte) sont disponibles et de telles données n'existent pas.

Dans le premier cas, le plus simple, l'effet « étude » u_c est prédit en appliquant directement la formule du chapitre 5.4.2 à partir des données de cette nouvelle étude. On peut imaginer en routine clinique, que l'ensemble des patients avec la maladie considérée au sein d'un hôpital constituent une nouvelle cohorte. *L'empirical Bayes factor* de l'effet cohorte pourrait être estimé [130], et mis à jour régulièrement afin de fournir des prédictions correctement calibrées pour un patient de l'hôpital.

En revanche, si de telles données n'existent pas, plusieurs solutions ont été recensées par [117] ; chacun repose sur une hypothèse concernant le nouveau patient.

1) Calculer un intercept moyen

Dans le cas du modèle linéaire à effets aléatoires, cela correspond à $u_c=0$ comme si le patient venait d'une étude similaire à l'effet moyen. Si on peut considérer que $u_c=0$ correspond aussi au mode de la distribution a posteriori sur la cohorte, et donc à la valeur maximum de la densité, c'est une approche qui ne fait aucun usage de ce qu'on peut savoir sur le nouveau patient. Pour le modèle de survie, on peut, de la même façon reprendre, la proposition de Royston et al. d'estimer une fonction de risque de base marginale [131] ; Royston et al utilisent des fonctions paramétriques flexibles, comme les Splines ou les polynômes fractionnaires pour ajuster les fonctions de risque de base dans chacune des études qui peuvent ensuite facilement être marginalisées en prenant l'intégrale sur l'espace des paramètres. Une approche alternative proposée par Peto pour les méta-analyses en présence de risques non proportionnels repose sur une estimation actuarielle du risque de base dans chaque étude qui permet de construire des courbes de survie moyennes [108]. Cette approche a l'avantage de rester plus proche de l'esprit d'une estimation non paramétrique d'Aalen tel que présenté en section 4.4. Quelle que soit l'approche retenue, elle impose d'avoir centré les différentes variables incluses dans le modèle pronostique pour que le risque de base corresponde bien au risque d'une population « moyenne » et non au risque instantané d'une population de référence qui pourrait ne pas exister (par exemple les patients avec un biomarqueur à 0).

2) Sélectionner un intercept

Pour éviter de considérer une valeur moyenne pour l'intercept spécifique au centre dans le modèle longitudinal, une autre approche consiste à sélectionner simplement un intercept estimé à partir de l'une des études IPD qui est la plus similaire à la nouvelle population étudiée. Cet intercept peut être directement obtenu à partir de u_c pour la partie modèle mixte sur le biomarqueur et de $h_{0c}(t)$ pour le risque stratifié d'événement. Ainsi, on pourrait sélectionner l'étude de la méta-analyse réalisée dans le même pays que celui du patient dont on cherche à faire la prédiction, ou à la même époque etc. Bien qu'il semble préférable que ce choix repose sur l'expertise clinique, il est possible de s'appuyer sur une approche purement statistique. Cette approche pourrait, par exemple, évaluer la similarité en comparant les caractéristiques

des patients inclus ou la fréquence des résultats de chaque IPD de dérivation avec la nouvelle population où le modèle doit être appliqué. Cette approche est adoptée par Steyerberg et al. qui ont choisi l'étude dont le pronostic était la plus proche de ce qui était attendu pour le nouveau patient [132].

3) Construction des paramètres propres au centre sur données externes

Un peu comme dans la situation où le nouveau patient ferait partie d'une étude permettant l'estimation d'un intercept spécifique pour le marqueur, si on dispose de certaines informations sur la population dont est échantillonné le nouveau patient, comme la distribution du biomarqueur à l'inclusion, ou l'incidence d'événements cliniques à un temps donné, on peut reconstruire l'intercept ainsi que le risque de base sous certaines hypothèses paramétriques, un peu à la façon dont on calcule des tailles d'échantillons.

La difficulté pour cette nouvelle population est de trouver de critères objectifs qui sont présents pour les deux populations comme par exemple le sexe ou la situation géographique du patient. Une attention particulière doit être apportée aux caractéristiques subjectives comme par exemple la douleur qui peut être perçue différemment d'un patient à un autre ainsi que par exemple la présence de « convulsions atypiques » qui peut être définie différemment par les médecins. La taille de l'échantillon et l'âge sont aussi des facteurs importants lors de la validation externe d'un modèle. [133]

Lorsque on s'intéresse à la validation du modèle, le problème de calibration est souvent présent. Au lieu de reconstruire un nouveau modèle ils existent des méthodes de recalibration. Ensor et al ont ciblé des méthodes de recalibration pour la validation externe et la recalibration d'un modèle existant dans le cadre des méta-analyses sur données individuelles.

La recalibration du modèle existant en combinant les informations provenant à la fois des patients historiques et des nouveaux patients peut-être une solution. Plusieurs possibilités sont envisageables : (i) le déplacement du risque de base par une constante, ii) la réestimation de la forme du risque de base, iii) ajustement de l'indice pronostique dans son ensemble et iv) ajustement des effets des prédicteurs individuels. Pour chaque stratégie, la méta-analyse IPD examine (l'hétérogénéité dans) la performance du modèle à travers les populations. En outre, la probabilité d'obtenir de bonnes performances dans une nouvelle population peut être calculée, ce qui permet de classer les méthodes de recalibrage.[16]

Dans le cadre d'une méta-analyse, un modèle peut être évalué plusieurs fois avec chaque étude. Par conséquent cela conduit à plusieurs statistiques de validation et donc une méta-analyse pour comparer et résumer la performance parmi les études et étudier l'existence d'hétérogénéité.

La question de la performance des modèles pronostiques est abordée dans le chapitre suivant.

5.7 La méta-analyse MAOV

5.7.1 Méthodes

La méta-analyse MAOV [134] a été initiée par le GCIg pour évaluer la PFS en tant que critère de substitution pour la OS chez les patientes atteintes d'un cancer de l'ovaire avancé. En septembre 2016, une recherche exhaustive des publications sur le cancer de l'ovaire avancé a été effectuée dans MEDLINE. Les groupes du GCIg ont été également interrogés sur les essais potentiellement finis mais non publiés.

Les essais éligibles étaient des essais cliniques randomisés de traitements systémiques chez des patientes atteintes d'un cancer de l'ovaire non traité auparavant (ou étudiant un traitement d'entretien après un traitement systémique de première ligne) avec une taille d'échantillon minimale de 60 patientes au total et publiés entre le 1er janvier 2001 et le 25 septembre 2016, avec des données disponibles sur la OS et la PFS. Les traitements expérimentaux pris en compte étaient les traitements initiaux, d'entretien et d'intensification. Les investigateurs de tous les essais identifiés qui répondaient aux critères d'éligibilité ont été contactés pour établir le partage des données

Nous avons demandé des données pour tous les patients individuels (qu'ils aient été inclus ou non dans l'analyse primaire) recrutés dans chaque essai. La survie globale a été définie comme le temps écoulé entre la randomisation et le décès toutes causes confondues ou la date du dernier suivi utilisée pour la censure. La survie sans progression a été définie comme le temps écoulé entre la randomisation et au premier événement survenu parmi : la progression ou le second cancer lorsque cette information était disponible, le décès toutes causes confondues ou la date du dernier suivi utilisée pour la censure. Des informations détaillées sur le type de progression ont été demandées ; elles comprenaient la définition de la progression, l'évaluation radiologique et/ou clinique qui documentait la progression, et les mesures en série des niveaux du biomarqueur CA125. Toutes les données ont été réanalysées de manière centralisée et vérifiées pour détecter les incohérences. En particulier, des outils de diagnostic de la qualité de la randomisation ont été systématiquement appliqués. [135], [136]

Des diagrammes de forêt ('forest' plot) ont été utilisés pour afficher les rapports de risque (RR) dans l'ensemble et pour les essais individuels. Les HR ont comparé le risque d'un événement chez les patients traités avec un traitement expérimental avec le risque chez les patients recevant le traitement de contrôle. Une approche à effet fixe a été mise en œuvre et les HR ont été obtenus à partir des nombres attendus et observés d'événements, vu au chapitre 5.2. Le HR regroupé a ensuite été ajusté en fonction de l'essai. Le test d'hétérogénéité et la statistique I^2 , définis dans le chapitre 5.2.3, ont été utilisés pour étudier l'hétérogénéité globale entre les essais. [137] Les courbes de survie ont été estimées à l'aide de l'approche actuarielle de Peto [138] pour tenir compte des essais multiples. L'évolution de la durée médiane de survie a été évaluée à l'aide d'un test de tendance linéaire au niveau de l'essai, pondéré par le nombre d'événements. La PFS a été évaluée en tant que critère de substitution de l'OS mais cette analyse ne sera pas présentée puisque la méthodologie utilisée n'a pas été détaillée dans ce manuscrit.

5.7.2 Résultats

37 essais ont été identifiés à partir de la recherche documentaire et leurs investigateurs ont été contactés. Des données individuelles sur 11 029 patientes ont été obtenues uniques issus de 17 essais cliniques randomisés éligibles uniques avec OS et PFS documentés. La table 5 (annexes) énumère les caractéristiques des 17 études au niveau de l'essai. Parmi les patientes, 7 436 ont connu une progression et 5 138 sont décédées au cours de l'étude. L'OS médiane a varié de 2.7 à 6.2 mois et la PFS médiane a varié de 0.9 à 2.3 mois (voir le tableau 5 annexes). La figure 2 (annexes) présente un diagramme de forêt des effets du traitement sur l'OS et la PFS pour tous les essais. Dans l'ensemble et au niveau des essais, les effets de la chimiothérapie expérimentale sur l'OS et la PFS étaient presque nuls (HR pour la PFS, 0.97 [IC à 95 %, 0.93-1.02] ; HR pour la SG, 0.99 [IC à 95 %, 0.94-1.05]). Aucune hétérogénéité entre les essais n'a été détectée pour aucun des critères d'évaluation ($I^2 = 0\%$ [$p = 0.70$] pour l'OS et $I^2 = 0\%$ [$p = 0.60$] pour la PFS) (annexes figure 2)

6. MESURES DE LA CAPACITE PRONOSTIQUE

La performance prédictive d'un modèle de prédiction est déterminée avec l'évaluation de sa capacité de discrimination et de calibration. [13] Plus précisément, il s'agit d'évaluer la capacité du modèle à classer les individus par ordre de risque et à prédire correctement la survenue d'un évènement. La calibration quantifie la différence entre le risque prédit par le modèle et le risque observé et la discrimination d'identifier les individus avec et sans évènement. Une simple évaluation de la qualité prédictive est possible avec la visualisation de la probabilité de survie prédite vs celle observée.

Plusieurs points doivent être prise en compte lors de l'évaluation des facteurs pronostics. Tout d'abord, l'état des patients qui change au fil du temps, la censure du temps jusqu'à l'évènement, et le délai entre l'entrée dans l'étude et l'apparition de la maladie peuvent varier. La nature des données répétées et l'existence d'évènements concurrentes doivent être également pris en compte. Blanche et al. ont proposé des méthodes pour pallier ces particularités. [139]

6.1 Evaluation de la calibration et la discrimination

6.1.1 L'aire sous la courbe ROC

Lorsqu'on s'intéresse à l'évaluation d'un modèle pronostique et prédictif, on parle souvent de leur capacité de calibration et discrimination. La discrimination permet d'évaluer les capacités diagnostiques d'un test et ici dans notre cas d'un biomarqueur. Plus précisément, il s'agit de sa capacité d'identifier et séparer les individus entre malades et non malades. Habituellement, l'aire sous la courbe ROC est utilisé pour évaluer la capacité discriminante d'un marqueur

continu. Cette méthode a été étendue dans le cadre des modèles multivariés avec les particularités suivantes : elle est invariante aux transformations monotones des valeurs des marqueurs et s'exprime donc avec les rangs des probabilités prédites. [13]

On suppose pour l'individu i , la covariable X_i ; et la réponse binaire.

$$D_i = \left\{ \begin{array}{l} 1, \text{evenement} \\ 0, \text{absence d'evenement} \end{array} \right\}$$

Pour un seuil $c \in R$ choisi, la sensibilité $Se(c, t)$ et spécificité $Sp(c, t)$, correspondant respectivement aux proportions des vrais positifs et des vrais négatifs, sont définies par les fonctions suivantes :

$$Se(c, t) = P(X > c \vee D(t) = 1)$$

Et

$$Sp(c, t) = P(X \leq c \vee D(t) = 0)$$

La courbe ROC est la représentation graphique de $Se(c, t)$ vs $1 - Sp(c, t)$ pour tous les possibles valeurs de c

$$ROC(t) = \{(1 - Sp(c, t), Se(c, t)), c \in R\} \quad (6.3)$$

La courbe ROC permet d'évaluer les probabilités prédites d'un modèle avec l'aide de deux indicateurs : la sensibilité et la spécificité. Elle est comprise dans le secteur $[0,1] \times [0,1]$ et plus elle est élevée dans ce secteur, meilleure est sa capacité à discriminer les sujets à haut risque d'évènement des sujets à bas risque d'évènement. L'aire sous cette courbe (AUC) est une métrique comprise entre $[0, 1]$ qui permet de résumer le pouvoir discriminant d'un outil de prédiction. Elle correspond à la probabilité de prédire plus vraisemblablement un événement pour un patient ayant l'évènement que pour un patient n'ayant pas l'évènement. La courbe ROC est représenté par la première bissectrice pour le cas où aucune discrimination est possible entre $D_i = 0 \wedge D_j = 1$. Une AUC supérieure à 0.7 est considérée comme un indicateur de bonne performance alors qu'un indicateur de 0.5 indique que le modèle ne fait pas mieux qu'un classement au hasard.

L'aire sous la courbe ROC est égal à la probabilité conditionnelle pour deux individus sélectionnés au hasard avec statut d'évènement différent [13]:

$$AUC(t) = \int_0^1 ROC(u) du = P\{X_i < X_j \vee D_i(t) = 0, D_j(t) = 1\} \quad (6.4)$$

Habituellement, l'aire sous la courbe ROC est utilisé pour évaluer la capacité discriminante d'un marqueur continu. Cette méthode a été étendue dans le cadre des modèles multivariés avec

les particularités suivantes : elle est invariante aux transformations monotones des valeurs des marqueurs et s'exprime donc avec les rangs des probabilités prédites.

Pour le cas d'un marqueur continu, par exemple pour le CA-125, des valeurs plus grandes sont liées à un risque d'évènement plus important alors qu'une relation inverse est possible pour d'autres marqueurs. La courbe ROC permet : (i) d'évaluer la capacité discriminante des prédictions et (ii) de déterminer un seuil $c \in R$ optimal qui maximise la proportion de vrais positifs et minimise la proportion des faux négatifs. [10]

Il est souvent intéressant de pouvoir quantifier l'évaluation des prédictions. Le choix d'un seuil c , permet de dichotomiser le résultat de cette évaluation et faciliter son interprétation. Par exemple, si la présence de faux négatifs compromet l'évolution de la maladie, un seuil bas est recommandé ; alors que si on s'intéresse à diminuer la présence de faux positifs, le choix d'un seuil plus élevé est recommandé.

Deux mesures sont en lien direct avec la sensibilité et la spécificité et ce seuil c : la valeur prédictive positive (VPP) et la valeur prédictive négative (VPN). Plus précisément, la valeur prédictive positive correspond à la probabilité d'avoir un évènement alors que la valeur du biomarqueur est supérieure au seuil fixé c . La valeur prédictive négative correspond à la probabilité de ne pas avoir un évènement alors que la valeur du biomarqueur est inférieure au seuil c . L'avantage de ces deux mesures porte sur l'évaluation de l'intérêt clinique du seuil fixé du biomarqueur.

Celles-ci sont définies par les probabilités suivantes :

$$VPP(c) = P(D(t) = 1 | X > c)$$

et

$$VPN(c) = P(D(t) = 0 | X \leq c)$$

Sur un échantillon de n individus, nous identifions les 4 situations possibles dans un tableau de contingence :

Tableau 1: Tableau de contingence VPP, VPN

	Présence d'évènement	
	Oui	Non
Seuil Positif ($X > c$)	$v_D^+ = \sum_{i=1}^n \mathbb{1}_{(X_i > c, D_{i=1})}$	$v_D^+ = \sum_{i=1}^n \mathbb{1}_{(X_i > c, D_{i=0})}$

Seuil Négatif ($X \leq c$)	$v_D^- = \sum_{i=1}^n \mathbb{1}_{(X_i \leq c, D_{i=1})}$	$v_D^- = \sum_{i=1}^n \mathbb{1}_{(X_i \leq c, D_{i=0})}$
------------------------------	---	---

Les estimateurs de \widehat{VPP} et \widehat{VPN} sont définis par les équations suivants [140]:

$$\widehat{VPP} = \frac{v_D^+}{v_D^+ + v_D^-}$$

$$\widehat{VPN} = \frac{v_D^-}{v_D^+ + v_D^-}$$

Des extensions de la sensibilité et spécificité sont proposées pour le cas d'une réponse continue en présence d'un événement censuré [141]. On peut alors distinguer plusieurs définitions de la sensibilité et la spécificité :

- (i) **Cumulative/Dynamique** : $Se^C(c, t) = P(X_i > c | T_i \leq t)$ et $Sp^D(c, t) = P(X_i \leq c | T_i > t)$

Cette méthode est plus appropriée pour un temps t spécifique ou un petit nombre de temps t_i différents. Chaque individu est classé en cas ou témoin par rapport à son statut au temps t . En outre, chaque individu joue le rôle de témoin pour les périodes $t < T_i$, mais contribue ensuite comme cas pour les périodes ultérieures, $t \geq T_i$

- (ii) **Incident/Statique** : $Se^I(c, t) = P(X_i > c | T_i = t)$ et $Sp^S(c, t) = P(X_i \leq c | T_i > t^*)$

Selon cette définition les individus ne changent pas de statut pathologique et ils sont traités comme 'cas'. Les individus 'cas' sont stratifiés en fonction du temps jusqu'à l'évènement et les 'témoins' sont considérés sans évènement jusqu'à un temps t fixé.

- (iii) **Incidente/Dynamique** : $Se^I(c, t) = P(X_i > c | T_i = t)$ et $Sp^D(c, t) = P(X_i \leq c | T_i > t)$

A cette approche, chaque individu peut jouer le rôle du 'témoin' pour un temps précoce $t < T_i$ et ensuite être dans la partie 'cas' pour le temps $t = T_i$.

Toutes les méthodes décrites sont utilisées pour définir la capacité de discrimination d'un marqueur sur un événement au cours du temps.

Dans le cas de la définition sensibilité incidente et spécificité statique, le groupe 'contrôle' reste inchangé, permettant d'identifier les individus à risque pendant une grande durée de temps.

La sensibilité incidente et spécificité dynamique permettent d'évaluer la spécificité au cours du temps et s'adresse à des cas où on souhaite évaluer un marqueur qui varie au cours du temps.

Dans ce cas, les groupes de 'cas et 'contrôle' changent avec chaque changement de temps, ce qui entraîne des difficultés à l'interprétation.

La définition de sensibilité cumulative et spécificité dynamique permet une évaluation de la discrimination pour une prédiction dans une fenêtre de temps et non seulement pour un temps fixe.

Pour un événement censuré, le AUC est défini par: [140]

$$AUC^{C/D}(t) = \int_0^1 ROC^{C/D}(u)du = P(X_i > X_j \vee T_i \leq t, T_j > t)$$

A partir des estimateurs de \hat{Se} et \hat{Sp} , on estime également l'aire sous la courbe ROC (AUC).

Plusieurs estimateurs sont proposés dans la littérature pour estimer la courbe ROC en fonction du temps. [142]

- L'estimateur 'naïf' : ou \hat{Se} et \hat{Sp} sont définis par la partie de vrais positifs et vrais négatifs observés dans la population excluant les individus censurés, si la censure est présente.
- Comme montré par Heagerty et al 2000, Chambless et Diao 2006,[10], [143] l'estimateur du Kaplan-Meier permet de tenir compte de la censure mais \hat{Se} et \hat{Sp} ne sont pas nécessairement monotones et inclus dans [0,1]. Par conséquent, la courbe ROC correspondante n'est pas monotone ni incluse dans [0,1] x [0,1]. Ces estimateurs sont influencés par la présence de censure, de valeurs extrêmes et sont caractérisés par manque de stabilité. Les deux auteurs proposent des approches bootstrap pour estimer la variance des estimateurs et leur intervalle de confiance.
- Les estimateurs IPCW [140], [144], [145] se basent sur une pondération de l'estimateur par la probabilité de non censure. \hat{Se} et \hat{Sp} sont monotones et bornés dans [0,1]. Ces propriétés facilitent l'interprétation de la courbe ROC. L'approche bootstrap est également proposée pour l'estimation de la variance et de l'intervalle de confiance de l'estimateur.

Nous avons choisi cette dernière pour évaluer la performance de nos prédictions.

Calibration

La courbe ROC permet d'évaluer les performances de discrimination du modèle mais n'assure pas que le modèle soit bien calibré, c'est à dire que la probabilité prédite par un modèle corresponde bien à la probabilité d'avoir l'événement.

6.1.2 Le score de Brier

Le score de Brier permet de mesurer l'exactitude des prédictions. En 1950, Brier a proposé un score dans un cadre d'évaluation de prédictions météorologiques estimé sous forme d'erreur quadratique définie par la moyenne entre la probabilité prédite et celle observée. [146]. Une généralisation de ce score, avec des résultats similaires, est proposée par Murphy et se compose en trois termes : a) une mesure de l'incertitude propre aux événements, ou états, b) une mesure de fiabilité des prévisions et c) une nouvelle mesure de la résolution des prévisions. [147]

Cet estimateur a été étendu pour les données de survie[148] avec une présence de censure à droite [11]

L'idée est de comparer à un temps donné la survenue d'un évènement observée (survie, progression) avec la prédiction de cet évènement à ce temps.

$$BS(t^*) = \frac{1}{n} [(I(T_i > t^*) - \hat{\pi}(t^*, X_i))^2]$$

Où $\hat{\pi}$ sont les probabilités estimées d'un évènement au temps t^* et I l'indicatrice des évènements observés pour $i = 1, \dots, n$.

Dans le contexte d'estimation d'un évènement avant le temps $t+s$, nous considérons pour un individu i avec φ_i est la fonction indicatrice de son statut d'évènement (survie ou progression) observé et $\hat{\pi}(t+s|X_i)$ la probabilité d'évènement prédite au temps $t+s$ sachant les individus encore à risque au temps s et X_i les covariables du modèle et n_s : les individus encore à risque. Le Brier score au temps $s+t$ est défini par :

$$BS(s, t) = \frac{1}{n_s} \sum_{i=1}^{n_s} \{\varphi_i(s+t) - \hat{\pi}_i^{(s)}(s+t|s)\}^2$$

Dans le cadre d'une analyse en survie avec présence de censure, le statut du patient au temps $s+t$ n'est pas nécessairement connu. Trois cas de censure existent :

- Le patient est encore sans évènement au temps $s+t$, i.e. $t_i > s+t$.
- Le patient a eu un évènement avant le temps $s+t$
- Le patient est censuré avant $s+t$, et par conséquent son statut au temps $t+s$ est inconnu.

Réduire l'estimation du Brier Score aux patients non censurés au temps $s+t$ induirait un biais. Afin d'éviter le biais de sélection dû à la censure avant $s+t$, Plusieurs auteurs ont proposé une pondération du score de Brier en utilisant la méthode de l'inverse de la probabilité de censure au cours du temps (IPCW 'Inverse Probability of Censoring Weighting') [11], [12], [148]. Dans ce cas, le Brier Score a la forme suivante :

$$BS(s, t) = \frac{1}{n_s} \sum_{i=1}^{n_s} \widehat{W}_i(s, t) \{ \mathbb{1}_{(s < T_i \leq s+t)} - \widehat{\pi}_i^{(s)}(s, t) \}^2$$

Dans ce cas, l'estimateur du Brier Score est pondéré par la fonction \widehat{W}_i conditionnellement du temps de censure. Cette dernière peut être estimée par l'estimateur de Kaplan Meier \widehat{G} , par une régression de Cox, une régression Aalen ou encore une régression paramétrique en utilisant la censure comme l'événement. (Gerds et al 2008). Le poids \widehat{W}_i est de la forme suivante :

$$\widehat{W}_i(s, t) = \frac{\mathbb{1}_{(T_i > t)}}{\widehat{G}(t)} + \frac{\mathbb{1}_{(T_i < t)} \Delta_i}{\widehat{G}(T_i)}$$

où $\Delta_i / \widehat{G}(T_i)$: le poids de pondération.

Il est important de noter que l'approche IPCW a l'avantage que la relation du temps jusqu'à l'événement et du vecteur des covariables n'a pas besoin d'être spécifiée et fait l'hypothèse que le temps de censure est indépendant du temps avant l'événement conditionnellement au vecteur des covariables X_i .

Le Brier score est une mesure de prédiction complète qui évalue la capacité du modèle de prédiction à la fois en termes de calibration et de discrimination. Cependant, c'est une mesure qui dépend de l'incidence cumulée de l'événement principal à cause de la présence de censure, et par conséquent son interprétation doit être faite avec précaution puisque l'échelle de la prédiction varie avec la variation du temps s [12]

Cet estimateur peut prendre des valeurs comprises entre 0 et 1 signifiant la parfaite discordance proche de 1 et la parfaite concordance proche de zero. La valeur de 0.25 est considérée comme une valeur seuil car elle correspond à ce que donnerait un modèle prédisant une probabilité d'événement de Par exemple, un risque de décès de 50% à tous les patients correspond à un Brier score de 0.25. Les valeurs du Brier score très inférieures à 0.25 indiquent une bonne prédiction. [13]

6.2 Validation interne et validation externe

En recherche clinique, il est indispensable de vérifier la validité des résultats obtenus dans une étude. Il existe deux principales méthodes pour vérifier l'exactitude des résultats : la validation interne et la validation externe. Ces méthodes permettent de vérifier des aspects différents lors de la construction d'une étude. [149], [150]

La validation interne permet de vérifier l'existence des preuves solides qui représentent la réalité et ne sont pas seulement dues au hasard. Les méthodes de « cross validation » et de « K-fold cross-validation » sont principalement utilisées pour ce type de validation. La validation croisée consiste à séparer la base existante en deux sous-échantillons aléatoirement : l'échantillon d'apprentissage, utilisé pour la construction du modèle, et l'échantillon de validation, utilisé pour évaluer les performances prédictives. [151] La séparation de l'échantillon

permet d'éviter le risque de surapprentissage que l'on pourrait avoir avec le modèle construit sur les mêmes données que les prédictions.

Le principe de la validation croisée est représenté dans la figure ci-dessous. Cette technique permet de calculer les prédictions sur les k sous-échantillons qui sont ensuite fusionnés. Ainsi, la qualité prédictive est évaluée sur l'échantillon de validation issu de k sous-échantillons.

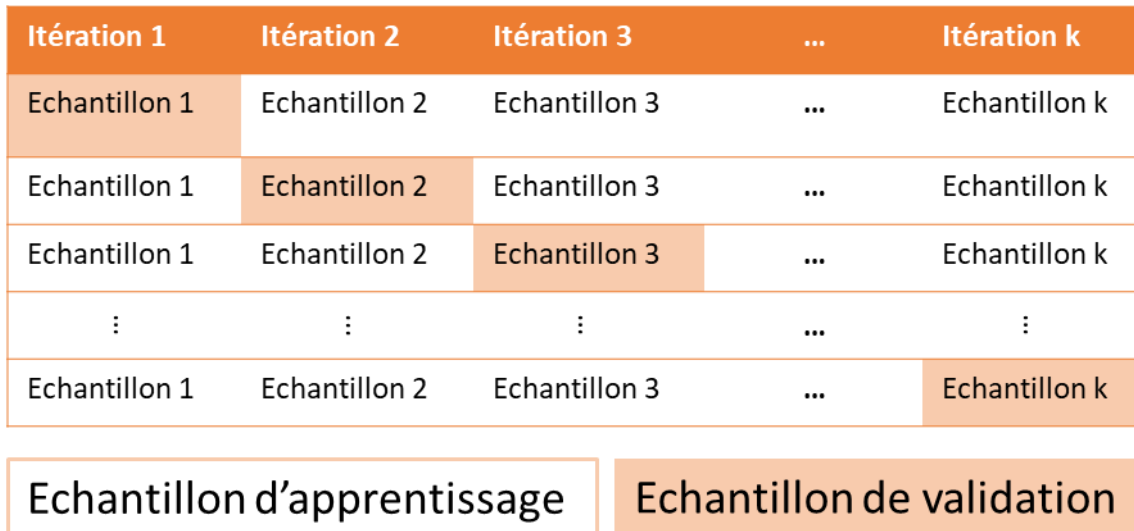


Figure 6.1: Le rééchantillonnage

Afin d'améliorer les méthodes de validation croisée, B. Efron a proposé l'approche bootstrap. Cette technique se base sur le principe que l'échantillon observé est la meilleure approximation de la distribution à priori d'une variable aléatoire [152]. En plus, elle permet de contrôler la présence de biais et de suroptimisation des prédictions. [153], [154]. Par conséquent, la variance et les intervalles de confiance des estimations sont calculés afin d'estimer la fiabilité et la précision du modèle.

Plus précisément, les mesures de dispersion et les intervalles de confiance des estimations sont calculés à partir d'un rééchantillonnage de l'échantillon observé.

La validation externe consiste à trouver une nouvelle base des données, indépendante de l'échantillon d'apprentissage, qui servira d'échantillon de validation. Idéalement cette nouvelle base des données doit avoir les mêmes caractéristiques que base de données utilisée pour développer le modèle. [155]. De plus, il est important que l'échantillon externe soit d'une taille importante afin d'avoir une puissance suffisante pour détecter les variations de la performance. [151]

La validation externe permet la généralisation de résultats dans une nouvelle population. Par exemple, dans le cadre des prédictions individuelles dynamiques et de la médecine personnalisée, il est indispensable de vérifier la validité des prédictions obtenues par les différents modèles statistiques dans une nouvelle population. Il existe plusieurs méthodes pour

prédire le risque d'un évènement mais l'évaluation de celles-ci reste un sujet délicat. La difficulté est due (a) à l'absence de données de validation indépendantes (externes), (b) à la censure présente dans les données de survie. Suite à la construction d'un modèle prédictif, il est indispensable d'évaluer sa capacité de pouvoir qualifier les nouveaux patients.

La validation externe permet de dévoiler une mauvaise performance du modèle et ensuite identifier et traiter les raisons de cette mauvaise performance. Par exemple ces raisons peuvent être dues à une mauvaise stratégie de construction du modèle, à l'existence des différences des caractéristiques parmi les deux échantillons ou encore des variations aléatoires dues aux petites tailles d'échantillon de certaines études. [75]

Dans le cas où la validation externe n'est pas toujours possible en raison du manque des données appropriées, une validation interne est recommandée. [149]

6.3 Validation dans le cadre d'une méta-analyse

Les meta-analyses permettent naturellement une validation externe. Par construction, elles disposent de plusieurs essais avec des critères d'inclusion et des caractéristiques de patients équivalents, qui visent à répondre à la même question scientifique. Néanmoins, il est important de faire attention à la taille d'échantillon de l'étude qui servira pour la validation externe. Steyerberg et al 2003 suggèrent des tailles d'échantillon importantes pour l'échantillon de validation externe pour obtenir une puissance suffisante pour détecter des changements de performance cliniquement importants par rapport à l'estimation validée en interne.[151]

Deux formes de validation croisée sont possibles dans le cadre d'une méta-analyse. La validation croisée « leave-one-study-out » et la validation croisée « leave-one-fraction-out ». Les deux méthodes présentent des différences concernant l'algorithme de fonctionnement et leur application est dépendante de la question de recherche. La première est utilisée pour étudier une association comme par exemple l'effet du traitement ou la validation d'un critère de substitution et la deuxième est principalement utilisée dans le cadre des prédictions.[156]

La validation croisée « leave-one-study-out » consiste à séparer l'échantillon total, constitué par k essais, en un échantillon d'apprentissage de k-1 essais et un échantillon de validation constitué de 1 essai. Ce type de cross validation permet une validation externe

Dans le cas leave-one-fraction-out, l'échantillon complet est séparé en k sous-échantillons ou chaque sous-échantillon dispose de patients de tous les essais. Ensuite le modèle d'apprentissage est construit dans les k-1 échantillons et l'échantillon de validation sur l'échantillon restant.

Avec la k-fold cross-validation, il y a une variabilité induite par la séparation aléatoire en k parties. Pour pallier cela, la procédure de la séparation de l'échantillon initial est répétée M fois avec un rééchantillonnage aléatoire sans remise des k sous-échantillons avec la méthode de bootstrap qui permet d'améliorer la procédure de la cross-validation.

Le bootstrap est une méthode non paramétrique qui est utilisée lorsque l'échantillon présente des particularités, par exemple : la distribution de l'échantillon est difficile à dériver même de façon approximative, l'échantillon est de petite taille ou il est constitué de données complexes (groupées et stratifiées)

Suite à ces M répétitions nous obtenons une distribution des mesures qui évaluent la qualité prédictive du modèle. La moyenne et l'intervalle correspondant au 95 % de la distribution bootstrap est ensuite présenté. Ceci permet d'évaluer avec une meilleure précision les fluctuations de la technique de validation croisée et tenir en compte les fluctuations des partitions.

Sachant que dans cette étude nous avons pris en compte l'effet essai à la construction de notre modèle, il est indispensable de le prendre en compte lors de la validation. Par conséquent, le modèle de prédiction dispose d'un paramètre qui dépend de l'essai. La question qu'on se pose est comment faire lors qu'on n'est pas dans le cadre d'un essai clinique et qu'on souhaite estimer le risque de résultats futurs pour des nouveaux individus sur la base d'un ou plusieurs caractéristiques sous-jacentes ? Comment alors peut-on estimer l'effet essai pour ces nouveaux individus ? Et surtout comment pouvoir prédire avec précision le risque pour ces nouveaux individus ? Nous sommes déjà intéressés à cette question au chapitre précédent (5.6.2)

Pour une validation externe, trois stratégies ont été proposées par Snell et al 2015. [157]

- (i) L'estimation d'un nouveau intercept à la population externe en proposant un test pour les estimations entre les deux populations afin de vérifier leur compatibilité.
- (ii) L'utilisation de la moyenne pondérée des intercepts estimées des différentes études à partir du modèle développé.
- (iii) L'utilisation de l'estimation de l'une des études du modèle développé, la plus similaire à celle de l'étude de validation externe, qui présente la même prévalence et les mêmes caractéristiques par rapport aux références représentatifs actuels de la maladie étudiée

En se basant sur cette dernière, Steyerberg et al 2008, ont développé leur modèle stratifié par étude. Cependant, dans leur cohorte de validation, ils ont choisi l'effet essai lié à un seul essai qui semblait correspondre à une représentation des proportions typiques de mortalité et de résultat défavorable. [158]

Il est important de noter que dans le cadre d'une validation externe pour un modèle de prédiction, plusieurs mesures statistiques de performance sont demandées. Plus particulièrement, il est conseillé d'utiliser au moins une mesure de calibration et une mesure

de discrimination. En effet, il est important de vérifier la cohérence entre les risques observés et ceux prédits ainsi que l'existence de corrélation inter-essais et intra-essais.

En deuxième partie de ce travail nous allons présenter l'application de ces méthodes aux données de la méta-analyse des données individuelles 'MAOV'. Au chapitre qui suit, nous allons évaluer la trajectoire du CA-125 les premiers mois après l'initiation du traitement de première ligne pour prédire la survie des patientes atteintes d'un cancer de l'ovaire avancé.

7. EVALUATION DE LA TRAJECTOIRE DU BIOMARQUEUR CA-125 LES PREMIERS MOIS APRES L'INITIATION AU TRAITEMENT POUR PREDIRE LA SURVIE GLOBALE DES FEMMES ATTEINTES D'UN CANCER DE L'OVAIRE AVANCE NOUVELLEMENT DIAGNOSTIQUE, SUR LA BASE DE DONNEES DE META-ANALYSE.

7.1 Introduction

Le cancer de l'ovaire (CO) est la septième cause de mortalité par cancer chez les femmes dans le monde, avec un taux de survie de 46 %, 5 ans après le diagnostic [159]. Le traitement de première intention consiste en une chirurgie de réduction primaire ou intermittente et une chimiothérapie à base de platine et de taxane, qui peuvent être associées à des traitements d'entretien comprenant du bevacizumab ou inhibiteurs de la poly (ADP-ribose) polymérase (PARP), le cas échéant. Récemment, les inhibiteurs de PARP ont démontré des gains très prometteurs en matière de survie sans progression et de survie globale, en particulier pour les patientes atteintes d'un cancer de l'ovaire muté BRCA [57], [59], [160], [161]. L'antigène CA-125 est un marqueur principalement lié au cancer de l'ovaire, mais il peut également être élevé dans d'autres pathologies. Selon les critères du 'Gynecologic Cancer Intergroupe' (GCIg), ce marker est utilisé pour suivre la réponse au traitement en cas de récurrence de la maladie ou pour définir l'évolution après un traitement de première intention [69], [162]. Malgré le fait qu'en 2010 l'utilisation systématique du CA-125 a été découragée [163] deux stratégies de suivi des femmes après le traitement primaire sont désormais utilisées : (i) un suivi clinique avec le CA-125 et une imagerie sur indication clinique ou (ii) des imageries tout au long du suivi et des tests facultatifs du CA-125 [66]. Cependant, la contribution du CA-125 est débattue, car les essais cliniques mettent en œuvre des surveillances avec CA-125 et CT-scan, en particulier dans les contextes de traitement de maintenance. Les meilleures façons d'analyser son délai jusqu'à la décroissance du CA-125, sa reproductibilité entre les essais, ainsi que le risque d'erreur, ne sont pas suffisamment documentées.

L'évaluation de la valeur pronostique de l'évolution précoce du CA-125 dans le cas où on s'intéresse au délai jusqu'à l'évènement d'intérêt, soulève des défis statistiques. La trajectoire du CA-125 peut être résumée de plusieurs manières (par exemple, la valeur du CA-125 au départ, KELIM à 3 mois, valeur du CA-125 ou diminution au temps de prédiction (par exemple 3 et 6 mois)), et son rôle pronostique a été étudié dans de nombreuses études isolées [5]–[7], [69], [164]. Néanmoins, les performances de ces méthodes n'ont jamais été évaluées dans une méta-analyse des données individuelles des patients qui constitue le plus haut niveau de preuve. Le CA-125, comme tout biomarqueur, peut être sujet à des erreurs de mesure et de variation biologique (c'est-à-dire que certaines valeurs peuvent sembler supérieures ou

inférieures à celles prévues par la trajectoire globale), de sorte que les résumés bruts du CA-125 observé peuvent être sous-optimaux pour évaluer la valeur pronostique du CA-125. De plus, une méthode d'imputation peut être nécessaire en cas des données manquantes du CA-125 pour un instant donné [165]. Par conséquent, certains auteurs ont proposé de s'appuyer sur des techniques de modélisation qui gèrent des biomarqueurs avec du bruit et des mesures manquantes pour mieux évaluer l'évolution du CA-125 [84], [86]

Récemment, l'utilisation de la constante K du taux d'élimination du CA-125 (KELIM) [87], basée sur des mesures longitudinales du CA-125 ainsi que sur des paramètres pharmacocinétiques et pharmacodynamiques du traitement, a suscité un intérêt important dans de multiples contextes (dans les maladies récurrentes, pour les traitements néoadjuvants et pour les traitements initiaux). En milieu adjuvant et néoadjuvant, le rôle pronostique du KELIM à 100 jours (soit 3 mois) a été testé dans plusieurs essais [92], [166]. Alternativement, des techniques statistiques telles que la modélisation à effets mixtes a été développée avec succès dans le cas du cancer de la prostate pour modéliser la dynamique de l'antigène PSA (Prostatic Specific Antigen) au fil du temps [10], [84]. Une étude récente a proposé une estimation du taux de CA-125 facile à quantifier afin d'aider à la prise de décision concernant le traitement de deuxième intention pour les patientes atteintes d'un cancer sévère de l'ovaire récurrent de haut grade [11]. Ces statistiques les méthodes de données longitudinales sont ainsi au cœur de l'évaluation de la valeur pronostique de la trajectoire estimée sur différentes périodes. Considérer des délais de moins de trois mois pourrait être difficile en raison de la rareté des mesures du CA-125 au cours de cette période. À l'inverse, des périodes de surveillance plus longues du CA-125, par exemple six mois, peuvent améliorer la précision du modèle et la valeur discriminatoire de la décroissance du CA-125. Un tel délai de surveillance peut entraîner une charge de collecte du CA-125, et par conséquent un nombre élevé de visites de suivi.

Lorsqu'on s'intéresse à prédire le délai jusqu'à un événement basé sur un biomarqueur, il est important de tenir en compte la nature des données et leurs particularités. Par exemple, le biomarqueur peut être sujet aux erreurs de mesures, le délai diffère entre le processus longitudinal (quelques mois) et les paramètres de survie (quelques années). L'approche de Landmark est une approche appropriée qui considère ces particularités. En bref, son principe est de fixer un temps repère-appelé temps de landmark- à partir duquel la prédiction doit être effectuée et de restreindre la population à des sujets sans événement à ce moment-là. Les résumés de la dynamique du biomarqueur jusqu'au temps de landmark servent à prédire la progression ou la mort dans un temps futur, après le temps de landmark [12], [167].

Lors du développement d'outils de prédiction, l'évaluation des performances prédictives doit être effectuée avec soin. La capacité du marqueur à prédire le risque de décès peut être mesurée à l'aide du score de Brier et de l'aire sous la courbe ROC (AUC). Ces estimateurs sont adaptés au contexte de d'analyse du temps jusqu'à l'évènement. [14]

Les performances prédictives peuvent être trop optimistes lorsqu'elles sont évaluées sur les mêmes données utilisées pour entraîner le modèle. Pour corriger ce biais, des techniques de validation externe ou de validation croisée peuvent être utilisées.

Ce travail vise à évaluer les performances de la dynamique précoce du CA-125 et à déterminer le temps de landmark le plus approprié (c'est-à-dire le délai requis pour enrichir les modèles de prédiction statistique) afin de prédire la survie globale au meilleur moment futur chez les patientes nouvellement diagnostiquées avec un cancer de l'ovaire avancé et traité par chimiothérapie à base de taxane et de platine. Nous avons utilisé la méta-analyse des données individuelles des patients MAOV [134] (IPD) de l'étude de groupe de méta-analyse GCIG, qui comporte un échantillon large et diversifié, permettant une validation croisée des résultats entre patients et études ainsi qu'entre sous-groupes en fonction des caractéristiques des patients.

7.2 Méthodes

Ce rapport suit les directives 'Preferred Reporting Items for Systematic Reviews and Meta-analyses' (PRISMA)-IPD et la déclaration 'Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis' (TRIPOD) pour l'enregistrement du protocole, l'identification de l'essai, la collecte des données, l'intégrité des données, l'évaluation des biais et les analyses de sensibilité [76], [121], [126], [168]

Cette méta-analyse a été enregistrée auprès de PROSPERO (CRD42017068135). Le comité d'éthique du Centre de lutte contre le cancer Gustave Roussy, Villejuif, France, a approuvé cette étude. Ainsi, la Commission nationale de l'informatique et des libertés a levé l'obligation de consentement éclairé en raison de l'utilisation de données dépersonnalisées.

7.2.1 Population d'analyse

Pour développer et valider un modèle statistique prédictif de la dynamique du CA-125 en relation avec la survie globale, nous avons utilisé les données de la méta-analyse du GCIG sur données individuelles MAOV. Celle-ci incluait des patientes atteintes d'un cancer de l'ovaire nouvellement diagnostiqué aux stades FIGO IC à IV, dont les données avaient été collectées dans des essais cliniques randomisés publiés entre janvier 2001 et septembre 2016 (méta-analyse du GCIG sur le cancer de l'ovaire [134]. Parmi l'ensemble initial de 17 essais, nous avons sélectionné ceux qui étudiaient les traitements systémiques initiaux après chirurgie (pas de traitements de maintenance).

Pour la présente analyse, seuls les essais qui ont recueilli les taux sériques du biomarqueur CA-125 au départ (avant le début du traitement) et à plusieurs reprises au cours du suivi ont été retenus. La collecte des données a suivi le protocole de recherche original. Pour être éligible, chaque patiente devait avoir au moins deux mesures du CA-125, dont une au départ, et des informations complètes concernant la survie globale et la progression.

Les données ont été vérifiées pour s'assurer leur qualité. Les niveaux du biomarqueur supérieurs à 15 000 ont été considérés comme des valeurs aberrantes, et les mesures peropératoires du CA-125 ont été exclues, car elles sont affectées par l'intervention chirurgicale.

Les valeurs du biomarqueur ont été normalisées par une transformation logarithmique afin de corriger la symétrie de la distribution des mesures du CA-125 et respecter l'hypothèse de normalité du modèle linéaire.

7.2.2 Critère principal

Le principal critère d'évaluation était la survie globale, définie comme le temps écoulé entre la randomisation et le décès, quelle qu'en soit la cause. Les patients en vie à la date limite ont été censurés à la dernière date de suivi

La valeur pronostique des niveaux de CA-125 à 3 et 6 mois à partir de la randomisation et le taux de diminution du CA-125 ont été évalués. Le modèle a été développé sur la base de patients traités soit par la « chimiothérapie standard » (Paclitaxel + Carboplatine et Epirubicine/Doxorubicine), soit par la même chimiothérapie avec un traitement expérimental, étant donné qu'aucun des essais inclus n'a montré de différences significatives dans l'effet du traitement entre les deux bras. Une analyse de sensibilité a été réalisée en excluant les patients des groupes expérimentaux. Les caractéristiques suivantes des patients étaient disponibles au départ : âge, état de performance, stade FIGO, sous-type et grade histologique, et maladie résiduelle après la chirurgie (≥ 1 cm, < 1 cm, inconnu).

7.2.3 Méthodes statistiques

Afin d'étudier l'association biomarqueur & la survie globale dans le temps, une analyse de landmark a été réalisée, en prenant en compte différents délais, notés s , après la randomisation. [92]. Toutes les données disponibles jusqu'à s ont été modélisées pour prédire le risque de décès entre s et $s + t$, t représentant l'horizon de prédiction.

Les facteurs pronostiques étudiés, basés sur la trajectoire précoce du CA-125, comprenaient (i) la valeur observée du CA-125 au temps de landmark, où, si cette valeur était manquante, la mesure antérieure la plus proche au temps de landmark était retenue (OVLT), et (ii) le déclin relatif observé (ORD) du biomarqueur entre la randomisation et la dernière mesure du CA-125 avant le temps de landmark qui est décrit par l'équation suivante :

$$\text{Vitesse de décroissance} = \frac{(ca125_{t_l} - ca125_{t_0})}{t_l - t_0}$$

t_0 : le temps à la randomisation, t_l : le temps de landmark, $ca125_{t_0}$: la valeur du CA-125 à la randomisation, $ca125_{t_l}$: la valeur du CA-125 au temps de landmark

En outre, trois résumés du CA-125 estimés à l'aide d'un modèle linéaire mixte hiérarchique ont été pris en compte : (iii) la valeur estimée au temps de landmark (EVLT) et (iv-v) la valeur estimée du CA-125 au temps de landmark (EVLT) et (iv-v) la pente estimée à la randomisation (ESB) et au temps de landmark (ESLT).

7.2.4 Modèle linéaire mixte hiérarchique

Pour ces trois résumés, les mesures répétées du CA-125 log-transformé ont été analysées à l'aide d'un modèle linéaire hiérarchique mixte (HLMM). Ce dernier traite la structure hiérarchique des données, l'erreur de mesure inhérente et les valeurs manquantes selon le mécanisme d'absence au hasard. Trois niveaux hiérarchiques ont été distingués : (i) la variabilité entre les essais, en utilisant des effets aléatoires au niveau de l'essai, (ii) la corrélation à l'intérieur du patient, en utilisant des effets aléatoires spécifiques au patient, en introduisant par conséquent une structure de corrélation entre les mesures répétées, et (iii) la variabilité due aux erreurs de mesure spécifiques à l'observation.

Nous supposons k le nombre d'essais disponibles dans une méta-analyse et N le nombre total d'individus. Pour chaque individu, les mesures du biomarqueur CA-125 au cours du temps correspondent au vecteur de réponses Y_{ik} , ou un patient i de l'essai k dispose n_i mesures. Le vecteur u_k correspond à l'effet aléatoire associée à l'essai. Les vecteurs X_{ik}^T et Z_{ik} sont des covariables dépendantes du temps, associées au vecteur des effets fixes β , au vecteur des effets aléatoires b , respectivement. [74]

$$Y_{ik}(t) = X_{ik}^L(t)^T \beta + Z_{ik}(t)^T b_{ik} + 1_{n_i} u_k + \varepsilon_{ik}(t)$$

Le terme ε_{ik} correspond à l'erreur et suit une distribution normale $N \sim (0, \sigma^2 I_{n_i})$, $b_{ik} \sim N_q(0, D)$ et $u_k \sim N_q(0, \sigma_u^2)$. Les effets aléatoires et le terme d'erreur sont indépendants. Pour le patient i de l'essai k , $X_{i,k}$ est de dimension $n_i \times p$ et le $Z_{i,k}$ est de dimension $n_i \times q$.

Trois niveaux sont définis au modèle linéaire mixte hiérarchique. Premièrement, l'intercept aléatoire u_k , spécifique à l'essai et correspondant à l'effet moyenne de chaque essai, ceci est la caractéristique spécifique du cadre méta-analytique. Deuxièmement, le vecteur $X_{ik}^T \beta$ correspond à l'effet moyen du CA-125. Troisièmement, le vecteur aléatoire $Z_{ik} b$ correspond à la fluctuation individuelle. Le vecteur des réponses y peut être continue ou discret, dans notre cas ce sont les données observées continues du biomarqueur.

Les données longitudinales du biomarqueur CA-125 ont été résumées par le modèle linéaire mixte décrit au chapitre 3.3 ont été également estimées. Plus précisément, ces résumés ont été décrits au chapitre 5.6.1 et il s'agit des résumés suivants : (i) la valeur du CA-125 estimée au temps de landmark (EVL), (ii) la pente du CA-125 estimée à la randomisation (ESB) et (iii) la pente du CA-125 estimée au temps de landmark (ESLT)

Afin d'obtenir les prédictions individuelles dynamiques, il est pertinent d'estimer les effets aléatoires. L'estimateur 'Best Linear Unbiased Predictor' (BLUP) estimé par l'espérance des effets aléatoires conditionnellement aux observations ; on le calcule comme suit à partir de la distribution marginale (*), vu dans le chapitre 5.4.2:

$$\hat{d}_{ik} = E(d_{ik} | y_{ik}) = B Z_k^T V_{ik}^{-1} (y_{ik} - X_{ik} \hat{\beta})$$

Il existe plusieurs packages R qui permettent une construction d'un modèle linéaire mixte comme par exemple `lcm`, `nlme`, `lme4`. Pour notre analyse, nous avons utilisé la fonction `lmer` du package `lme4` puisque celle-ci permet l'ajout de plusieurs effets aléatoires emboîtés, indispensable pour prendre en compte des effets aléatoires à l'essai ainsi qu'au niveau de l'individu.

7.2.5 Modélisation de la trajectoire avec des splines

Pour tenir en compte la dynamique non linéaire du log-CA-125 dans le temps et à la vue de nos données, nous avons testé plusieurs fonctions multivariées (F) qui nous semblaient pouvoir traduire de plus près la trajectoire du biomarqueur. Nous avons testé plusieurs fonctions du temps: $temps$, $temps^2$, $\frac{1}{temps}$, $\log(temps)$ et des splines.

Finalement, nous avons choisi une base de splines cubiques naturelles pour la modélisation du temps avec 3 ou 4 nœuds internes pour le temps repère à 3 ou 6 mois, respectivement [76], [169]. Les nœuds ont été placés au début du suivi (c'est-à-dire à 0,5, 1 et 2 ou 3 mois pour 3 et 4 nœuds, respectivement) afin de saisir d'éventuelles décroissances rapides du biomarqueur. Les nœuds externes des splines ont été placés aux percentiles 2 % et 98 % du temps. La sélection des nœuds était basée sur la minimisation du critère d'Akaike (AIC) du modèle. Pour chaque patient, nous avons inclus des effets aléatoires sur chaque fonction spline cubique naturelle afin de capturer l'écart du CA-125 spécifique au patient par rapport à la trajectoire moyenne. Pour chaque étude, un intercept aléatoire a également été utilisé pour capturer l'écart spécifique à l'essai.

Pour générer les splines dans notre analyse, nous avons utilisé le logiciel R et le package 'splines' qui permet de générer des B-splines. Pour générer des splines cubiques naturelles, la fonction 'ns()' a été utilisée. Cette fonction génère une matrice de base pour les splines. La fonction 'nsxD()' permet d'avoir la dérivée première d'une base de spline cubique naturelle.

7.2.6 Évaluation de la performance prédictive des résumés du CA-125

Les résumés ont été inclus, un-par-un, en tant qu'effets fixes dans un modèle de Cox à risques proportionnels, prenant en compte un risque de base spécifique à l'étude. La méthodologie de ces analyses a été détaillée en section 4.4.1

Plus précisément, nous avons étudié l'effet de chaque résumé de la trajectoire du CA-125, représenté par le vecteur Γ_{ik} . Nous avons aussi ajouté les caractéristiques des patientes représentées par le vecteur X_i . Par exemple pour un individu $i = (1, \dots, n_k)$ de l'essai k , le risque instantané est exprimé par la fonction suivante :

$$\lambda_i(t/\Gamma_{ik}, X_i; U_k) = \lambda_0(t, k) \exp(\gamma \Gamma_{ik} + \vartheta X_i)$$

Ou $\lambda_0(t, k)$ représente le risque de base instantané au temps t pour l'essai k

γ et ϑ : les coefficients des résumés du CA-125 et des caractéristiques des patientes respectivement.

L'hétérogénéité entre les études de l'association entre les résumés et la survie a été testée à l'aide d'un test de rapport de vraisemblance qui comparait la vraisemblance partielle du modèle stratifié à la vraisemblance partielle du modèle stratifié utilisant des effets de résumé spécifiques à l'essai [168], [170]

Comme le montre la figure 8, la probabilité de décès a été prédite à partir du temps de landmark s à l'horizon de prédiction s + t selon le résumé CA-125 calculé au moment s, avec s = 3 et 6 mois et t = 24, 36, 48, 60 dans un sous-ensemble d'études avec un suivi suffisant.

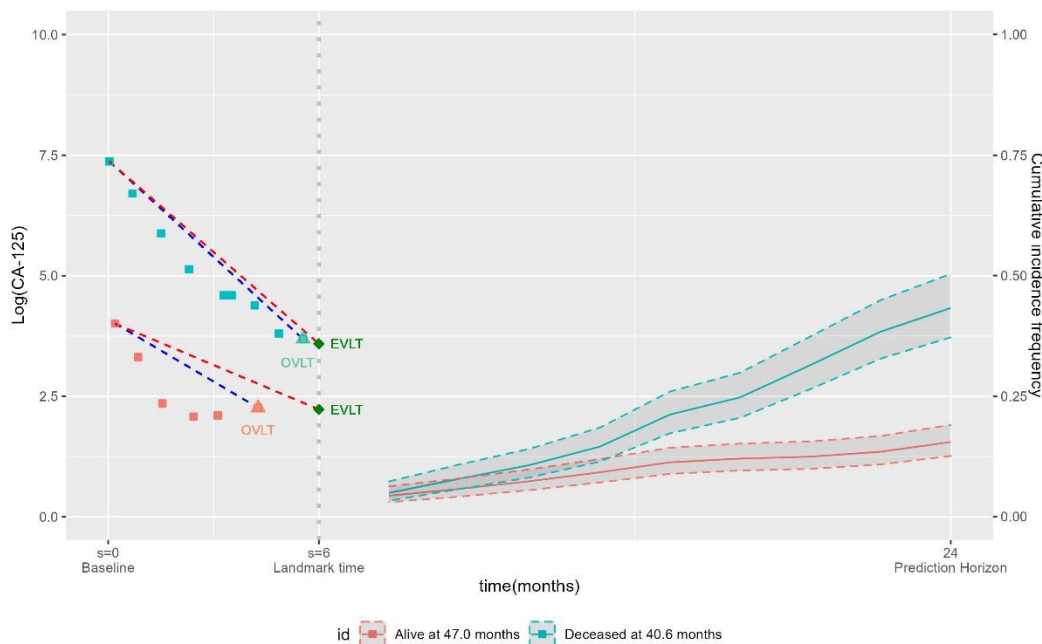


Figure 7.1. Illustration du cadre de Landmark. Trajectoires des valeurs observées du CA-125 jusqu'au temps de Landmark s = 6 mois pour deux patients présentant un risque de décès au temps de Landmark (n° 1 en vie après 47 mois et n° 2 décédée après 40.6 mois). Leur survie globale a été évaluée à t = 24 mois.

Des résumés de la trajectoire précoce du CA-125 ont été obtenus à partir des valeurs jusqu'au temps de Landmark (les lignes pointillées bleues représentent la pente observée, et les lignes pointillées rouges représentent la pente estimée).

La performance prédictive de ces prédictions a été évaluée à l'aide de l'AUC et du score de Brier afin de déterminer le temps de landmark le plus approprié et d'explorer la robustesse pour différents horizons de prédiction. Pour tenir en compte des temps censurés entre le temps

s et le temps $s + t$, les estimateurs de ces deux quantités ont été pondérés par l'inverse de la probabilité de censure (IPCW) [140]. Les deux mesures ont été utilisées pour évaluer la qualité de l'outil de prédiction. L'AUC peut être considérée comme une mesure de concordance entre les patients présentant un risque élevé ou faible de décès[169], tandis que le score de Brier mesure la calibration et la discrimination. Les deux mesures sont comprises entre 0 et 1. Un AUC inférieur à 0,7 indique une performance discriminatoire modérée.

La calibration a été évaluée graphiquement : les patients ont été répartis en groupes définis par les quantiles des probabilités d'événements prédits, et les résultats ont été tracés en fonction du risque de décès observé.

Pour corriger les performances trop optimistes obtenues sur les ensembles de données d'apprentissage, nous avons appliqué une validation croisée à 10 blocs au niveau du patient. L'ensemble de données a été divisé en 10 sous-échantillons de même taille. Pour chaque sous-échantillon, les prédictions ont été calculées en utilisant le modèle formé sur les 9 sous-échantillons restants. Les prédictions des 10 sous-échantillons ont ensuite été regroupées pour l'évaluation des performances. La moyenne et l'écart-type du score de Brier et l'AUC sur 50 répétitions de la technique de validation croisée ont finalement été rapportés pour tenir en compte les fluctuations dans les partitions à l'évaluation de la performance.

Pour étudier la valeur ajoutée d'un résumé de la dynamique du CA-125, nous avons comparé le modèle de Cox stratifié par étude et ajusté pour le résumé évalué soit (i) au modèle nul, c'est-à-dire à un modèle de survie stratifié par étude qui a servi de référence, soit (ii) au même modèle de Cox avec le résumé de la dynamique et avec ou sans le taux de CA-125 au début de l'étude.

Dans une analyse secondaire, la valeur pronostique du CA-125 a été évaluée en fonction de sous-groupes de patientes définis par leurs caractéristiques de base, telles que le stade FIGO et la maladie résiduelle.

Enfin, pour le seuil connu de 35 pour la valeur du CA-125, nous avons évalué la valeur prédictive positive (VPP) et la valeur prédictive négative (VPN) aux horizons temporels analysés.

Les analyses statistiques ont été réalisées à l'aide de R (version 4.1.2, R foundation for Statistical Computing, Vienne, Autriche), avec le package R RiskRegression pour l'évaluation des performances prédictives et le package lme4 pour le HLMM.

7.3 Les résultats

7.3.1 Sélection et caractéristiques des études

Parmi les 17 essais de la méta-analyse du GCIg, des mesures répétées du CA-125 étaient disponibles pour 13 essais. Quatre essais portant sur des traitements de maintenance ont été exclus. Neuf essais au total ont ensuite été sélectionnés, dont quatre et cinq essais, respectivement, portaient sur de nouveaux traitements initiaux et sur l'intensification d'un

traitement initial, ce qui a permis de constituer un échantillon total de 5573 patients. (Figure 7.2)

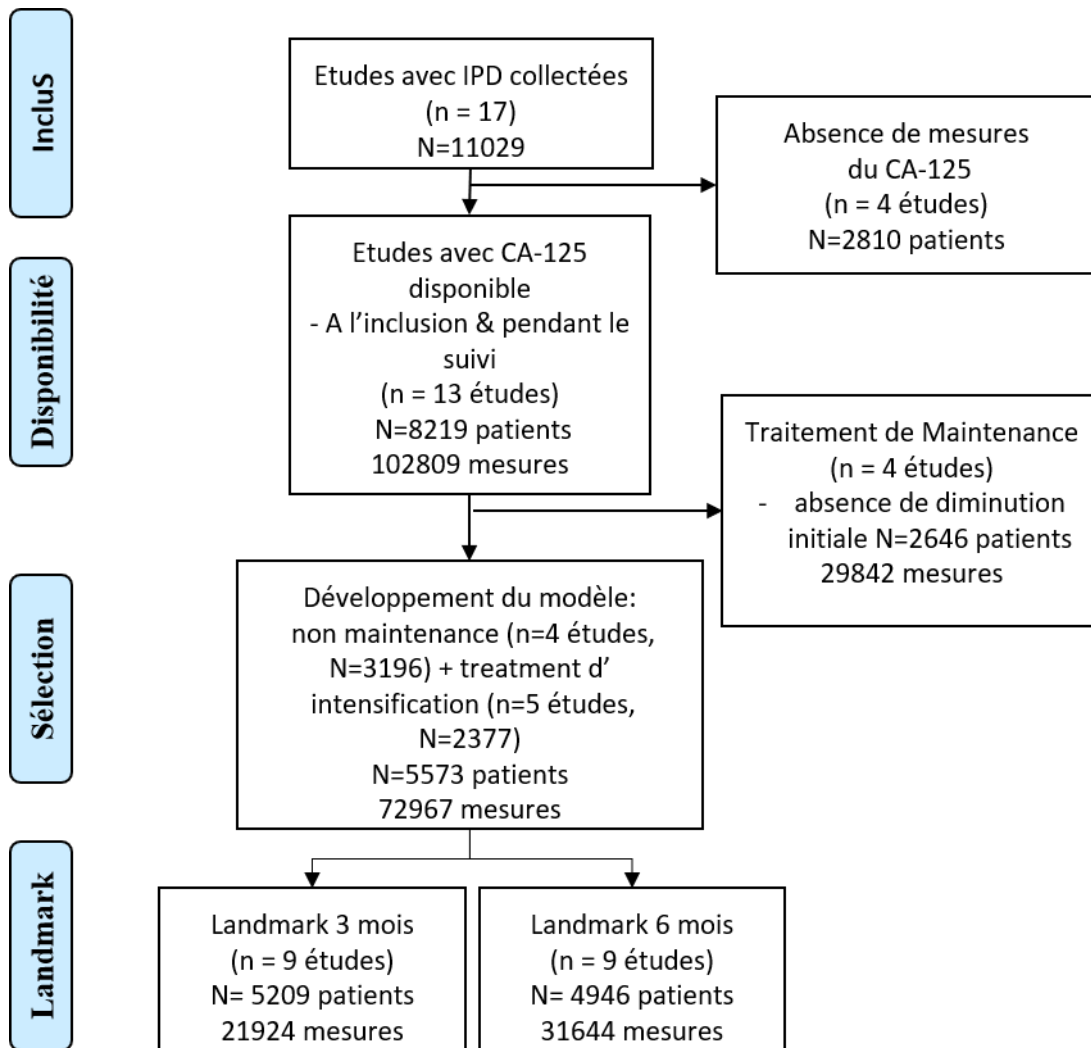


Figure 7.2 : Sélection des études

Tableau 7.1 : Description des études

	Essai	Nb de pts	# CA-125 mesures	Médiane de Survie
Treatment d'intensification	Total	2377	21978	3.87 (3.6 ; 4.11)
	MITO-7	529	1429	NA
	GINECO-2007	109	634	2.63 (2.25 ; 3.39)
	TURBO-2014	253	3421	3.74 (3.23 ; 4.30)
	JGOG-3016	557	3475	6.72 (5.30 ; NA)
	SCOTROC-4	929	14019	2.75 (2.54 ; 2.97)
Sans maintenance	total	3196	49989	4.33 (4.08 ; 4.66)
	JGOG-3017	632	9403	NA
	NSGO-2012	774	9904	3.60 (3.21 ; 3.98)
	SCOTROC-1	1052	12582	2.86 (2.73 ; NA)
	CCTG-OV.16	737	18100	3.84 (3.49 ; 4.28)
Total		5573	72967	4.11 (3.94 ; 4.33)

Abréviations : NA, non atteint.

Tableau 7.2: Caractéristiques des patients par essai (N=5209)

Essai	FIGO Status					Maladie Résiduelle			
	I	II	III	IV	.	<1cm	>= 1cm	NA	.
CCTG-OV.16	0	65(9.4%)	468(67.4%)	161(23.2%)		325(57%)	238(41.8%)	7(1.2%)	124
JGOG-3016	0	102(18.6%)	364(66.5%)	81(14.8%)		258(47.2%)	275(50.3%)	14(2.6%)	
JGOG-3017	395(66.1%)	69(11.5%)	112(18.7%)	22(3.7%)		548(91.6%)	48(8%)	2(0.3%)	
GINECO-2007	0	0	78(80.4%)	19(19.6%)	1	14(14.4%)	83 (85.6%)		1
MITO-7	41(8.9%)	42(9.1%)	273(59.2%)	105(22.8%)		NA	NA		
NSGO-2012	0	94(12.7%)	533(71.8%)	115(15.5%)		317(42.7%)	425(57.3%)		
SCOTROC-1	70(7.2%)	127(13%)	645(65.9%)	137(14%)		332(33.9%)	647(66.1%)		
SCOTROC-4	118(13.9%)	75(8.8%)	549(64.6%)	108(12.7%)		300(35.3%)	550(64.7%)		
TURBO	0	19(7.9%)	156(65%)	65(27.1%)		173(72.7%)	63(26.5%)	2(0.8%)	2

Tableau 7.3: Description de caractéristiques des patientes N=5573

Variable		Total (9 essais)		Population en vie à 3 mois		Population en vie à 6 mois	
Age	n	5573		5209		4946	
	Moyenne (sd)	58.4 (10.8)		58.3 (10.8)		58.2 (10.6)	
	Médiane (Q1 ; Q3)	59 (51 ; 66)		58.3 (51 ; 66)		58.1 (51 ; 66)	
	Min ; Max	18 ; 88		18 ; 88		21 ; 88	
Performance status	n (m.d.)	5568 (5)		5204 (5)		4941 (5)	
	0	2740	49.2%	2597	49.9%	2523	51.1%
	1	2376	42.7%	2196	42.2%	2063	41.7%
	2+	452	8.1%	411	7.9%	355	7.2%
Figo	n (m.d.)	5572(1)		5208(1)		4945(1)	
	Stage I	657	11.8%	624	12%	625	12.6%
	Stage II	610	10.9%	593	11.4%	581	11.7%
	Stage III	3401	61%	3178	61%	2993	60.5%
	Stage IIV	904	16.2%	813	15.6%	746	15.1%
Type Histologique	n (m.d.)	5568 (5)		5204 (5)		4941 (5)	
	Séreux	2778	49.9%	2641	50.7%	2546	51.5%
	Endométrioïde	504	9.1%	475	9.1%	461	9.3%
	Mucineux	154	2.8%	127	2.4%	113	2.3%
	Cellules Claires	881	15.8%	822	15.8%	787	15.9%

	Autre	1111	20%	1009	19.4%	917	18.6%
	Inconnu	140	2.5%	130	2.5%	117	2.4%
Grade histologique	n (m.d.)	4275 (1298)		4019 (1190)		3813 (1133)	
	Bien différencié	451	10.5%	426	10.6%	408	10.7%
	Modérément différencié	1112	26%	1051	26.2%	991	26%
	Faible/Non différencié	2712	63.4%	2542	63.2%	2414	63.3%
Procédure chirurgicale	n (m.d.)	5020(553)		4727 (482)		4450 (496)	
	Résection primaire complète	4366	87%	4111	87%	3885	87.3%
	Chirurgie retardée	161	3.2%	144	3.0%	124	4.8%
	Sans chirurgie	493	9.8%	472	10%	441	9.9%
Maladie résiduelle maximale	n (m.d.)	4909(664)		4621 (588)		4350 (596)	
	<1cm	2356	48%	2267	49.1%	2209	50.8%
	>=1cm	2526	51.5%	2329	50.4%	2119	48.7%
	Inconnu	27	0.6%	25	0.5%	22	0.5%
Évaluation de la progression	n	5573		5209		4946	
	CT-scan	632	11.3%	598	11.5%	598	12.1%
	CA-125-Clinique (confirmation avec scanner	3712	66.6%	3483	66.9%	3247	65.6%
	Les deux	1229	22.1%	1128	21.7%	1101	22.3%

CA-125 à l'initiation	n	5573	5209	4946
	Moyenne(sd)	687.50 (1552.73)	683.05 (1542.01)	653.64 (1481.97)
	Médiane (Q1 ; Q3)	144 (44 ; 539.24)	142 (43 ; 532)	136 (41 ; 501)
	Min ; Max	0.82 ; 14400	0.82 ; 14400	0.82 ; 14400
log(CA-125 +1) à l'initiation	n	5573	5209	4946
	Moyenne (sd)	5.12 (1.69)	5.11 (1.69)	5.06 (1.69)
	Médiane (Q1 ; Q3)	4.98 (3.81 ; 6.29)	4.96 (3.78 ; 6.28)	4.92 (3.74 ; 6.22)
	Min ; Max	0.60 ; 9.58	0.60 ; 9.58	0.60 ; 9.58
nombre de mesures répétées du CA-125 par patient	n	5573	5209	4946
	Moyenne (sd)	13.09	13.58	13.74
	Médiane (Q1 ; Q3)	11 (7 ; 17)	12 (7 ; 17)	12 (7 ; 18)
	Min ; Max	2 ; 100	2 ; 80	2 ; 80

Les essais et les caractéristiques des patients sont décrits dans les tableaux 7.1 et 7.2, respectivement, et dans le tableau 7.3.

Plus de 72.967 mesures du CA-125 ont été recueillies (voir l'organigramme de la figure 7.2). Le nombre médian de mesures du CA-125 par patient était de 11 (IQR [7- 17]) (Tableau 7.3), avec une fourchette comprise entre 2 et 100 valeurs, comme décrit dans le tableau 7.3. Au total, 2 570 décès ont été signalés et la durée médiane de survie était de 4,11 ans [3,94-4,33]

7.3.2 Critère principal et cadre temporel de Landmark

Nous avons calculé la survie aux horizons de 24, 36, 48 et 60 mois pour deux temps de Landmark, s = 3 et 6 mois. Au total, 5209 et 4946 patients étaient en vie à ces deux dates de Landmark, respectivement, et ont été inclus dans l'ensemble de l'analyse. (Tableau 7.3)

Performances des résumés du CA-125

Les performances validées de manière croisée des différents résumés du CA-125 sont présentées dans les figures 7.3 et 7.4, et les résultats des AUCs et Brier score sont décrits dans la table 7.4 et l'image 7.3 ci-dessus.

7.3.3. Temps de Landmark à 3 mois

La valeur estimée à 3 mois et la dernière valeur observée avant ou à 3 mois ont fourni les meilleures performances prédictives à tous les horizons avec, par exemple, AUC de 0,745 IC 95% [0,699, 0,796] et 0,749 [0,702, 0,796], respectivement, et une hétérogénéité statistiquement significative entre les essais pour les deux résumés. Ces AUC peuvent être comparées à l'AUC du modèle nul (qui ne tient compte que de l'effet de l'essai), qui était de 0,648 [0,593 ; 0,702] à l'horizon de 24 mois (figure 7.3, tableau 7.4).

La pente estimée a montré une précision prédictive modeste. Par exemple, à 24 mois, la pente initiale estimée, sans ajouter la valeur du CA-125 au départ, a fourni un AUC de 0,685 [0,637 ; 0,739], avec une hétérogénéité significative entre les essais ($p = 0,0009$).

La pente estimée et l'ORD à 3 mois ont montré des AUC de 0,668 [0,616 ; 0,722] (avec une hétérogénéité ($p = 0,005$)) et 0,646 [0,590, 0,697] (avec une hétérogénéité non significative entre les essais ($p = 0,102$)), respectivement.

À l'horizon de 24 mois, l'impact pronostique du CA-125 au départ (avec une seule variable incluse dans le modèle stratifié par essai) était de 0,703 [0,659, 0,752]. Après l'ajout de la valeur du CA-125 au départ, les performances prédictives des valeurs estimées et observées du CA-125 au moment du repère sont restées pratiquement les mêmes que celles sans la valeur du CA-125 au départ (0,745 [0,701, 0,794] et 0,748 [0,705, 0,796], respectivement).

Cependant, la pente et l'ORD estimés à 3 mois, ainsi que le CA-125 au départ, ont montré des performances prédictives améliorées à tout moment avec, par exemple, des AUC de 0,703 [0,658-0,752] et 0,738 [0,693, 0,789] à 24 mois.

Il est important de noter que l'AUC de 0,75 indique que pour deux patientes choisies au hasard, il existe une probabilité de 75 % que la patiente présentant la valeur de CA-125 la plus basse à 3 mois ait la survie la plus longue lorsque l'on se limite à la fenêtre comprise entre 3 et 24 mois.

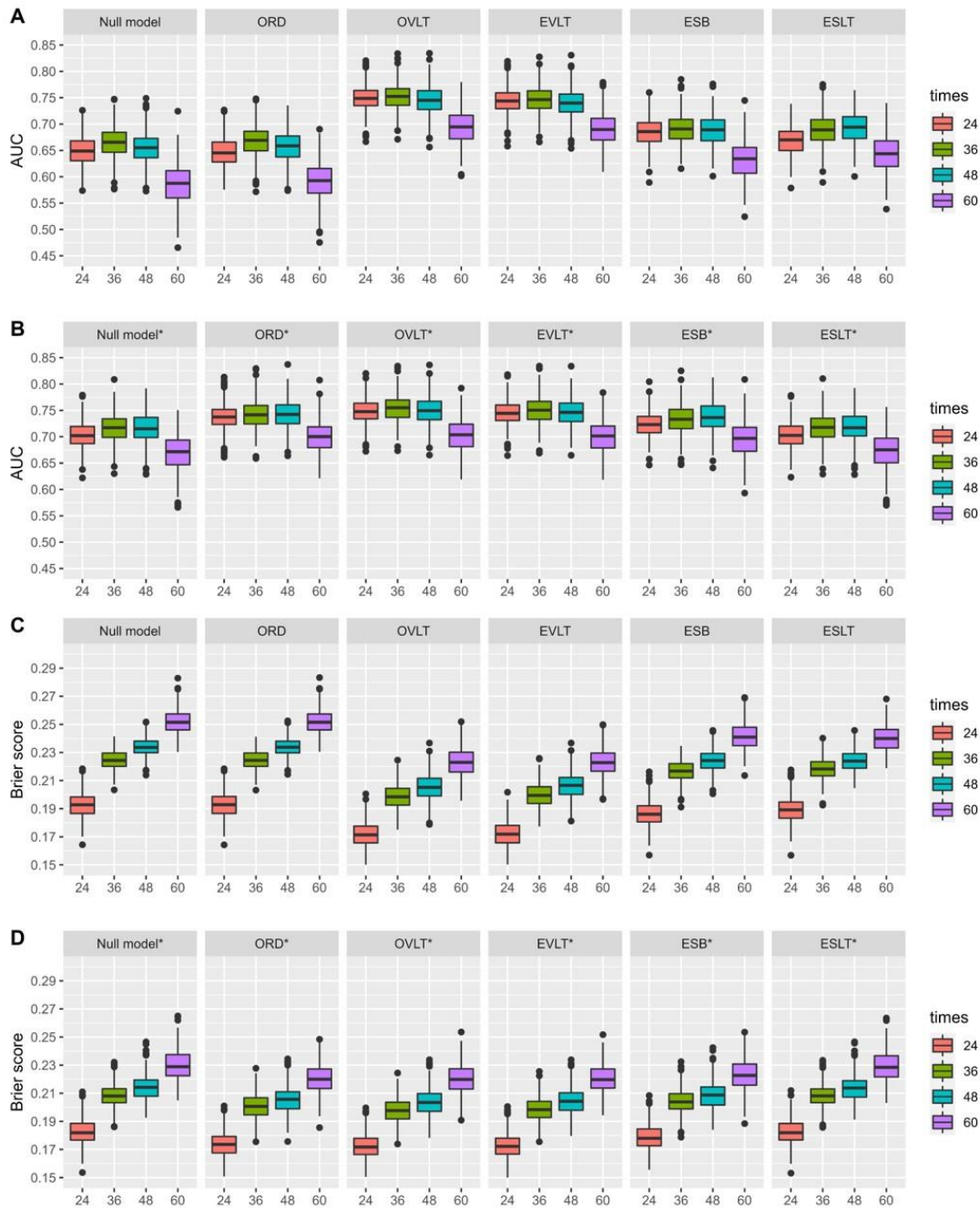


Figure 7.3. AUC (A,B) et score de Brier (C,D) validés par validation croisée pour la prédiction de la survie globale à 24, 36, 48 et 60 mois à partir de l'historique du CA-125 jusqu'à 3 mois en considérant différents résumés du CA-125 (ORD, OVLT, EVLT, ESB,ESLT) ajustés (B,D) ou non (A,C) sur la valeur du CA-125 à l'inclusion. Modèle nul : modèle de survie stratifié à l'étude, ORD : déclin relatif observé, OVLT : valeur observée au temps de Landmark, EVLT : valeur estimée au temps de Landmark, ESB : pente estimée au départ, ESLT : pente estimée au temps de Landmark, * modèles ajustés sur la valeur du CA-125 au départ.

Note : L'essai SCOTROC-1 a été exclu pour le temps de Landmark de 6 mois et le temps de prédiction à 36 mois, et pour le temps de Landmark à 3 mois et le temps de prédiction de 48 mois ; les essais SCOTROC-4 et MITO-7 sont exclus pour les deux temps de Landmark et pour le temps de prédiction de 60 mois.

A. AUC	Time	T=24	T=36	T=48	T=60
	Null model	0.648 [0.593 ; 0.702]	0.665 [0.608 ; 0.717]	0.654 [0.594 ; 0.711]	0.586 [0.51 ; 0.652]
	ORD	0.646 [0.59 ; 0.697]	0.668 [0.611 ; 0.724]	0.657 [0.599 ; 0.711]	0.591 [0.514 ; 0.659]
	OVLT	0.749 [0.702 ; 0.796]	0.751 [0.704 ; 0.801]	0.746 [0.69 ; 0.799]	0.695 [0.634 ; 0.753]
	EVLT	0.745 [0.699 ; 0.796]	0.746 [0.698 ; 0.794]	0.741 [0.688 ; 0.795]	0.691 [0.631 ; 0.75]
	ESB	0.685 [0.637 ; 0.739]	0.691 [0.635 ; 0.749]	0.689 [0.633 ; 0.745]	0.632 [0.572 ; 0.701]
	ESLT	0.668 [0.616 ; 0.722]	0.689 [0.635 ; 0.743]	0.694 [0.639 ; 0.748]	0.643 [0.583 ; 0.707]
B. AUC*	Null model*	0.703 [0.658 ; 0.752]	0.717 [0.664 ; 0.771]	0.716 [0.663 ; 0.77]	0.669 [0.604 ; 0.725]
	ORD*	0.738 [0.693 ; 0.789]	0.742 [0.69 ; 0.791]	0.743 [0.689 ; 0.798]	0.7 [0.64 ; 0.763]
	OVLT*	0.749 [0.705 ; 0.796]	0.754 [0.704 ; 0.804]	0.75 [0.696 ; 0.801]	0.703 [0.644 ; 0.762]
	EVLT*	0.745 [0.701 ; 0.794]	0.75 [0.701 ; 0.799]	0.746 [0.691 ; 0.801]	0.7 [0.641 ; 0.756]
	ESB*	0.723 [0.677 ; 0.77]	0.734 [0.682 ; 0.787]	0.738 [0.68 ; 0.792]	0.696 [0.635 ; 0.752]
	ESLT*	0.704 [0.658 ; 0.752]	0.718 [0.665 ; 0.772]	0.719 [0.665 ; 0.773]	0.672 [0.608 ; 0.73]
C. Brier	Null Model	0.193 [0.174 ; 0.212]	0.225 [0.212 ; 0.239]	0.234 [0.222 ; 0.247]	0.252 [0.236 ; 0.269]
	ORD	0.193 [0.174 ; 0.212]	0.225 [0.212 ; 0.239]	0.234 [0.222 ; 0.247]	0.252 [0.236 ; 0.269]
	OVLT	0.171 [0.153 ; 0.191]	0.199 [0.181 ; 0.216]	0.205 [0.188 ; 0.223]	0.223 [0.204 ; 0.243]
	EVLT	0.172 [0.153 ; 0.191]	0.2 [0.183 ; 0.217]	0.206 [0.189 ; 0.224]	0.223 [0.204 ; 0.242]
	ESB	0.186 [0.169 ; 0.206]	0.217 [0.202 ; 0.231]	0.224 [0.209 ; 0.239]	0.241 [0.225 ; 0.261]
	ESLT	0.189 [0.171 ; 0.209]	0.218 [0.204 ; 0.233]	0.224 [0.21 ; 0.238]	0.24 [0.224 ; 0.257]
D. Brier*	Null model*	0.182 [0.164 ; 0.2]	0.208 [0.192 ; 0.225]	0.214 [0.198 ; 0.232]	0.23 [0.212 ; 0.252]
	ORD*	0.173 [0.154 ; 0.192]	0.201 [0.183 ; 0.218]	0.205 [0.186 ; 0.224]	0.22 [0.201 ; 0.24]
	OVLT*	0.172 [0.153 ; 0.19]	0.198 [0.18 ; 0.216]	0.204 [0.186 ; 0.223]	0.22 [0.202 ; 0.24]
	EVLT*	0.172 [0.153 ; 0.191]	0.199 [0.181 ; 0.216]	0.204 [0.186 ; 0.224]	0.22 [0.202 ; 0.24]
	ESB*	0.178 [0.16 ; 0.197]	0.204 [0.188 ; 0.221]	0.208 [0.19 ; 0.227]	0.223 [0.205 ; 0.244]
	ESLT*	0.182 [0.164 ; 0.201]	0.208 [0.192 ; 0.225]	0.214 [0.197 ; 0.231]	0.229 [0.212 ; 0.251]

Table 7.4 : AUC (A,B) et score de Brier (C,D) validés par validation croisée 10 fois pour la prévision de la survie globale à 24, 36, 48 et 60 mois à partir de l'historique du CA125 jusqu'à 3 mois en considérant différents résumés du CA-125.

Null model: model de survie stratifié à l'essai, ORD: Décroissance relative Observée, OVLT: Valeur observée du CA-125 au temps de landmark, EVLT: Valeur estimée au temps de landmark, ESB: Pente estimée à l'initiation du traitement, ESLT: Pente estimée au temps de landmark

* tous les modèles sont ajustés au CA-125 à l'initiation du traitement

7.3.4 Temps de Landmark à 6 mois

Les analyses du temps de Landmark à six mois ont donné des résultats similaires à ceux des analyses à trois mois, avec une performance plus faible pour un horizon de prédiction plus long. La valeur estimée à 6 mois et la dernière valeur observée avant 6 mois avaient des AUC de 0,731 [0,678 ; 0,778] et 0,735 [0,683 ; 0,783], respectivement, à 24 mois et des AUC de 0,656 [0,586 ; 0,734] et 0,66 [0,59 ; 0,734] à 60 mois (Figure 7.4, Tableau 7.5).

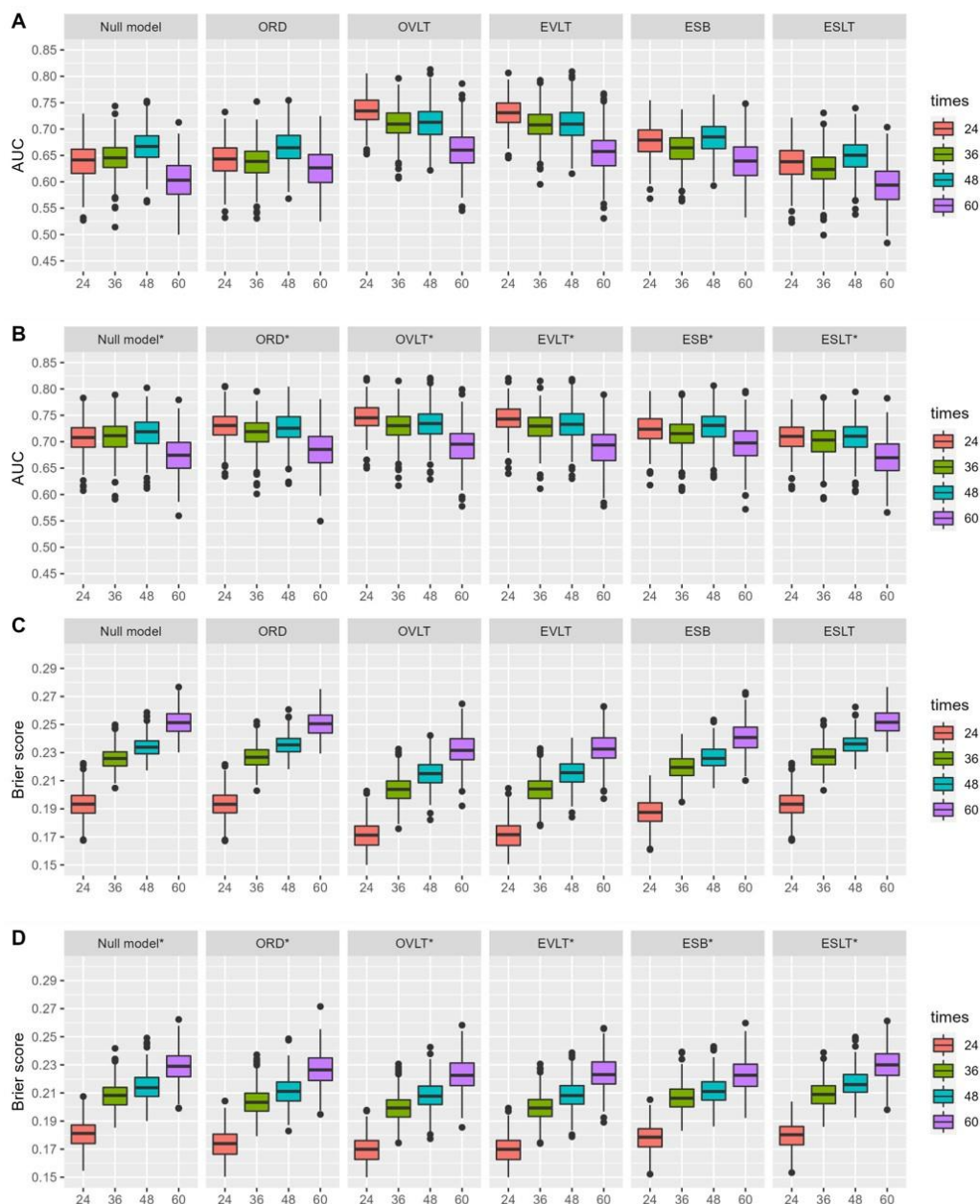


Figure 7.4. AUC (A,B) et scores de Brier (C,D) validés par validation croisée pour la prédiction de la survie globale à 24, 36, 48 et 60 mois à partir de l'historique du CA-125 jusqu'à 6 mois, en considérant différents résumés du CA-125 (ORD, OVLT, EVLT, ESB, ESLT) ajustés (B,D) ou non (A,C) en fonction de la valeur du CA-125 à l'inclusion. Modèle nul : modèle de survie stratifié à l'étude, ORD : déclin relatif observé, OVLT : valeur observée au temps de Landmark. Modèle nul : modèle de survie stratifié par étude, ORD : déclin relatif observé, OVLT : valeur observée au temps de Landmark, EVLT : valeur estimée au temps de Landmark, ESB : pente estimée au départ, ESLT : pente estimée au temps de Landmark. * modèles ajustés pour le CA-125 au début de l'étude. Note : L'essai SCOTROC-1 a été exclu pour le temps repère à 6 mois et le temps de prédiction à 36 mois, ainsi qu'au temps de Landmark à 3 mois et au temps de prédiction à 48 mois. Les essais SCOTROC-4 et MITO-7 ont été exclus pour les deux temps de Landmark et pour le temps de prédiction de 60 mois.

Table 7.5: AUC (A,B) et score de Brier (C,D) validés par validation croisée 10 fois pour la prédiction de la survie globale à 24, 36, 48 et 60 mois à partir de l'historique du CA125 jusqu'à 6 mois en considérant différents résumés du CA125 Landmark 6 mois

	Time	T=24	T=36	T=48	T=60
A. AUC	Null model	0.639 [0.574 ; 0.699]	0.646 [0.58 ; 0.704]	0.667 [0.606 ; 0.728]	0.603 [0.527 ; 0.672]
	ORD	0.643 [0.579 ; 0.703]	0.638 [0.575 ; 0.698]	0.665 [0.606 ; 0.726]	0.625 [0.556 ; 0.696]
	OVLТ	0.735 [0.683 ; 0.783]	0.71 [0.653 ; 0.762]	0.712 [0.654 ; 0.765]	0.66 [0.59 ; 0.734]
	EVLТ	0.731 [0.678 ; 0.778]	0.708 [0.648 ; 0.757]	0.709 [0.652 ; 0.762]	0.656 [0.586 ; 0.734]
	ESB	0.677 [0.615 ; 0.731]	0.663 [0.598 ; 0.72]	0.683 [0.618 ; 0.735]	0.639 [0.564 ; 0.706]
	ESLT	0.636 [0.574 ; 0.696]	0.624 [0.555 ; 0.686]	0.649 [0.591 ; 0.714]	0.593 [0.511 ; 0.667]
B. AUC*	Null model*	0.708 [0.65 ; 0.761]	0.709 [0.652 ; 0.762]	0.717 [0.656 ; 0.771]	0.675 [0.609 ; 0.74]
	ORD*	0.73 [0.675 ; 0.779]	0.717 [0.66 ; 0.767]	0.727 [0.67 ; 0.783]	0.686 [0.616 ; 0.756]
	OVLТ*	0.746 [0.698 ; 0.791]	0.73 [0.672 ; 0.778]	0.733 [0.674 ; 0.786]	0.692 [0.621 ; 0.755]
	EVLТ*	0.744 [0.695 ; 0.791]	0.729 [0.672 ; 0.776]	0.732 [0.673 ; 0.784]	0.69 [0.619 ; 0.76]
	ESB*	0.723 [0.67 ; 0.773]	0.714 [0.655 ; 0.767]	0.728 [0.668 ; 0.777]	0.697 [0.636 ; 0.761]
	ESLT*	0.709 [0.652 ; 0.762]	0.7 [0.644 ; 0.752]	0.709 [0.648 ; 0.767]	0.671 [0.604 ; 0.74]
C. Brier	Null Model	0.193 [0.174 ; 0.212]	0.226 [0.212 ; 0.24]	0.234 [0.221 ; 0.247]	0.251 [0.235 ; 0.269]
	ORD	0.193 [0.174 ; 0.212]	0.227 [0.212 ; 0.241]	0.236 [0.223 ; 0.249]	0.25 [0.235 ; 0.268]
	OVLТ	0.171 [0.151 ; 0.191]	0.204 [0.187 ; 0.224]	0.215 [0.199 ; 0.233]	0.232 [0.211 ; 0.254]

	EVLT	0.171 [0.151 ; 0.192]	0.204 [0.187 ; 0.224]	0.216 [0.199 ; 0.234]	0.233 [0.211 ; 0.254]
	ESB	0.188 [0.169 ; 0.207]	0.22 [0.205 ; 0.237]	0.227 [0.213 ; 0.245]	0.241 [0.222 ; 0.263]
	ESLT	0.193 [0.174 ; 0.212]	0.227 [0.212 ; 0.242]	0.236 [0.223 ; 0.25]	0.252 [0.235 ; 0.269]
D. Brier*	Null model*	0.181 [0.161 ; 0.2]	0.208 [0.191 ; 0.228]	0.214 [0.196 ; 0.234]	0.229 [0.209 ; 0.251]
	ORD*	0.174 [0.153 ; 0.194]	0.204 [0.186 ; 0.225]	0.211 [0.193 ; 0.231]	0.227 [0.203 ; 0.25]
	OVLT*	0.169 [0.149 ; 0.188]	0.199 [0.182 ; 0.219]	0.208 [0.191 ; 0.227]	0.223 [0.202 ; 0.246]
	EVLT*	0.169 [0.148 ; 0.189]	0.199 [0.182 ; 0.219]	0.209 [0.192 ; 0.226]	0.224 [0.202 ; 0.247]
	ESB*	0.178 [0.159 ; 0.196]	0.207 [0.189 ; 0.227]	0.212 [0.194 ; 0.232]	0.223 [0.202 ; 0.245]
	ESLT*	0.18 [0.16 ; 0.198]	0.209 [0.192 ; 0.228]	0.217 [0.199 ; 0.236]	0.23 [0.209 ; 0.253]

Null model: model de survie stratifié à l'essai, ORD: Décroissance relative Observée, OVLT: Valeur observée du CA-125 au temps de landmark, EVLT: Valeur estimée au temps de landmark, ESB: Pente estimée à l'initiation du traitement, ESLT: Pente estimée au temps de landmark

* tous les modèles sont ajustés au CA-125 à l'initiation du traitement

Le modèle linéaire mixte hiérarchique était bien calibré, comme le montrent les figures 7.5-7.9.

Calibration - ORD

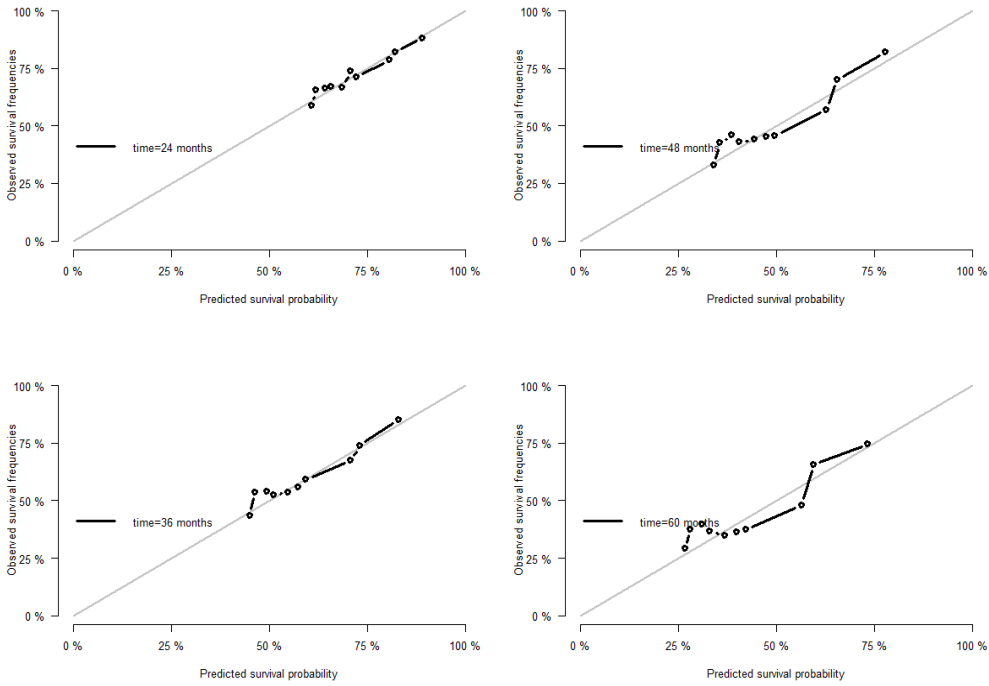


Figure 7.5: Calibration pour taux de déclin observé (ORD) - 3 mois

Calibration - OVLT

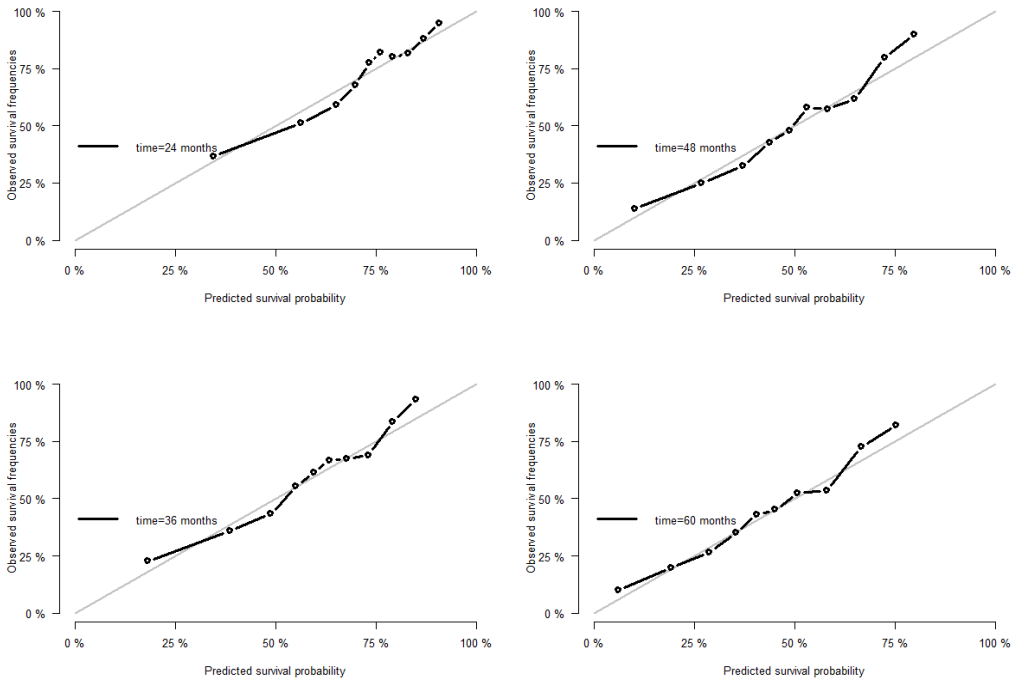


Figure 7.6: Calibration pour la valeur observée du CA-125 (OVLT) à 3 mois

Calibration - EVLT

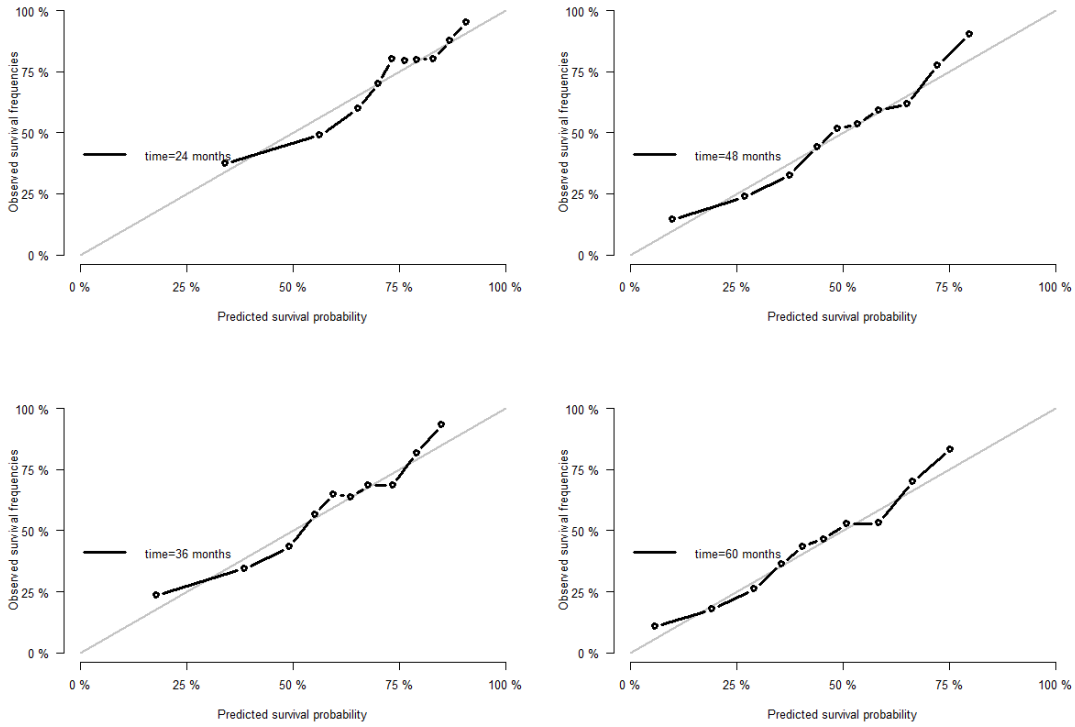


Figure 7.7: Calibration pour la valeur estimé (EVL) à 3 mois

Calibration - ESB

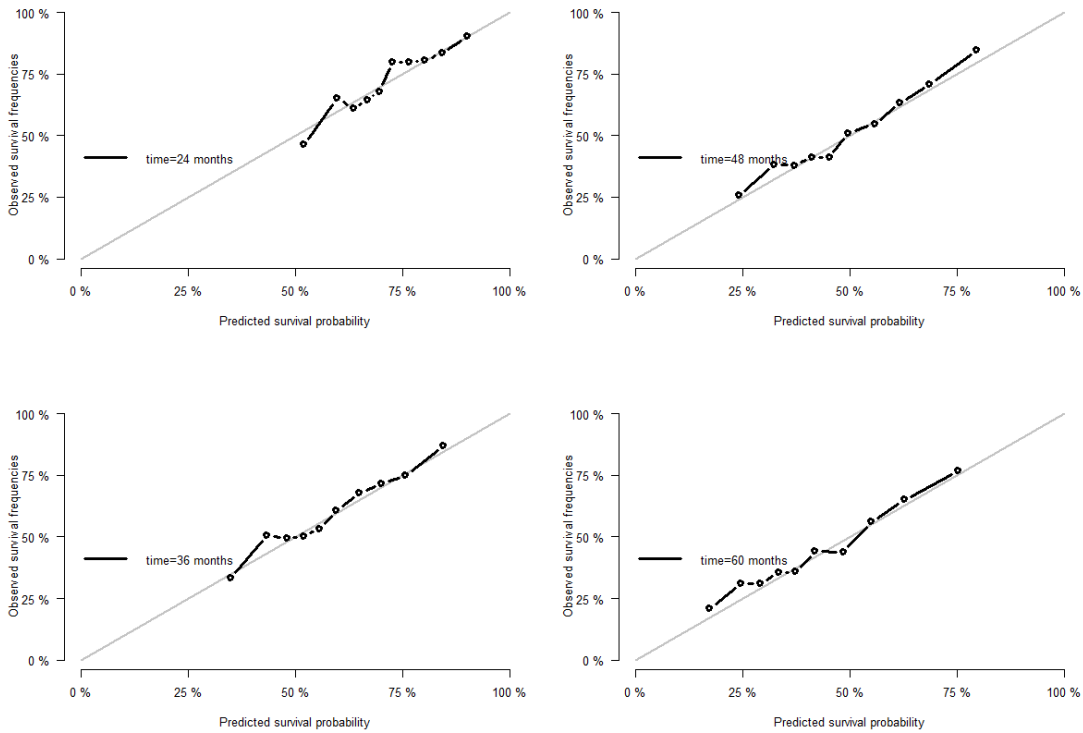


Figure 7.8: Calibration pour la pente estimée à l'initiation du traitement (ESB)

Calibration - ESLT

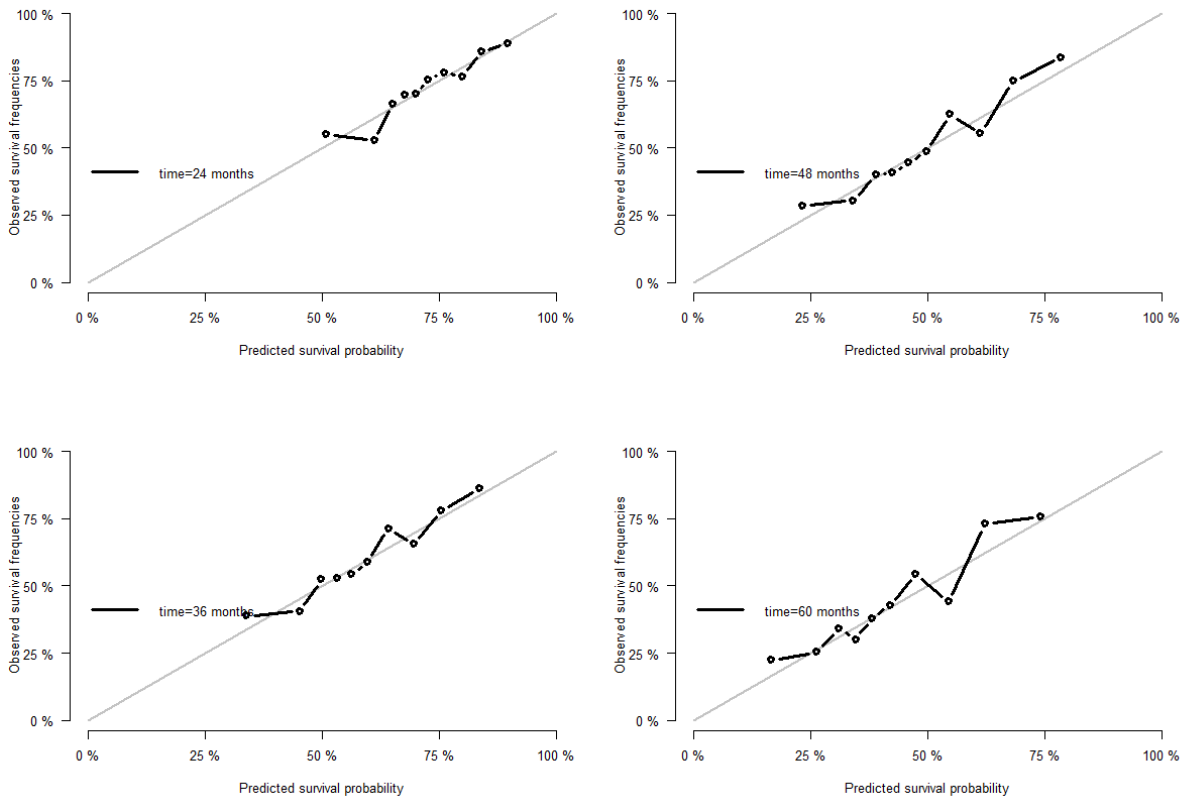


Figure 7.9: Calibration pour la pente estimée au temps de landmark (ESLT) 3 mois

7.3.5. Performance de la synthèse du CA-125 pour les sous-groupes de patients en fonction des caractéristiques

Dans le cas des patients au stade FIGO III, la valeur estimée à 3 mois a fourni une valeur d'AUC de 0,709 [0,640 ; 0,772] et un score de Brier de 0,190 [0,167 ; 0,214] contre 0,684 [0,446 ; 0,878] et un score de Brier de 0,09 [0,062 ; 0,123] pour les stades FIGO IC et II à 24 mois. Pour la maladie résiduelle ≥ 1 cm vs. < 1 cm à 24 mois, les AUC, sans ajouter le CA-125 au départ, étaient respectivement de 0,707 [0,634 ; 0,768] et 0,687 [0,589 ; 0,790]. De plus le modèle a fonctionné de manière équivalente après l'ajout de la valeur de base du CA-125 (voir les tableaux 7.6 et 7.7 et les figures 7.10 et 7.11).

Tableau 7.6: Performances prédictives par sous-groupes pour la valeur estimée du CA-125 à 3 mois

A. AUC	Time	T=24	T=36	T=48	T=60
	Figo I & II	0.684 [0.446 ; 0.878]	0.653 [0.467 ; 0.839]	0.683 [0.495 ; 0.861]	0.581 [0.338 ; 0.833]
	Figo III	0.709 [0.64 ; 0.772]	0.694 [0.621 ; 0.766]	0.714 [0.645 ; 0.78]	0.651 [0.569 ; 0.725]
	Figo IV	0.688 [0.563 ; 0.798]	0.692 [0.548 ; 0.816]	0.698 [0.524 ; 0.847]	0.69 [0.506 ; 0.861]
	Res Dis <1cm	0.687 [0.589 ; 0.79]	0.679 [0.577 ; 0.771]	0.698 [0.606 ; 0.788]	0.649 [0.553 ; 0.748]
	Res Dis >=1cm	0.707 [0.634 ; 0.768]	0.707 [0.629 ; 0.782]	0.72 [0.632 ; 0.796]	0.662 [0.548 ; 0.763]
B. Brier	Figo I & II	0.091 [0.062 ; 0.123]	0.132 [0.097 ; 0.168]	0.162 [0.124 ; 0.207]	0.204 [0.156 ; 0.258]
	Figo III	0.19 [0.167 ; 0.214]	0.221 [0.197 ; 0.246]	0.216 [0.196 ; 0.239]	0.231 [0.21 ; 0.254]
	Figo IV	0.235 [0.187 ; 0.286]	0.23 [0.189 ; 0.276]	0.201 [0.156 ; 0.241]	0.181 [0.142 ; 0.228]
	Res Dis <1cm	0.124 [0.099 ; 0.151]	0.175 [0.148 ; 0.2]	0.2 [0.175 ; 0.224]	0.23 [0.203 ; 0.256]
	Res Dis >=1cm	0.218 [0.193 ; 0.247]	0.224 [0.201 ; 0.249]	0.205 [0.182 ; 0.233]	0.211 [0.182 ; 0.243]

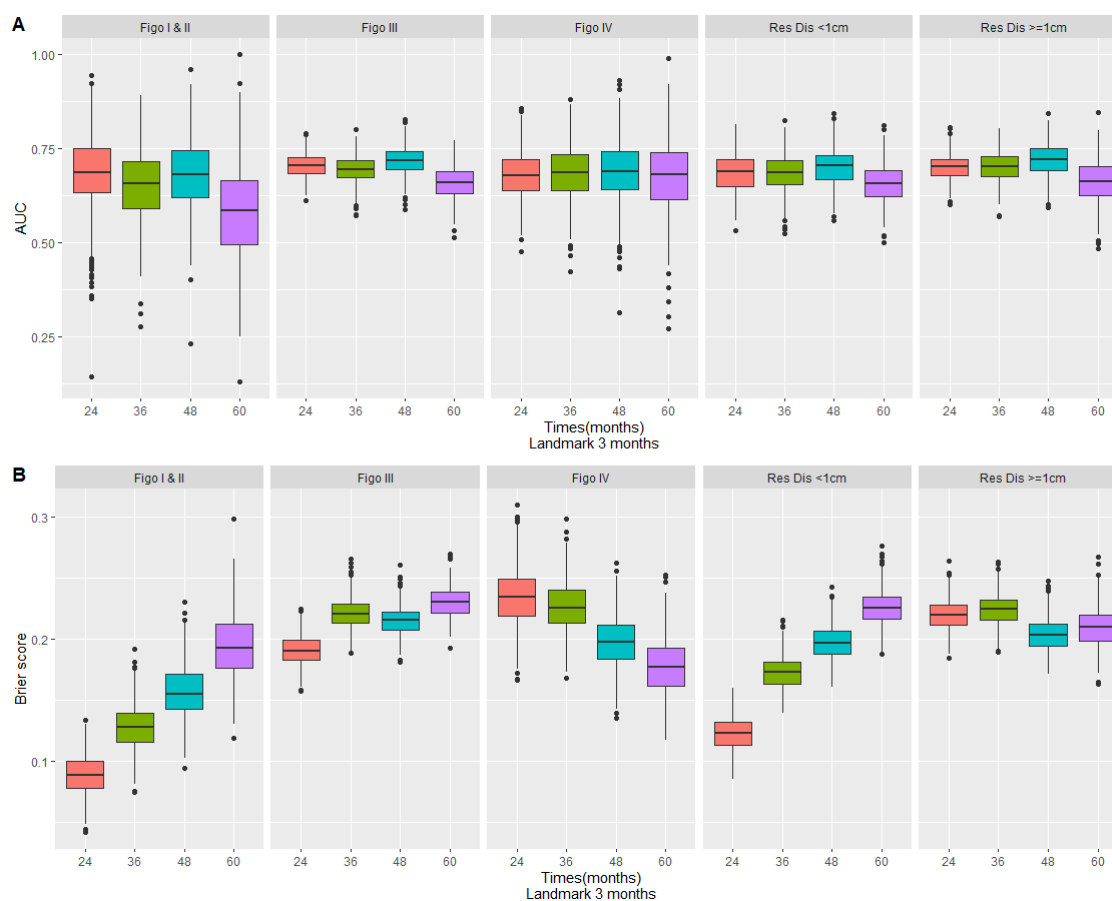


Figure 7.10: Performances prédictives mesurées par l'AUC (A) et le score de Brier (B) par sous-groupes pour la valeur estimée du CA-125 à 3 mois

Tableau 7.7: Performances prédictives par sous-groupes pour la valeur estimée du CA-125 à 6 mois

A. AUC	Time	T=24	T=36	T=48	T=60
	Figo I & II		0.649 [0.442 ; 0.85]	0.656 [0.435 ; 0.843]	0.665 [0.443 ; 0.85]
Figo III		0.7 [0.636 ; 0.767]	0.682 [0.599 ; 0.745]	0.681 [0.61 ; 0.758]	0.627 [0.53 ; 0.712]
Figo IV		0.688 [0.55 ; 0.821]	0.678 [0.518 ; 0.818]	0.684 [0.499 ; 0.858]	0.653 [0.433 ; 0.877]
Res Dis <1cm		0.682 [0.591 ; 0.771]	0.674 [0.587 ; 0.759]	0.683 [0.596 ; 0.765]	0.622 [0.519 ; 0.721]
Res Dis >=1cm		0.704 [0.63 ; 0.775]	0.691 [0.592 ; 0.774]	0.687 [0.589 ; 0.777]	0.632 [0.509 ; 0.741]
B. Brier	Figo I & II	0.093 [0.063 ; 0.123]	0.134 [0.099 ; 0.172]	0.171 [0.13 ; 0.219]	0.223 [0.169 ; 0.282]

	Figo III	0.19 [0.166 ; 0.215]	0.222 [0.198 ; 0.246]	0.225 [0.203 ; 0.245]	0.237 [0.213 ; 0.263]
	Figo IV	0.233 [0.177 ; 0.295]	0.236 [0.191 ; 0.285]	0.212 [0.166 ; 0.252]	0.198 [0.148 ; 0.246]
	Res Dis <1cm	0.126 [0.103 ; 0.152]	0.177 [0.154 ; 0.203]	0.207 [0.184 ; 0.232]	0.241 [0.212 ; 0.272]
	Res Dis \geq 1cm	0.218 [0.192 ; 0.25]	0.229 [0.201 ; 0.26]	0.217 [0.191 ; 0.243]	0.219 [0.188 ; 0.251]

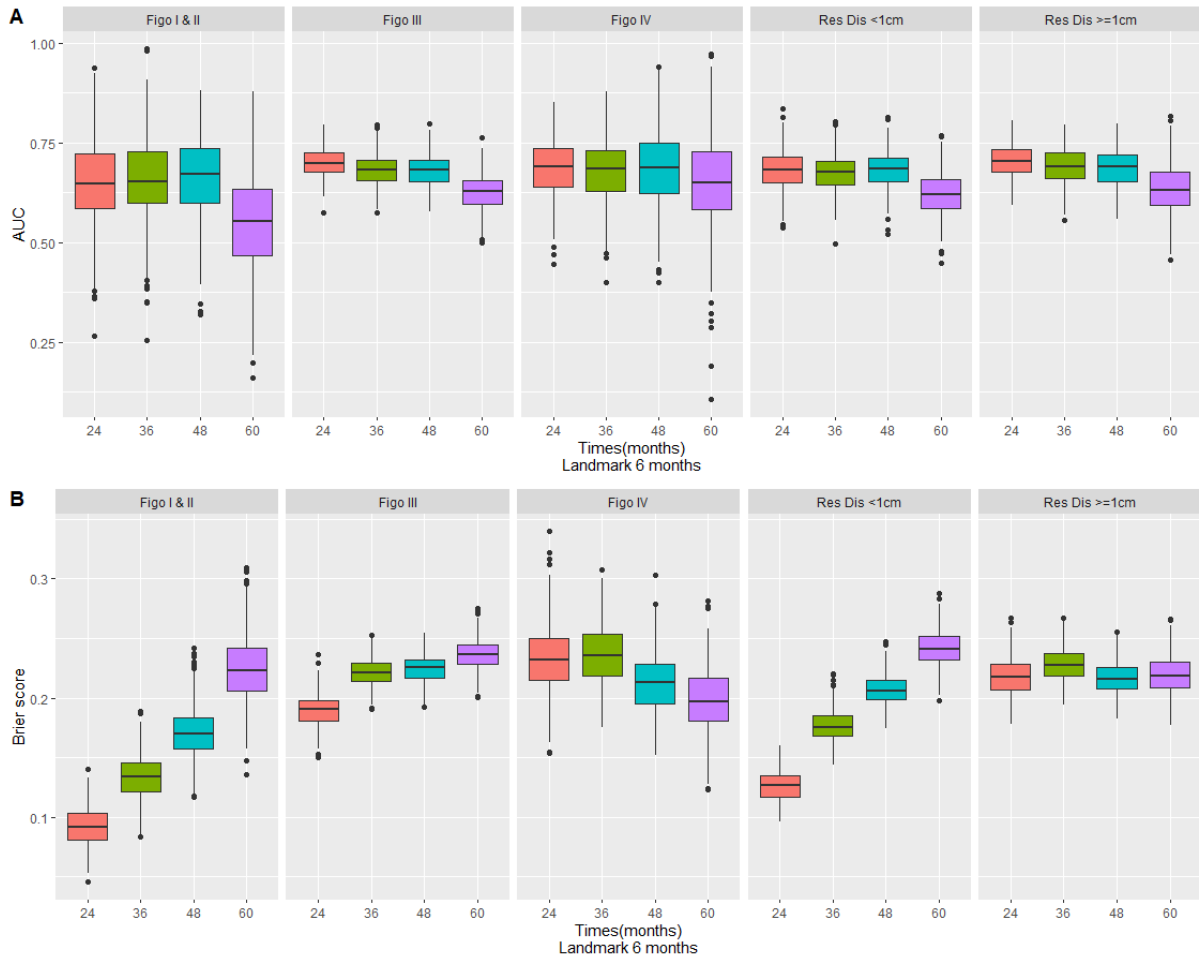


Figure 7.11: Performances prédictives mesurées par l'AUC (A) et le score de Brier (B) par sous-groupes pour la valeur estimée du CA-125 à 6 mois

7.3.6 Analyse de sensibilité et valeurs prédictives de la valeur normale du CA-125

L'analyse de sensibilité des patientes du groupe de traitement standard a montré des résultats équivalents pour les deux temps de Landmark. (Tables des annexes 6 et 7 et les figures 7.12 et 7.13).

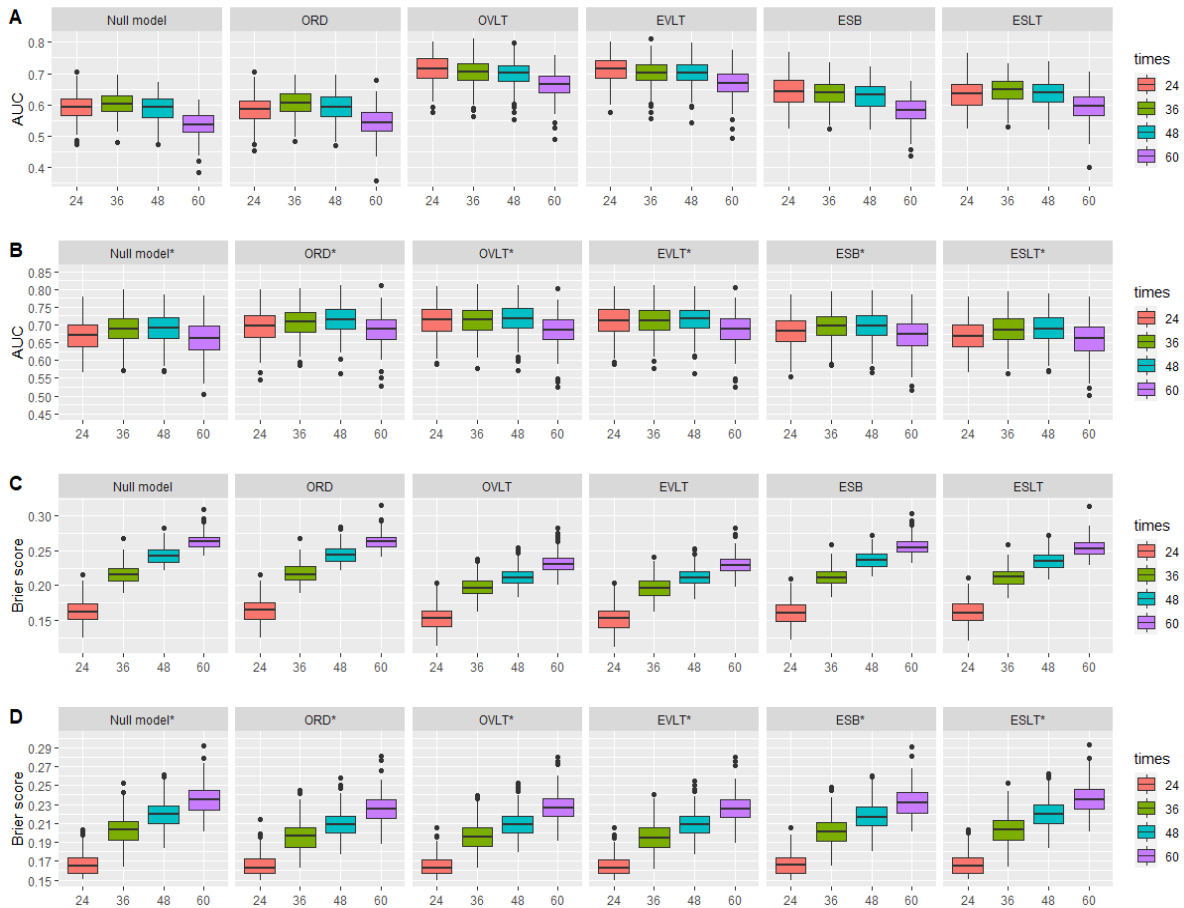


Figure 7.12: AUC (A,B) et score de Brier (C,D) validés par validation croisée 5 fois pour la prédiction de la survie globale à 24, 36, 48 et 60 mois à partir de l'historique du CA125 jusqu'à 3 mois dans le bras Régime standard en considérant différents résumés du CA125

Null model: model de survie stratifié à l'essai, ORD: Décroissance relative Observée, OVLT: Valeur observée du CA-125 au temps de landmark, EVLT: Valeur estimée au temps de landmark, ESB: Pente estimée à l'initiation du traitement, ESLT: Pente estimée au temps de landmark

* tous les modèles sont ajustés au CA-125 à l'initiation du traitement

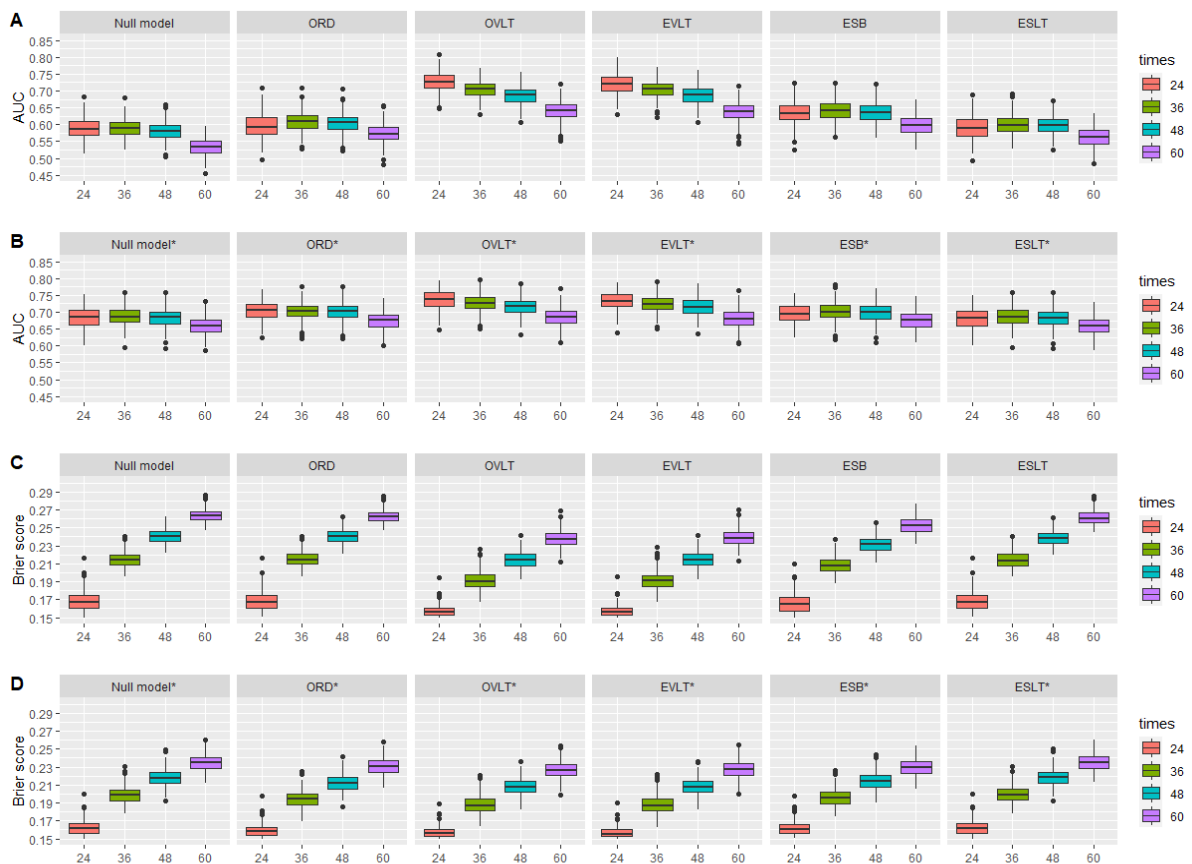


Figure 7.13 : AUC (A,B) et score de Brier (C,D) validés par validation croisée 5 fois pour la prédiction de la survie globale à 24, 36, 48 et 60 mois à partir de l'historique du CA125 jusqu'à 6 mois dans le bras Régime standard en considérant différents résumés du CA125

Null model: model de survie stratifié à l'essai, ORD: Décroissance relative Observée, OVLT: Valeur observée du CA-125 au temps de landmark, EVLT: Valeur estimée au temps de landmark, ESB: Pente estimée à l'initiation du traitement, ESLT: Pente estimée au temps de landmark

* tous les modèles sont ajustés au CA-125 à l'initiation du traitement

Enfin, nous avons examiné la sensibilité et la spécificité de la valeur du CA-125 à 3 mois. Un patient avec une valeur de CA-125 < 35 avait une probabilité de décès de 20% à 24 mois, et de 30%, 38%, et 66% à 36, 48, et 60 mois, respectivement. Au contraire, un patient avec une valeur de CA-125 > 35 avait une probabilité de décès de 51% à 24 mois, et de 66%, 75%, et 79% à 36, 48, et 60 mois, respectivement.

7.4 Discussion

Chez les patientes atteintes d'un cancer de l'ovaire de stade FIGO II à IV nouvellement diagnostiqué, le taux sérique du biomarqueur CA-125 à 3 mois après la fin du traitement initial a montré la meilleure prédiction de la probabilité de survie globale à 24, 36 et 48 mois par rapport à diverses mesures de résumé de la décroissance du CA-125. Quel que soit le résumé en question, l'AUC était relativement similaire aux horizons de 24, 36 et 48 mois, avec une

baisse de performance à 60 mois. En revanche, plus l'horizon de prédiction était court, plus l'erreur de prédiction était faible, tel que mesurée par le score de Brier. La dernière valeur observée du CA-125 avant ou à 3 mois et l'estimation basée sur le modèle à mesures répétées ont été associées à des AUC à 24 mois de 0,749 et 0,745, respectivement. Un délai de référence de 6 mois n'a pas amélioré les performances prédictives, ce qui suggère qu'une évaluation précoce offre le meilleur compromis entre l'application clinique et les performances statistiques. La capacité prédictive des deux mesures du résumé du CA-125 est restée équivalente après l'ajout de la valeur initiale du CA-125. Une amélioration importante des mesures de taux a été observée après l'ajout de la valeur initiale du CA-125. Par exemple, le taux de déclin observé a atteint 0,738 à 24 mois contre 0,646 sans l'ajout de la valeur de référence du CA-125. Ceci reflète l'importance de prendre en compte la valeur initiale du CA-125 afin de mieux interpréter la cinétique du CA-125. Malgré la présence d'hétérogénéité, probablement due à des critères d'inclusion ou à des normes de soins de santé différents d'une étude à l'autre, le contexte méta-analytique et la validation croisée ont permis de tirer des résultats robustes.

Une valeur de CA-125 supérieure à 35 à 3 mois entraînait un risque de décès plus élevé, comme 51 % à 24 mois ou jusqu'à 79 % à 60 mois.

Dans le cas d'une surveillance à 3 mois, la cinétique du CA-125 pourrait aider à prendre des décisions concernant l'intensification du suivi ou l'anticipation d'un changement de traitement.

Plus précisément, un patient dont le taux de CA-125 est supérieur à 35 IU/mL à 3 mois bénéficierait d'une surveillance plus étroite, avec la mesure de la concentration de CA-125 et des scanners, afin d'anticiper la nécessité d'un possible changement de traitement. Ces patients pourraient être de bons candidats pour des essais cliniques en cas de progression.

Les prédictions à plus de 60 mois, après le temps de Landmark, doivent être interprétées avec précaution, car toutes les mesures ont donné des résultats médiocres par rapport aux temps de prédiction antérieurs. Le manque de suivi pourrait expliquer cette performance.

La principale limitation est que les médicaments récemment approuvés, tels que les inhibiteurs de PARP, n'ont pas été inclus dans cette méta-analyse, puisque les essais ont été publiés entre 2001 et 2016. Bien que la norme soit toujours la chimiothérapie, le traitement d'entretien comprenant des inhibiteurs de PARP est désormais la norme de soins.

Cela peut conduire à des associations différentes entre le CA-125 et la survie globale. Une mise à jour incluant des essais avec des médicaments nouvellement approuvés, tels que les inhibiteurs de PARP, serait intéressante pour quantifier davantage la capacité pronostique du CA-125 dans ce contexte. En outre, nous nous sommes appuyés sur des modèles linéaires mixtes pour surmonter les problèmes de données manquantes. Bien que nous ayons considéré des fonctions du temps flexibles et que nous ayons soigneusement évalué la qualité de l'ajustement, cette approche est toujours basée sur des hypothèses paramétriques.

Récemment, le sous-groupe de patientes ayant un statut HRD- ou BRCA-positif a suscité un grand intérêt en tant que sous-type distinct, ce qui peut nous amener à remettre en question

la robustesse de nos résultats pour ce sous-groupe [171]. Cette question pourrait être étudiée par une mise à jour de la méta-analyse avec les agents les plus récents. Enfin, cette méta-analyse est limitée au CA-125 sérique, bien que d'autres biomarqueurs ayant des capacités pronostiques potentielles, tels que la protéine de l'épididyme humain de type 4 (HE4), l'ostéopontine, la mésothéline (MSLN) et le récepteur α des folates (FOLR1), puissent également être mesurés en fonction des circonstances et de l'état des patients [172]. Par conséquent, nous avons concentré notre recherche sur le CA-125 sérique, qui est utilisé en routine, et nous avons essayé de quantifier sa capacité pronostique. Néanmoins, le même type de travail pourrait être étendu, en combinant d'autres marqueurs dans des contextes spécifiques pour évaluer le risque de progression ou de décès.

L'évaluation de KELIM chez les patients traités par un traitement de maintenance par veliparib ou bevacizumab ont donné des résultats prometteurs qui suggèrent une performance similaire entre KELIM et les agents chimio-thérapeutiques rapportés. Néanmoins, une évaluation plus poussée de la valeur prédictive du CA-125 à 3 mois dans le cadre d'un traitement d'entretien par PARP est nécessaire.

Le développement d'approches basées sur des modèles a été motivé par l'impact de la variabilité entre les différents tests utilisés dans la pratique courante sur l'interprétation des résultats des stratégies à un et deux points dans le temps. Il est intéressant de noter que la fenêtre temporelle de 3 mois trouvée dans la présente étude est cohérente avec la période de 100 premiers jours de traitement utilisée pour le calcul de KELIM.

7.5. Conclusions

En conclusion, la surveillance du CA-125 d'une durée de trois mois après le début du traitement pourrait contribuer à fournir des informations individuelles aux patients sur la base de la trajectoire initiale du CA-125. Plus précisément, cela pourrait aider à fournir un outil informatif afin de guider les cliniciens dans la prise de décision concernant la survie globale jusqu'à 48 mois.

Le chapitre suivant a évalué la trajectoire du CA-125 les premiers mois après l'initiation du traitement de première ligne pour prédire la résistance des patientes atteintes d'un cancer de l'ovaire avancé au traitement en base de platine. Pour cette étude, nous avons utilisé les essais de la méta-analyse MAOV qui ont testé un traitement en base de platine.

8. LE ROLE DU CA-125 DANS LA DETECTION DE LA RESISTANCE A LA CHIMIOTHERAPIE A BASE DE PLATINE DANS LE CANCER DE L'OVAIRE AVANCE : META-ANALYSE DES DONNEES INDIVIDUELLES DES PATIENTS DU GCIG

8.1 Introduction

Le cancer de l'ovaire est le huitième cancer le plus fréquent et figure parmi les 14 causes les plus fréquentes de mortalité par cancer chez les femmes dans le monde entier, avec une incidence plus élevée dans les pays développés. [1] Cette maladie est caractérisée par de nombreux sous-types avec une grande hétérogénéité et divers phénotypes et génotypes avec des mécanismes biologiques potentiellement différents. [34], [173] En outre, en raison de symptômes silencieux, le cancer de l'ovaire est diagnostiqué principalement à des stades avancés. [174]. La norme de soins comprend une chirurgie cytoréductive suivie d'une chimiothérapie cytotoxique à base de platine et de taxane, et récemment suivie de traitements par inhibiteurs de parp avec ou sans bevacizumab chez les patients présentant une mutation BRCA ou des maladies HRD [61], [175]. Même si la réponse au traitement de première ligne est d'environ 60 à 80 % [176], la plupart des patients verront leur maladie progresser avec une résistance au traitement à base de platine qui est définie comme une progression de la maladie dans les six mois suivant la fin de la chimiothérapie de première ligne à base de platine. [3], [177], [178] Les patients « résistants au platine » présentent un faible taux de réponse (<15%) avec une survie médiane de 12 mois tandis que les patients « sensibles au platine » présentent des taux de réponse variables avec une survie médiane de deux ans [173]. Pour détecter une résistance précoce au traitement de platine, de grands efforts sont faits pour identifier des outils pronostiques et des biomarqueurs possibles. Cela permettrait de proposer aux patients un traitement alternatif, éventuellement dans le cadre d'un essai clinique.

L'antigène 125 du cancer (CA-125) est une protéine produite principalement par les cellules cancéreuses de l'ovaire et, selon les critères du Gynecologic Cancer Intergroup (GCIG), [67] c'est utilisé pour surveiller les patients en routine.

Le CA-125 continue d'attirer une attention particulière car son association avec la progression et la survie globale dans le cancer de l'ovaire a été prouvée dans plusieurs études individuellement [179], [180] et il constitue un moyen simple et rapide d'évaluer les changements de la maladie [181]. Cependant, la question de savoir si la dynamique du changement du CA-125 est suffisamment associée à la résistance au platine, c'est-à-dire à la progression dans les 12 premiers mois après l'initiation du traitement systémique, reste en suspens. Cette question est fortement liée au choix du meilleur résumé de la dynamique du CA-125 dans le temps.

L'association de la dynamique précoce du CA-125 avec la survie globale a été récemment évaluée [182], [183] dans le cadre d'une méta-analyse de données individuelles. De multiples façons de résumer la trajectoire précoce du CA-125 ont été explorées (par exemple, la valeur du CA-125 au départ, la valeur du CA-125 ou la diminution à 3 mois, ainsi que des résumés

basés sur un modèle, par exemple, avec l'utilisation d'un modèle linéaire mixte ou basé sur les paramètres pharmacocinétiques et pharmacodynamiques du traitement, la constante de taux d'ELIMination K (KELIM) à 3 mois).

L'analyse de la valeur pronostique des trajectoires des biomarqueurs soulève des questions méthodologiques (données manquantes, erreurs de mesure, corrélation intra-individuelle, endogénéité du biomarqueur) qui ont conduit au développement de modèles de prédiction basés sur l'approche du landmarking ou de modèles conjoints [166], [184], [185] et d'estimateurs appropriés de la précision prédictive [12], [13]. Par exemple, Karamouza (2023) a appliqué l'approche de repérage pour prédire la survie globale à partir de résumés de la trajectoire précoce du CA-125 jusqu'à 3 mois qui tenait compte des mesures du CA-125 sujettes aux erreurs et irrégulières chez les femmes atteintes d'un cancer de l'ovaire avancé nouvellement diagnostiqué [182]. Notre approche a donné une valeur discriminante à 2 ans de 0,75 IC 95% [0,70-0,80] et KELIM, en utilisant les valeurs CA-125 des 100 premiers jours pendant le traitement, était aussi bas que 0,62 IC 95% [0,60-0,64]. Cela signifie qu'une femme dont la dynamique du CA-125 est moins bonne a 25 % (ou 38 %) de chances de survivre plus longtemps qu'une femme dont la dynamique du CA-125 est favorable. Un indice discriminatoire AUC inférieur à 0,8 (c'est-à-dire un risque d'erreur de 20 % dans l'identification du pire pronostic) est probablement insuffisant pour guider la décision thérapeutique.

L'objectif de cette étude est d'évaluer si les mesures du CA-125 au cours des 3 premiers mois de traitement peuvent servir de prédicteur fiable de la résistance des patients à la chimiothérapie à base de platine. Nous avons exploité la méta-analyse des données individuelles des patients (IPD) de l'étude du groupe de méta-analyse GCIg [134], et nous nous sommes appuyés sur une approche de repérage [166], [182] dans laquelle les résumés de la trajectoire précoce du CA-125, tels que la valeur prédite à 3 mois ou la pente dérivée des modèles mixtes linéaires, sont utilisés pour prédire la survenue d'une progression à 1 an.

8.2 Matériel et méthodes

8.2.1 Méta-analyse sur données individuelles des patients

Cette étude est basée sur la méta-analyse MAOV, initié par le GCIg, qui a inclus 17 essais contrôlés et randomisés avec des données provenant de 11029 patientes. Cette méta-analyse a étudié les traitements de première ligne chez les patientes atteintes d'un cancer de l'ovaire de stade FIGO IIc à IV [134]. Seuls les essais comprenant un bras de traitement à base de platine ont été sélectionnés. Tous les essais devaient recueillir le CA-125 sérique à l'inclusion dans l'étude et pendant le suivi, avec au moins une deuxième mesure disponible dans les 3 mois suivant la randomisation. Les données extraites pour chaque patiente étaient les suivantes : âge, stade FIGO, maladie résiduelle, type et grade histologique, traitement attribué, survie et état de progression, date de la randomisation, date de la première progression, date du décès ou date du dernier suivi.

Toutes les données et les analyses sont conformes aux procédures standards des méta-analyses sur données individuelles, en suivant les directives PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) [112] et la déclaration TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) [14], [126], [186]. Cette méta-analyse est enregistrée auprès du registre PROSPERO avec le numéro : CRD42017068135.

8.2.2 Mesures du CA-125

Les mesures per-opératoires du CA-125 ont été exclues, car elles sont affectées par l'intervention chirurgicale. Le biomarqueur a été transformé sur une échelle logarithmique naturelle pour obtenir une distribution des valeurs proche de la normale.

8.2.3 Résultats

Le critère d'évaluation principal était la survie sans progression à 12 mois, définie comme le délai entre la randomisation et la progression ou le second cancer, ou le décès, selon ce qui est survenu en premier. Les patients ont été censurés à 18 mois. Deux critères d'évaluation supplémentaires à 15 et 18 mois ont été étudiés dans le cadre d'une analyse de sensibilité.

8.2.4 Analyse statistique

Nous avons suivi la même stratégie d'analyse de Landmark telle que celle décrite dans KARAMOUZA et al (2023) pour prédire le risque de progression six mois après la fin du traitement de première ligne pour les patients ayant reçu une chimiothérapie à base de platine. [182]

8.2.5 Calcul du résumé du CA-125

Nous avons calculé différents résumés de la trajectoire du CA-125 à trois mois, illustrée dans la figure 8.1, à savoir : (i) la valeur observée du CA-125 la plus proche à trois mois (OVLT), (ii) le déclin relatif observé (ORD) du biomarqueur de la mesure à l'initiation du traitement et la dernière mesure du CA-125 la plus proche à trois mois, (iii) la valeur prédite au temps de landmark (EVLT) et (iv-v) les pentes prédites au moment de l'initiation du traitement (ESB) et au temps de landmark à trois mois (ESLT). Les résumés (i) et (ii) ont été directement calculés à partir des observations, tandis que les résumés (iii) à (v) ont été dérivés d'un modèle linéaire mixte hiérarchique (HLMM) ajusté sur les mesures répétées du CA-125 jusqu'à 3 mois. Ce modèle a permis de dériver des résumés à partir de l'ensemble de données avec l'aide des effets emboîtés à trois niveaux. Plus précisément, un effet aléatoire pour la variabilité entre

les essais, un effet aléatoire pour la variabilité entre les patients et une erreur de mesure au niveau de l'observation. La trajectoire du CA-125 a été supposée non linéaire dans le temps. Par conséquent, elle a été capturée par une fonction spline, cubique et naturelle avec trois nœuds internes. Le premier nœud interne a été placé près du début de la surveillance (0,5 mois) afin de détecter une diminution rapide du CA-125.

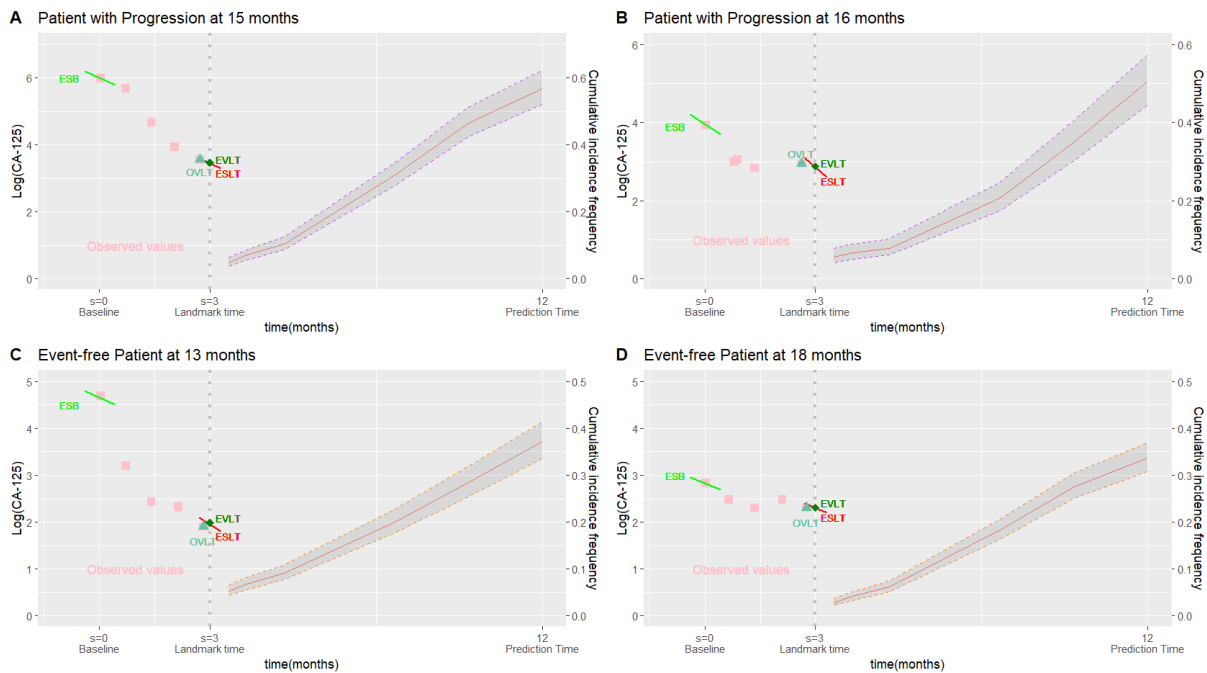


Figure 8.1. Illustration du cadre Landmark : Trajectoires des valeurs observées du CA-125 jusqu'au temps de landmark de $s = 3$ mois pour quatre patients (A, B, C, D) présentant un risque de décès au temps de landmark (#A et #B ont progressé à 15 et 16 mois, respectivement ; #C et #D étaient exempts d'événement après 13 mois et 18 mois, respectivement). La ligne verte représente la pente observée et la ligne rouge représente la pente estimée. Le risque prédit de progression est calculé à partir du temps de landmark jusqu'au temps de prédiction de 12 mois, avec son intervalle de prédiction de 95 %.

8.2.6 Association des résumés du CA-125 avec la progression

La capacité des résumés observés et estimés à prédire la survie sans progression à 12 mois a été évaluée dans un modèle de régression semi-paramétrique de Cox avec un risque de base spécifique à l'étude. Dans l'analyse principale, un résumé du CA-125 à la fois, était le seul facteur prédictif de la survie sans progression. Dans des analyses supplémentaires, le modèle a été ajusté au taux du CA-125 et aux caractéristiques cliniques (statut FIGO et maladie résiduelle) à l'initiation du traitement.

8.2.7 Évaluation des capacités prédictives des résumés du CA-125

La précision prédictive a été évaluée à l'aide de l'aire sous la courbe ROC (AUC) et du score de Brier à 12 mois (ou 15 et 18 mois dans les analyses de sensibilité). Le score AUC permet de mesurer la capacité de discrimination du modèle, c'est-à-dire sa capacité à distinguer avec précision les patients présentant un risque élevé ou faible de progression. Le score de Brier permet d'évaluer à la fois la discrimination et la calibration, c'est-à-dire la capacité à prédire avec précision le niveau de risque absolu. Des estimateurs non paramétriques, basés sur la probabilité inverse des poids de censure (IPCW), ont été utilisés pour tenir en compte les données censurées à droite [140], [148], [187]. Un AUC de 0,5 et un score de Brier de 0,25 suggèrent l'absence de discrimination, tandis qu'une valeur proche de 1 et de zéro, respectivement, indique une capacité de prédiction parfaite.

La calibration de chaque résumé a également été vérifiée visuellement. Les patients ont été répartis en groupes définis par les quantiles des probabilités d'événement prédites et le risque prédit a été représenté graphiquement par rapport au risque observé pour chaque groupe. L'hétérogénéité entre les études de l'association entre chaque résumé et la progression a été quantifiée avec un test de rapport de vraisemblance. Ce dernier a comparé la vraisemblance partielle du modèle pour chaque étude avec la vraisemblance partielle du modèle stratifié utilisant les effets de résumé spécifiques à l'essai [120]

Afin d'éviter tout excès d'optimisme, d'améliorer la généralisation des résultats et de tenir en compte l'hétérogénéité entre les essais, 50 répétitions d'une validation croisée de 10 échantillons a été utilisée. Dans chaque répétition, l'ensemble de données initial a été divisé de manière aléatoire en 10 sous-échantillons de même taille. Dans le contexte d'une méta-analyse, chaque échantillon doit inclure des patients de tous les essais. Les estimations obtenues sur un ensemble de données de formation, constitué de 9 sous-échantillons, ont été utilisées pour calculer les prédictions sur le 10e sous-échantillon servant d'ensemble de données de test. Les mesures de performance ont été calculées sur les 10 ensembles de données de test regroupés. La distribution de l'AUC(t) et du score de Brier sur les 50 répétitions a été présentée dans des diagrammes en boîte et résumée par leur valeur moyenne et l'intervalle correspondant de 95 %. La validation croisée a également été utilisée pour estimer les différences moyennes entre les modèles en termes de AUC et de score de Brier (ainsi que l'intervalle de confiance à 95 % de la distribution des estimations). Premièrement, la valeur ajoutée de chaque résumé a été évaluée par rapport au modèle de référence ; la référence était un modèle de Cox stratifié par étude, qui soit n'incluait aucune covariable (cas A), soit était ajusté pour le CA-125 à l'initiation du traitement (cas A*). Deuxièmement, les différents résumés du CA-125 ont été comparés entre eux en termes d'AUC et de score de Brier.

Les mêmes analyses ont été effectuées dans une sous-population avec des cas complets sur les caractéristiques des patients telles que le stade FIGO et la maladie résiduelle dans le modèle de prédiction. Dans la même étude, nous avons décrit des modèles de Cox qui incluent le

résumé de la trajectoire précoce du CA-125, stratifié sur l'étude et avec ou sans le CA-125 à l'initiation du traitement et les caractéristiques des patientes.

8.2.8 Quantification de la performance

Étant donné que l'indicateur le plus important de la sensibilité au platine est la normalisation du CA-125, nous avons évalué la performance prédictive pour différents seuils de valeur du CA-125 à 3 mois (c'est-à-dire 15, 35, 70, 80, 90, 100, 120, 140 U/mL). Il est important de noter qu'une valeur du CA-125 supérieure à 35 U/mL peut indiquer la présence d'un cancer ou d'autres pathologies. Plus précisément, nous avons évalué la valeur prédictive positive (VPP), c'est-à-dire le ratio des patients qui ont réellement progressé par rapport à tous ceux prédits comme progressifs et la valeur prédictive négative (VPN), c'est-à-dire le ratio des patients réellement diagnostiqués comme non progressifs par rapport à tous ceux prédits comme non progressifs à 12, 15 et 18 mois après le début du traitement. [142] Enfin, nous avons exploré la VPN et la VPP pour les patients en fonction de leur statut FIGO et du statut de la maladie résiduelle.

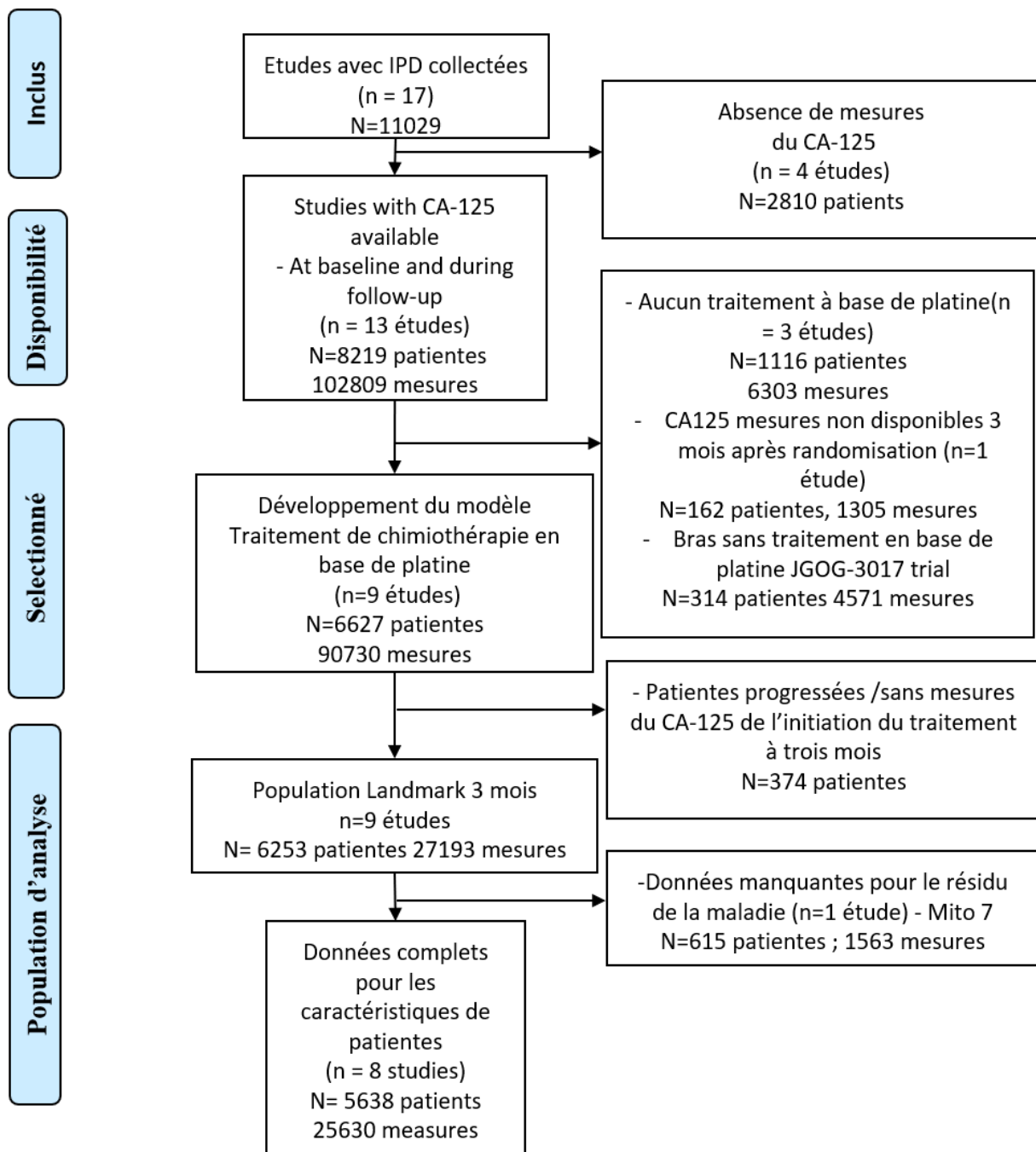
Les analyses statistiques ont été réalisées avec le logiciel R (version 4.1.2, R foundation for Statistical Computing, Vienne, Autriche), avec le package R RiskRegression pour l'évaluation des performances prédictives [11], [167], [188], lme4 pour le HLMM [189], et timeROC pour la sensibilité et la spécificité (Blanche 2013).

8.3 Résultats

8.3.1 Description de la population

Les données individuelles des patients de 17 études ont été rassemblées dans la méta-analyse MAOV initié par le GCIG. Parmi les 13 essais comportant des mesures répétées du CA-125. (Figure 8.2), neuf essais ont utilisé une stratégie de traitement de fond à base de platine. Trois essais ont été exclus en raison de l'utilisation d'une chimiothérapie cytotoxique non basée sur le platine et un essai a été exclu en raison de la non-disponibilité des mesures du CA125 au cours des 3 mois suivant la randomisation. L'analyse secondaire ajustée sur les covariables de base a exclu un essai supplémentaire car les informations sur la maladie résiduelle après la chirurgie n'avaient pas été collectées. (Figure 8.2). Les neuf essais ont recruté un total de 6 627 patientes avec 90 730 mesures répétées du CA-125 dans le temps avec des valeurs entre 2 et 100 (table 8.1) et les caractéristiques des patientes par essai (Table 8.2).

Figure 8.2: Organigramme de la population éligible incluse dans cette étude.



Chez les 6253 patientes ayant au moins deux mesures du CA125 et ne présentant pas d'événement à 3 mois, 27193 mesures du CA-125 étaient disponibles. Parmi elles, 5638 patientes (25630 mesures) disposaient de données cliniques.

Les caractéristiques des essais et des patientes sont décrites respectivement dans le tableau 8.1 et dans le tableau 8.3 des matériaux complémentaires. Le statut FIGO et la maladie résiduelle par essai sont présentés dans le tableau 8.2. La médiane de la survie sans progression était de 18 mois [17,6-18,4]. Au total de 4271 progressions ont été enregistrées avec 1809, 2459 et 2951 progressions à 12, 15 et 18 mois, respectivement, et 2849 événements mortels ont été rapportés (voir tableau 8.3 et tableau 8.4).

Table 8.1: Description des études.

Essai	Nb de pts	# CA-125 measures	Médiane PFS(mois)		Survie sans Progression (Oui)
			Délai entre la randomisation et le premier évènement	Délai entre le temps de landmark time et le premier évènement	
TURBO-2014	240	1139	19.1 (17.1;22.1)	16.1(14.1;19.1)	203
JGOG-3017	303	1247	NA*	NA*	74
MITO-7	461	922	20.5(18.5;25.1)	17.5(15.5;22.1)	238
JGOG-3016	547	2139	28.7 (24;8;33.1)	25.7(21.8;30.1)	358
CCTG-OV.16	694	3009	16.6 (15.5;17.7)	13.6(12.5;14.7)	589
NSGO-2012	742	3052	17.7 (16.7;18.9)	14.7(13.7;15.9)	599
SCOTROC-4	850	3965	13.7 (13.0;14.9)	10.7(10.0;11.9)	586
SCOTROC-1	979	5021	16.6(15.4;18.1)	13.6(12.4;15.1)	598
MRC-ICON7	1437	6699	19.2(18.5;20.6)	16.2(15.5;17.6)	1026
Total	6253	27193	18 (17.6; 18.4)	15.6(15.2;16.2)	4271

*Abréviations: NA, non atteint.

Table 8.2 : Caractéristiques des patients par essai (N=6253)

Essai	Statut FIGO					Maladie résiduelle				Total
	I	II	III	IV	total	<1cm	>= 1cm	ND	.	
TURBO	0	19(8%)	156(65%)	65(27%)	240	173(73%)	63(26%)	2(1%)	2	238
JGOG-3017	202(67%)	37(12%)	55(18%)	9(3%)	303	277(91%)	25(8%)	1(0%)		303
MITO-7	41(9%)	42(9%)	273(59%)	105(23%)	461	ND	ND	ND	ND	ND
JGOG-3016	0	102(19%)	364(67%)	81(15%)	547	258(47%)	275(50%)	14(3%)		547
CCTG-OV.16	0	65(9%)	468(67%)	161(23%)	694	325(57%)	238(42%)	7(1%)	124	570
NSGO-2012	0	94(13%)	533(72%)	115(15%)	742	317(43%)	425(57%)			742
SCOTROC-4	118(14%)	75(9%)	549(65%)	108(13%)	850	300(35%)	550(65%)			850
SCOTROC-1	70(7%)	127(13%)	645(66%)	137(14%)	979	332(34%)	647(66%)			979
ICON-7	112(8%)	158(11%)	987(69%)	180(13%)	1437	1044(74%)	365(26%)		28	1409

TOTAL	565(9%)	736(11%)	4265(64%)	1061(16%)	6627	3121(5 3%)	2788(47%)	26(0%)	692	5935
-------	---------	----------	-----------	-----------	------	---------------	-----------	--------	-----	------

Abréviations : ' ' : Informations manquantes, ND : Non disponible

Tableau 8.3: Description des caractéristiques des patients pour la population totale (N=6627) et pour les ensembles de données d'analyse en fonction de la disponibilité des données au moment du repère (3 mois) et des cas complets de maladie résiduelle.

Variable		Total (9 essais)		Population en vie sans progression à 3 mois		Population en vie sans progression à 3 mois (cas complets maladie résiduelle)	
Age	n	6627		6253		5638	
	Moyenne (sd)	58.2 (10.8)		58.2 (10.7)		58.1 (10.7)	
	Médiane (Q1 ; Q3)	59 (51 ; 66)		58 (51 ; 66)		58 (51 ; 66)	
	Min ; Max	18 ; 88		18 ; 88		18 ; 88	
Performance status	n (m.d.)	6602 (25)		6228		5614 (24)	
	0	3106	47%	2970	47.7%	2555	45.5%
	1	2969	45%	2775	44.5%	2584	46%
	2+	527	8%	483	7.8%	475	8.5%
Figo	n (m.d.)	6627		6253		5638	
	Stage I	565	8.5%	543	8.7%	502	8.9%
	Stage II	736	11.1%	719	11.5%	671	11.9%
	Stage III	4265	64.4%	4030	64.4%	3681	65.3%
	Stage IV	1061	16%	961	15.4%	784	13.9%

Type	n (m.d.)	6622 (5)		6248 (5)		5633 (5)	
Histologique	Séreux	3750	56.6%	3596	57.6%	3169	56.3%
	Endométrioïde	608	9.2%	578	9.3%	527	9.4%
	Mucineux	181	2.7%	149	2.4%	139	2.5%
	Cellules Claires	690	10.4%	650	10.4%	621	11%
	Autre	1173	17.7%	1069	17.1%	976	17.3%
	Inconnu	220	3.3%	206	3.3%	201	3.6%
Grade histologique	n (m.d.)	5540 (1087)		5259 (994)		4751 (887)	
histologique	Bien différencié	499	9%	476	9.1%	453	9.5%
	Modérément différencié	1367	24.7%	1296	24.6%	1227	25.8%
	Faible/Non différencié	3674	66.3%	3487	66.3%	3071	64.6%
Procédure chirurgicale	n (m.d.)	6080 (547)		5776 (477)		5637(1)	
chirurgicale	Résection primaire complète	5411	89%	5144	89.1%	5079	90.1%
	Chirurgie retardée	161	2.6%	144	2.5%	144	2.6%
	Sans chirurgie	508	8.4%	488	8.4%	414	7.3%
Maladie résiduelle maximale	n (m.d.)	5935(692)		5638 (615)		5638	
résiduelle maximale	<1cm	3121	52.6%	3026	53.7%	3026	53.7%
	>=1cm	2788	47%	2588	45.9%	2588	45.9%
	Inconnu	26	0.4%	24	0.4%	24	0.4%

Évaluation de la progression	n	6627		6253		5638	
	CT-scan	318	4.8%	303	4.8%	303	5.4%
	CA-125-Clinique (confirmation avec scanner)	3603	54.4%	3385	54.1%	3356	59.5%
	Les deux	2706	40.8%	2565	41%	1979	35.1%
CA-125 at baseline	n	6627		6253		5638	
	Mean (sd)	643.98 (1481.366)		639.29 (1475.524)		625.57 (1461.589)	
	Median (Q1 ; Q3)	142.4 (44; 482)		140 (43; 473)		137 (43; 451)	
	Min ; Max	0.82; 14400		0.82; 14400		0.82; 14400	
log(CA-125 +1) à l'initiation	n	6627		6253		5549	
	Moyenne(sd)	5.08 (1.66)		5.07 (1.66)		5.05 (1.64)	
	Médiane (Q1 ; Q3)	4.97 (3.81; 6.18)		4.95 (3.78; 6.16)		4.93 (3.79; 6.11)	
	Min ; Max	0.60 ; 9.58		0.60; 9.58		0.60; 9.58	
nombre de mesures répétées du CA-125 par patient	n	90730		27193		25630	
	Moyenne (sd)	13.69 (8.7)		14.15(8.55)		14.89 (8.06)	
	Médiane (Q1 ; Q3)	12 (8 ; 18)		13 (9 ; 19)		13 (9 ; 19)	
	Min ; Max	2 ; 100		2 ; 80		2 ; 80	

Tableau 8.4: Description du critère d'évaluation principale

Variable		Total (9 trials)		Population en vie sans progression à 3 mois		Population en vie sans progression à 3 mois (cas complets maladie résiduelle)	
		6627		6253		5638	
Progression free survival	No	2040	30.8%	1982	31.7%	1753	31.1%
	Yes	4586	69.2%	4271	68.3%	3885	68.9%
Progression or censoring at 12 months	No	4411	66.5%	4335	69.3%	3939	69.9%
	Censoring before 12 months	131	2%	109	1.7%	84	1.5%
	Yes	2085	31.5%	1809	29%	1615	28.6%
Progression or censoring at 15 months	No	3707	55.9%	3642	58.2%	3334	59.1%
	Censoring before 15 months	177	2.7%	152	39.4%	105	39%
	Yes	2743	41.4%	2459	2.4%	2199	1.9%
Progression or censoring at 18 months	No	3084	46.5%	3028	48.4%	2801	49.7%
	Censoring before 18 months	301	4.5%	274	4.4%	193	3.4%
	Yes	3242	48.9%	2951	47.2%	2644	46.9%
Survie globale	Alive	3516	53.1%	3404	54.4%	2999	53.2%
	Dead	3110	46.9%	2849	45.6%	2639	46.8%

8.3.2. Performances prédictives des résumés du CA-125

Les performances des résumés du CA-125 pour prédire la résistance au platine à 12 mois sont représentées dans la figure 8.3 et reportées dans le tableau 8.5. La valeur observée du CA-125 la plus proche de 3 mois (OVLT) et la valeur prédite à 3 mois (EVL) ont eu la meilleure performance prédictive avec un AUC à 12 mois de 74,5 % (IC 95 % [70,7 ;78,6]) et 74,3 % (IC 95 % [70,3 ;78,4]), respectivement. L'ajout de la valeur du CA-125 au départ a amélioré les capacités de prédiction des mesures OVLT et EVLT de 0,04 [0,02 ; 0,07] points. L'analyse de sensibilité pour la prédiction à 15 et 18 mois a montré des performances similaires pour les deux résumés (Figure 8.3, Table 8.5).

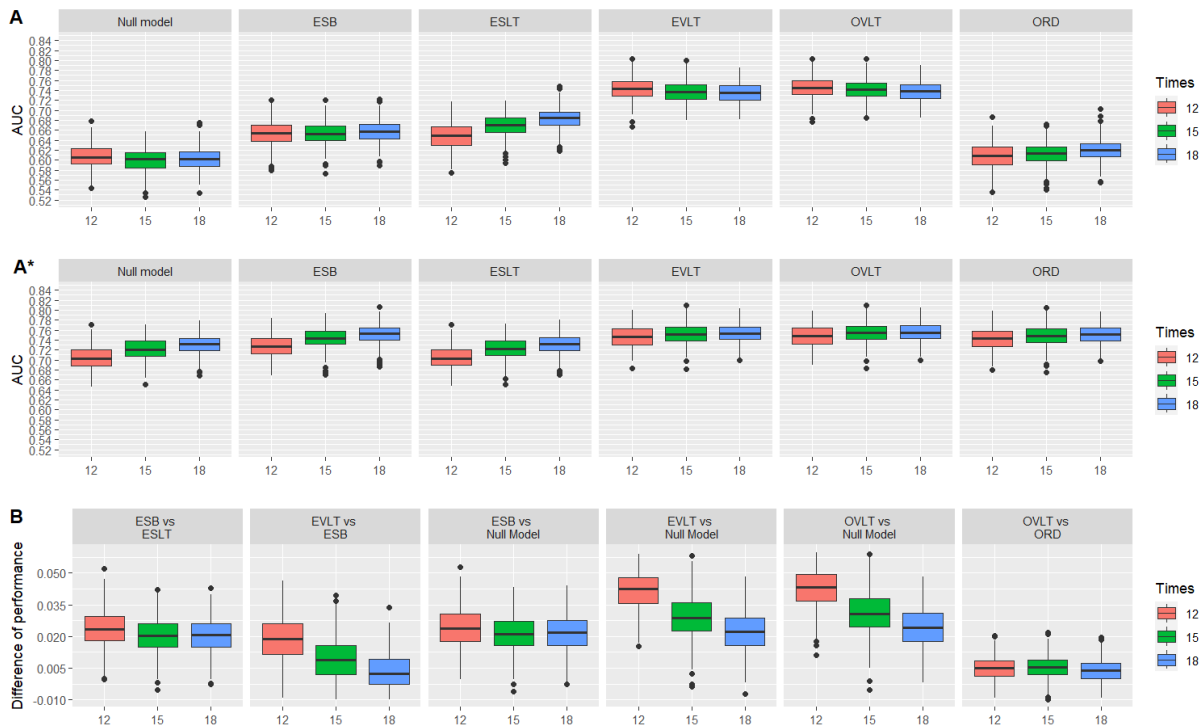


Figure 8.3 : Les AUC (A, A*) obtenus par 10-fold cross validation pour la prédiction de la progression à 12, 15 et 18 mois à partir de l'historique du CA-125 jusqu'à 3 mois, en considérant différents résumés du CA-125 (ORD, OVLT, EVLT, ESB, ESLT) ajustés (A*) ou non (A) pour la valeur du CA-125 à l'inclusion, et les différences d'AUC obtenues par 10-fold cross validation(B) entre les différents modèles. Modèle nul : modèle de survie stratifié sur l'étude, ORD : déclin relatif observé, OVLT : valeur observée au temps de landmark, EVLT : valeur estimée au temps de landmark, ESB : pente estimée à l'inclusion, ESLT : pente estimée au temps de landmark.

Null model: model de survie stratifié à l'essai, ORD: Décroissance relative Observée, OVLT: Valeur observée du CA-125 au temps de landmark, EVLT: Valeur estimée au temps de landmark, ESB: Pente estimée à l'initiation du traitement, ESLT: Pente estimée au temps de landmark

* tous les modèles sont ajustés au CA-125 à l'initiation du traitement

Table 8.5: 10-fold cross-validation AUC (A,A*) et Brier score (C,C*) pour la prédiction de la progression à 12, 15 and 18 mois à partir de l'historique du CA-125 jusqu'à 3 mois en considérant différents résumés du CA125 Point de repère 3 mois (Population en vie et n'ayant pas progressé à 3 mois N=6253)

	Time	T=12	T=15	T=18
A. AUC	Null model	0.606 [0.558 – 0.653]	0.600 [0.557 ; 0.643]	0.603 [0.562 ; 0.644]
	ORD	0.608 [0.557 – 0.656]	0.613 [0.569 ; 0.658]	0.620 [0.578 ; 0.663]
	OVLТ	0.745 [0.707 – 0.786]	0.741 [0.7 ; 0.778]	0.738 [0.701 ; 0.776]
	EVLT	0.743 [0.703 – 0.784]	0.737 [0.697 ; 0.774]	0.734 [0.698 ; 0.774]
	ESB	0.654 [0.606 – 0.704]	0.653 [0.607 ; 0.695]	0.657 [0.614 ; 0.699]
	ESLT	0.649 [0.600- 0.696]	0.669 [0.622 ; 0.710]	0.683 [0.638 ; 0.724]
A* AUC*	Null model	0.704 [0.659 – 0.746]	0.722 [0.681 ; 0.758]	0.731 [0.691 ; 0.768]
	ORD	0.743 [0.702 – 0.783]	0.748 [0.707 ; 0.784]	0.751 [0.715 ; 0.787]
	OVLТ	0.748 [0.707 – 0.789]	0.753 [0.714 ; 0.788]	0.755 [0.719 ; 0.790]
	EVLT	0.747 [0.705 – 0.787]	0.751 [0.712; 0.786]	0.753 [0.717 ; 0.789]
	ESB	0.728 [0.683 – 0.766]	0.743 [0.701 ; 0.778]	0.752 [0.713 ; 0.787]
	ESLT	0.704 [0.660 – 0.745]	0.723 [0.682 ; 0.759]	0.732 [0.693 ; 0.768]
B Difference *	ESB vs ESLT	0.0237 [0.001 – 0.04]	0.0203 [0.004-0.04]	0.0205 [0.004-0.04]
	EVLT vs ESB	0.0191[-0.001-0.04]	0.0080 [-0.01 - 0.03]	0.0008[-0.02 - 0.02]
	ESB vs Null Model	0.024 [0.006 – 0.04]	0.021 [0.004-0.04]	0.022 [0.004-0.04]
	EVLT vs Null Model	0.0430 [0.02 – 0.07]	0.0292 [0.01-0.05]	0.0224 [0.005-0.04]
	OVLТ vs Null Model	0.044 [0.02 – 0.07]	0.0311 [0.01-0.05]	0.024 [0.006-0.04]
	OVLТ vs ORD	0.005 [-0.007 – 0.015]	0.005 [-0.006 - 0.02]	0.004[-0.007 - 0.01]
C Brier	Null Model	0.201 [0.187 – 0.216]	0.232 [0.222 ; 0.241]	0.24 [0.233 ; 0.247]
	ORD	0.201 [0.187 – 0.217]	0.231 [0.221 ; 0.241]	0.238 [0.231 ; 0.246]

	OVLT	0.175 [0.160- 0.191]	0.199 [0.185 ; 0.213]	0.207 [0.196 ; 0.218]
	EVLT	0.175 [0.160- 0.191]	0.199 [0.186 ; 0.213]	0.208 [0.196 ; 0.219]
	ESB	0.194 [0.178 - 0.210]	0.223 [0.213 ; 0.235]	0.230 [0.221 ; 0.24]
	ESLT	0.196 [0.180 - 0.212]	0.221 [0.21 ; 0.233]	0.227 [0.217 ; 0.237]
C* Brier*	Null model	0.184 [0.170 - 0.2]	0.206 [0.193 ; 0.219]	0.211 [0.198 ; 0.224]
	ORD	0.175 [0.161 - 0.192]	0.197 [0.184 ; 0.21]	0.203 [0.191 ; 0.215]
	OVLT	0.175 [0.161 - 0.190]	0.196 [0.182 ; 0.209]	0.202 [0.191 ; 0.215]
	EVLT	0.175 [0.161 - 0.191]	0.196 [0.183 ; 0.208]	0.202 [0.191 ; 0.215]
	ESB	0.181 [0.165 - 0.197]	0.200 [0.188 ; 0.215]	0.204 [0.193 ; 0.218]
	ESLT	0.185 [0.170 - 0.201]	0.206 [0.193 ; 0.219]	0.21 [0.198 ; 0.224]
D. Difference *	ESB vs ESLT	-0.0039[-0.007 - -0.0001]	-0.0051 [-0.009 - -0.001]	-0.0060 [-0.01 - -0.001]
	EVLT vs ESB	0.0060 [0.0001-0.01]	0.0046 [-0.002 - 0.01]	0.0020 [-0.004 - 0.008]
	ESB vs Null Model	0.0038[-0.0001-0.007]	0.0052 [0.0002-0.01]	0.0062 [0.001-0.01]
	EVLT vs Null Model	0.0097 [0.003-0.01]	0.0098 [0.003 -0.02]	0.0082 [0.002 -0.01]
	OVLT vs Null Model	0.0095[0.003-0.02]	0.0099 [0.003-0.02]	0.0084 [0.001-0.01]
	OVLT vs ORD	0.0004[-0.002-0.003]	0.0012 [-0.003 - 0.004]	0.0008 [-0.002 - 0.004]

La baisse du CA-125 prédite soit à 3 mois (ESLT) soit juste après le début du traitement (ESB) avait des performances prédictives plus faibles (AUC=64,9 % (IC 95 % [60;69,6]) et 65,4 % (IC 95 % [60,6;70,4]), respectivement) que la valeur observée ou attendue à 3 mois. Le taux de décroissance a dû être combiné avec le CA-125 au départ dans le modèle de prédiction pour atteindre une valeur discriminatoire intéressante avec une AUC de 70,4 % (IC 95 % [66;74,5]) et 72,8 % (IC 95 % [68,3;76,6]), respectivement. La pente prédite au départ (ESB) est plus performante qu'à 3 mois (ESLT) avec une augmentation de 2,4 points de l'AUC (IC 95 % [0,1 ; 4,0]). L'élargissement de l'horizon de prédiction à 15 et 18 mois a amélioré la capacité prédictive de la pente estimée au départ (ESB) de 74,3 % (IC 95 % [70,1 ;77,8]) et de 75,2 % (IC 95 % [71,3;78,7]) à 15 et 18 mois, respectivement.

Le taux de déclin observé (ORD) seul a donné de mauvais résultats avec une AUC de 60,8 % (IC à 95 % [55,7 ; 65,6]) à 12 mois par rapport à tous les autres résumés. Cependant, lorsqu'il est combiné avec la valeur initiale du CA-125, il atteint presque la même performance que les

valeurs observées et prédites à 3 mois (également combinées avec le CA-125 initial) (différence d'AUC = 0,5 (IC à 95 % [-0,7;1,5]).

Le score de Brier a conduit aux mêmes conclusions (figure 8.4. et tableau 8.5)

8.3.3 Performance des résumés du CA-125 combinée avec le stade FIGO et la maladie résiduelle.

Nous avons étudié la combinaison des facteurs cliniques et biologiques chez les 5638 patientes présentant des caractéristiques cliniques documentées ; il est à noter que dans ce sous-ensemble, nous avons obtenu la même performance du CA-125 seul que celle que nous avons rapportée pour l'ensemble des patientes. Un ajustement supplémentaire sur la maladie résiduelle et le stade FIGO dans le modèle de prédiction a conduit à une amélioration significative de la capacité de prédiction pour toutes les mesures à des degrés divers (figure 8.4, table 8.6). Plus précisément, le modèle nul (caractéristiques des patientes et CA-125 au départ) présentait une performance prédictive de 73,9 % (IC à 95 % [69,9 ; 77,9]). L'ajout de la valeur observée du CA-125 à 3 mois (OVLT) ou de la valeur prédite à 3 mois (EVLT) a amélioré la performance prédictive avec une augmentation de 3 points et une AUC de 77,3 % (IC 95 % [73,2 ; 81,1]) à 12 mois et jusqu'à 78,1 % (IC 95 % [74,4 ; 81,9]) à 18 mois. Tous les autres résumés présentaient une amélioration lorsque l'on ajoutait les caractéristiques des patients et le CA-125 à l'initiation du traitement, mais leur performance restait légèrement inférieure à celle du OVLT et du EVLT (figures 8.4 ; 8.5, table 8.6).

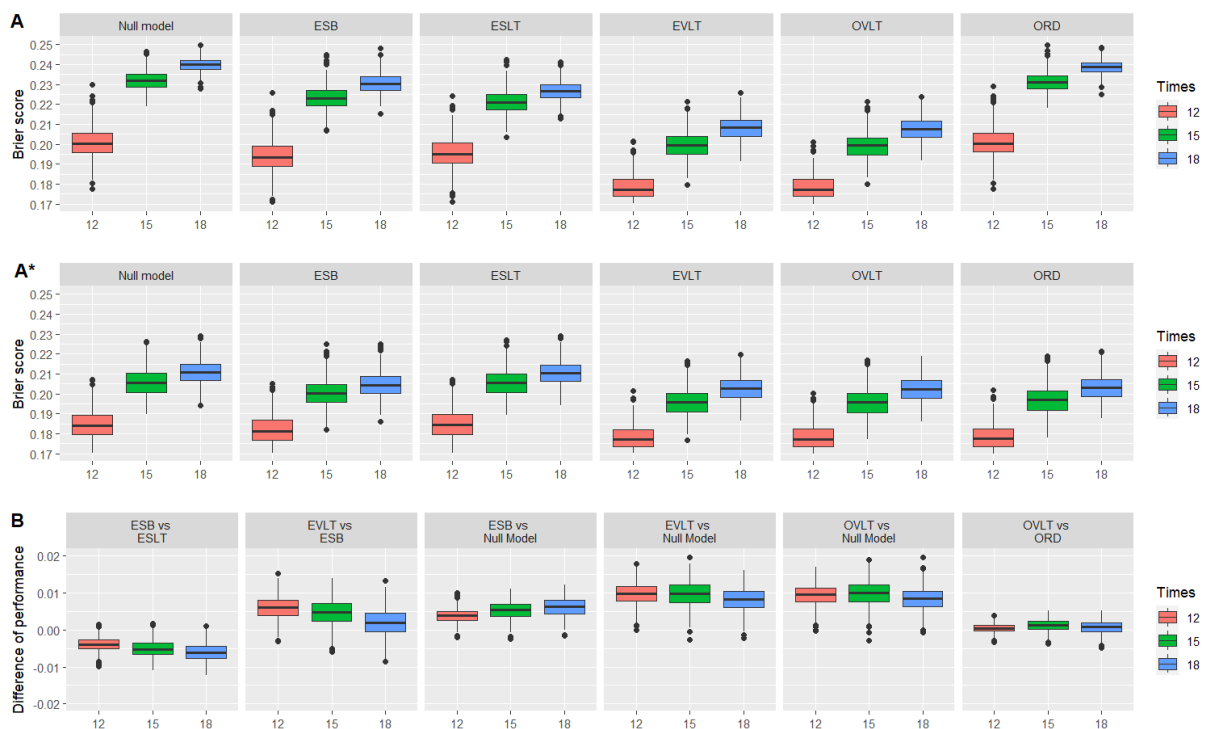


Figure 8.4 : Les Brier (A, A*) obtenus par 10-fold cross validation pour la prédiction de la progression à 12, 15 et 18 mois à partir de l'historique du CA-125 jusqu'à 3 mois, en considérant différents résumés du CA-125 (ORD, OVLT, EVLT, ESB, ESLT) ajustés (A*) ou non (A) pour la valeur du CA-125 à l'inclusion, et les différences d'AUC obtenues par 10-fold cross validation(B) entre les différents modèles. Modèle nul : modèle de survie stratifié sur l'étude, ORD : déclin relatif observé, OVLT : valeur observée au temps de landmark, EVLT : valeur estimée au temps de landmark, ESB : pente estimée à l'inclusion, ESLT : pente estimée au temps de landmark

* tous les modèles sont ajustés au CA-125 à l'initiation du traitement

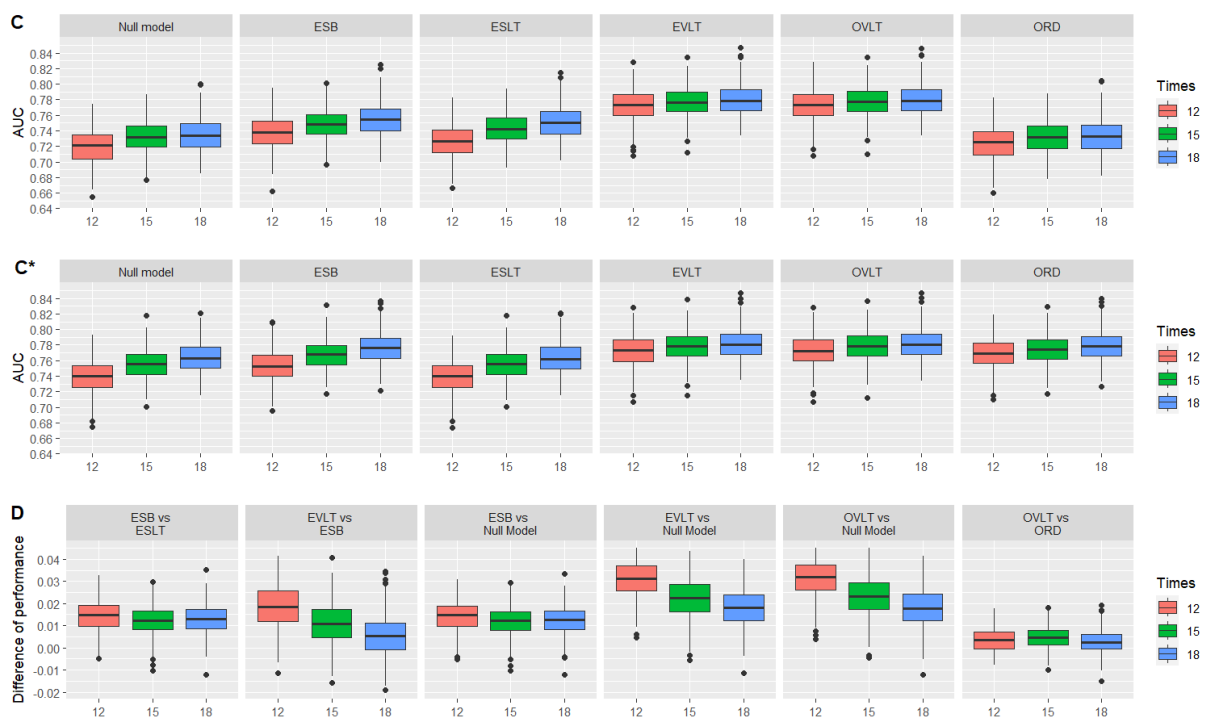


Figure 8.5 : AUCs (C, C*) obtenus par 10-fold cross validation pour la prédiction de la progression à 12, 15 et 18 mois à partir de l'historique du CA-125 jusqu'à 3 mois pour la population restreinte N=5638, en considérant différents résumés du CA-125 (ORD, OVLT, EVLT, ESB, ESLT) ajustés pour le stade FIGO et la maladie résiduelle et ajustés (C*) ou non (C) pour la valeur du CA-125 à l'inclusion, et différences de AUCs obtenues par 10-fold cross validation (B) entre les différents modèles. Modèle nul : modèle de survie stratifié sur l'étude, le stade FIGO et la maladie résiduelle. ORD : déclin relatif observé, OVLT : valeur observée au temps de landmark, EVLT : valeur estimée au temps de landmark, ESB : pente estimée à l'inclusion, ESLT : pente estimée au temps de landmark.

Table 8.6: 10-fold validation croisée AUC (A,A*) & Brier score (C,C*) pour la prédiction de la progression à 12, 15 and 18 mois à partir de 3 mois d'historique du CA-125 en considérant différents résumés du CA125 Point de repère 3 mois Population en vie et n'ayant pas progressé

à 3 mois . Population en vie et sans progression à 3 mois (cas complets maladie résiduelle) Cox avec figo & maladie résiduelle N=5638)

	Time	T=12	T=15	T=18
A. AUC	Null model	0.719 [0.677 ; 0.762]	0.732 [0.692 ; 0.769]	0.734 [0.696 ; 0.774]
	ORD	0.724 [0.683 ; 0.767]	0.732 [0.691 ; 0.769]	0.733 [0.695 ; 0.772]
	OVLТ	0.773 [0.733 ; 0.810]	0.777 [0.740 ; 0.812]	0.779 [0.743 ; 0.818]
	EVLT	0.773 [0.732 ; 0.810]	0.777 [0.741 ; 0.812]	0.779 [0.744 ; 0.817]
	ESB	0.738 [0.701 ; 0.779]	0.748 [0.709 ; 0.786]	0.754 [0.717 ; 0.793]
	ESLT	0.726[0.685 ; 0.768]	0.743[0.704 ; 0.780]	0.750[0.713 ; 0.789]
A* . AUC	Null model*	0.739 [0.699 ; 0.779]	0.755 [0.720 ; 0.791]	0.763 [0.727 ; 0.803]
	ORD*	0.769 [0.729 ; 0.807]	0.774 [0.738 ; 0.808]	0.779 [0.747 ; 0.816]
	OVLТ*	0.773 [0.733 ; 0.811]	0.779 [0.741 ; 0.812]	0.781 [0.744 ; 0.820]
	EVLT*	0.773 [0.732 ; 0.811]	0.778 [0.742 ; 0.811]	0.781 [0.746 ; 0.819]
	ESB*	0.754 [0.715 ; 0.795]	0.767 [0.733 ; 0.801]	0.776 [0.740 ; 0.812]
	ESLT*	0.739 [0.699 ; 0.779]	0.755 [0.720 ; 0.791]	0.763 [0.726 ; 0.803]
B. Differe nce	ESB vs ESLT	0.01 [0.0001 – 0.03]	0.01 [-0.001 – 0.03]	0.01 [-0.0005 – 0.03]
	EVLT vs ESB	0.02 [0.001- 0.04]	0.01 [-0.006 – 0.03]	0.005 [-0.01 – 0.02]
	ESB vs Null Model	0.01 [0.0002 -0.03]	0.01 [-0.0002 – 0.02]	0.01 [-0.0006 – 0.02]
	EVLT vs Null Model	0.03[0.01 – 0.05]	0.02 [0.004 – 0.04]	0.02 [0.001 – 0.03]
	OVLТ vs Null Model	0.03[0.02 – 0.05]	0.02 [0.005 – 0.04]	0.02 [0.002 – 0.03]
	OVLТ vs ORD	0.003[-0.005 – 0.01]	0.004 [-0.005 – 0.01]	0.003 [-0.007 – 0.01]
C. Brier	Null Model	0.180 [0.165 ; 0.196]	0.202 [0.189 ; 0.216]	0.207 [0.192 ; 0.220]
	ORD	0.179 [0.164 ; 0.195]	0.202 [0.189 ; 0.216]	0.207 [0.193 ; 0.220]
	OVLТ	0.166 [0.151 ; 0.182]	0.186 [0.173 ; 0.200]	0.191 [0.177 ; 0.205]
	EVLT	0.166 [0.151 ; 0.82]	0.186 [0.173 ; 0.200]	0.192 [0.177 ; 0.205]
	ESB	0.176 [0.161 ; 0.191]	0.197 [0.184 ; 0.211]	0.201 [0.187 ; 0.215]
	ESLT	0.180[0.165 ; 0.195]	0.200 [0.187 ; 0.214]	0.203 [0.190 ; 0.217]
C* . Brier	Null model*	0.176 [0.161; 0.191]	0.194 [0.181 ; 0.207]	0.198 [0.183 ; 0.212]
	ORD*	0.167 [0.152 ; 0.182]	0.187 [0.174 ; 0.200]	0.192 [0.177 ; 0.205]
	OVLТ*	0.167 [0.151 ; 0.183]	0.185 [0.172 ; 0.199]	0.191 [0.176 ; 0.205]
	EVLT*	0.167 [0.151 ; 0.182]	0.186 [0.172 ; 0.200]	0.191 [0.176 ; 0.204]
	ESB*	0.173 [0.158 ; 0.188]	0.191 [0.177 ; 0.204]	0.194 [0.180 ; 0.207]
	ESLT*	0.176 [0.161 ; 0.191]	0.194 [0.181 ; 0.207]	0.198 [0.184 ; 0.212]
D. Differe nce	ESB vs ESLT	0.003 [-0.0009 – 0.006]	0.004 [-0.0007 – 0.008]	0.005 [0.00004 – 0.009]
	EVLT vs ESB	0.02 [-0.0006- 0.01]	0.005 [-0.002 – 0.01]	0.003 [-0.004 – 0.01]
	ESB vs Null Model	0.003 [-0.001 – 0.006]	0.004 [-0.0007 – 0.008]	0.004 [-0.00003 – 0.009]
	EVLT vs Null Model	0.009 [0.003-0.01]	0.009 [0.002 – 0.01]	0.007 [0.0009 – 0.01]
	OVLТ vs Null Model	0.009 [0.003- 0.01]	0.009 [0.002 – 0.02]	0.007 [0.0009 – 0.01]

	OVLT vs ORD	0.0004[-0.002 – 0.003]	–	0.001 [-0.002 – 0.005]	–	0.0009 [-0.003 – 0.004]
--	-------------	------------------------	---	------------------------	---	-------------------------

Lors de l'étude des performances des différents seuils du CA-125, pour faciliter l'interprétation, nous fournissons les valeurs prédictives positives et négatives lorsque le CA125 est inférieur à 35 U/mL. Par exemple, à 3 mois, 81 % des patientes dont la valeur du CA-125 est inférieure à 35 sont susceptibles d'être sensibles au platine, tandis que 56 % des patientes dont la valeur du CA-125 est supérieure à 35 sont résistantes au platine. Les performances des autres seuils du biomarqueur sont présentées dans le tableau 8.7 ci-dessous.

Tableau 8.7: Positive predictive value (PPV) and negative predictive value (NPV) for different thresholds of the CA-125 biomarker for prediction horizons of 12, 15 and 18 months. Population N=5638

Cut point CA-125	PPV	NPV	PPV	NPV	PPV	NPV
	T=12		T=15		T=18	
15	44.5%	86.2%	57.7%	78%	66.1%	70.4%
35	55.8%	80.6%	69.4%	71%	77.4%	62.9%
70	59.6%	76.9%	73.3%	66.8%	80.3%	58.6%
80	60.2%	76.4%	74.5%	66.4%	81.4%	58.2%
90	61.2%	76.1%	75%	66%	82.1%	57.8%
100	62.3%	75.9%	75.2%	65.5%	82.1%	57.3%
120	63%	75.3%	75.2%	64.8%	81.8%	56.6%
140	64.4%	74.9%	77%	64.5%	83.6%	56.3%

Si l'on tient compte des caractéristiques des patients, environ 80 % d'eux ont un statut FIGO 3 et 4. Plus précisément, 77,1 % des patientes ayant un statut FIGO 3 et une valeur de CA-125 inférieure à 35 sont susceptibles d'être sensibles au platine, tandis que 55,9 % des patientes ayant une valeur de CA-125 >35 seront résistantes au platine à 12 mois. 68,4 % des patientes ayant un stade FIGO 3 et une maladie résiduelle ≥ 1 et une valeur de CA-125 inférieure à 35 sont susceptibles d'être sensibles au platine, tandis que 59,2 % des patients ayant une valeur de CA-125 >35 seront résistantes au platine à 12 mois. 68,4 % des patients ayant un stade FIGO 4 et une valeur de CA-125 inférieure à 35 sont susceptibles d'être sensibles au platine, tandis que 57,8 % des patients ayant une valeur de CA-125 >35 seront résistantes au platine à 12 mois.

8.4 Discussion

Les résultats de cette étude confirment la valeur ajoutée de l'évaluation du CA-125 à 3 mois pour prédire raisonnablement bien les patients résistants au platine. La dernière valeur observée du CA-125 à 3 mois a amélioré la capacité prédictive par rapport au CA-125 à l'inclusion avec une capacité discriminatoire de 74,5% (IC 95% [70,7% ; 78,6%]) pour la progression à 12 mois. La valeur estimée sans erreur à 3 mois a montré la même performance (74,3 % (IC 95 % [70,3 % ; 78,4 %])). En ajoutant le statut FIGO et la maladie résiduelle après la chirurgie, la capacité prédictive avec la valeur du CA-125 à 3 mois a légèrement augmenté à 77,3 % (IC 95 % [73,3 ; 81,1]) pour prédire la progression à 12 mois. La capacité prédictive du déclin relatif observé, de la pente prédite à l'initiation de l'étude et de la pente à 3 mois étaient médiocres avec un AUC de 60,8 % à 65,4 % lorsqu'elles étaient considérées seules à 12 mois. En ajoutant des informations supplémentaires, telles que le CA-125 à l'inclusion et les caractéristiques de la patiente, leur performance a été améliorée et elle a atteint la même capacité de discrimination que les résumés basés sur le niveau du CA-125 à 3 mois (AUC de 73,9 % jusqu'à 76,9 % à 12 mois). Quel que soit le résumé, une hétérogénéité significative de la valeur discriminante a été détectée entre les essais. Cependant, aucune de ces valeurs n'a atteint 80%, ce qui est souvent considéré comme un indicateur d'un facteur pronostique "fort". [141]

Notre étude précédente a conclu que la surveillance du CA-125 à 3 mois après le début du traitement montrait de bonnes capacités de prédiction avec une valeur d'AUC à 0,75 [0,70-0,80] en termes de survie globale jusqu'à 48 mois après la randomisation. [182] Récemment, le score KELIM sur les données de la même méta-analyse et en partie dans la même population, a rapporté une capacité prédictive plus faible avec un C-index (AUC moyenne sur le suivi complet) de 0,62 (95% CI, 0,60-0,64) pour la survie globale, et 0,59 pour la survie sans progression (95% CI, 0,57-0,60).[183] Le score KELIM en association avec les inhibiteurs de PARP a montré des résultats prometteurs en termes de valeur prédictive chez les patients traités par chimiothérapie à base de platine. [88]

Cette étude, qui est basée sur un cadre méta-analytique sur données individuelles, s'appuie sur une méthodologie rigoureuse qui gère des données hétérogènes et montre de bonnes indications pour alerter d'une éventuelle réorientation des patientes. Plus précisément, les patientes présentant une augmentation de la valeur du CA-125 deux fois supérieure à la valeur normale devraient bénéficier d'une augmentation de leur surveillance tout en préparant un traitement de seconde ligne. D'autre part, après trois mois de traitement, 81% des patientes avec une valeur de CA-125 inférieure à 35 seront sensibles au platine et pourront donc continuer à bénéficier du traitement initial et du plan de surveillance actuel.

Une limite notable de notre méta-analyse est la période de recrutement qui était antérieure à 2015. Par conséquent, cette étude n'a pu étudier ni la valeur pronostique du CA-125 chez les patientes ayant reçu des inhibiteurs de PARP, ni au sous-groupe des patients avec mutations BRCA positif, qui n'a pas été recueilli dans les études incluses [58].

Récemment, un essai randomisé de phase III évaluant une nouvelle thérapie anticancéreuse basée sur des gènes (ofranergene - obadenovec) en plus du paclitaxel hebdomadaire chez des patientes atteintes d'un cancer de l'ovaire "résistant au platine" a conclu que le CA-125 était

un biomarqueur pronostique important pour la survie sans progression et la survie globale [190].

Dans un contexte de surveillance, l'utilisation du CA-125 dans une approche en deux étapes pendant le dépistage et le suivi des patients, qui suit les changements du CA-125 et intensifie la surveillance, a montré des résultats intéressants pour la détection de la maladie à un stade précoce. [191]

Des études récentes ont montré un grand intérêt dans l'association du CA-125 avec la protéine de l'épididyme humain (HE4). Par exemple, l'HE4 et le CA-125 ont présenté un AUC de 0,81 et 0,74, respectivement, pour prédire les patients sensibles et résistants au platine [192]. Cette étude a conclu pour les deux une bonne capacité pronostique au troisième cycle de chimiothérapie. Des résultats pronostiques prometteurs ont été démontrés en essayant d'identifier les seuils de CA-125 et HE4 et leur association avec la mortalité par cancer avec des AUC de 63,1 (95% CI : 53,6-72,6) et 64,2 (95% CI : 54,7-73,6) respectivement. [193]. En combinant le CA-125 et le HE4, Rong et al 2021 ont rapporté un AUC de 0,788 avec une sensibilité de 100 % et une spécificité de 57,5 %.[194] Cependant, une analyse plus approfondie devrait être réalisée pour éclairer le rôle de la protéine HE4 en association avec le CA-125 en considérant un plus grand nombre de patientes. En outre, Samborski et al ont récemment tenté d'identifier des biomarqueurs ayant des capacités pronostiques potentielles, tels que la protéine humaine de l'épididyme de type 4 (HE4), afin d'améliorer le modèle de prédiction et d'aider à la compréhension de la maladie [195] Malheureusement, cette information n'est pas collectée dans les essais que nous avons inclus dans cette méta-analyse.

Pour faciliter l'interprétation médicale, la définition d'un point limite de la trajectoire du biomarqueur pourrait faciliter la prise de décision. Cependant, nous avons observé la difficulté d'identifier un seuil du CA-125 permettant de maximiser à la fois la spécificité et la sensibilité. Même si nous avons essayé d'élargir ce seuil, le seuil optimal qui maximise à la fois la VPP et la VPN n'a pas été identifié. Ce comportement est également observé dans plusieurs études récentes. [191], [196], [197]. Cela est probablement dû au fait que le cancer de l'ovaire est caractérisé par une grande hétérogénéité et un mécanisme d'action complexe [34].

D'autres approches basées sur des modèles, telles qu'une extension de l'approche de landmark, une modélisation conjointe ou des méthodes prenant en compte les risques concurrents, pourraient être des méthodes appropriées pour modéliser les caractéristiques multiples des biomarqueurs dans des études qui s'intéressent au temps jusqu'à l'événement et pourraient aider à mieux comprendre cette attitude. [185]

En conclusion, le CA-125 est un élément important dans la surveillance des patientes atteintes d'un cancer de l'ovaire. D'autres analyses avec une mise à jour de la méta-analyse incluant des études testant de nouveaux agents et de nouvelles thérapies devraient être réalisées.

Le chapitre suivant présente l'évaluation du score KELIM sur une partie d'essais de la méta-analyse MAOV.

9. EVALUATION DU SCORE KELIM SUR LA META-ANALYSE MAOV

Le score KELIM a été évalué sur la méta-analyse MAOV et en collaboration avec l'équipe de recherche du Centre pour l'Innovation en Cancérologie de Lyon (CICLY). Cette collaboration a conduit à la publication d'un article : « Survie et constante K du taux d'ELIMination de l'antigène 125 modélisé chez les patientes atteintes d'un cancer de l'ovaire en première ligne avant l'ère des inhibiteurs de la poly (ADP-ribose) polymérase : Une méta-analyse du Gynécologique Cancer Intergroup» [183]. J'ai contribué à la préparation et standardisation des données de ce travail qui a été mené par Mme Corbeaux et le Pr Benoît You : il est présent ci-dessous.

9.1 Introduction

Dans plusieurs études, KELIM a été identifié comme un indicateur de la chimio sensibilité intrinsèque de la tumeur [61], [84], [198].: Dans de nombreuses études, KELIM a été identifié comme un indicateur fiable de la chimio sensibilité intrinsèque de la tumeur. En effet, il s'est avéré qu'il était associé à : (1) la réponse radiologique pendant la chimiothérapie néo-adjuvante ; (2) la probabilité d'une résection complète lors de la chirurgie d'intervalle (3) la probabilité d'une rechute ultérieure résistante au platine ; (4) la survie sans progression (PFS) et la survie globale (OS) du patient ; ainsi que (5) la probabilité de rester longtemps sans maladie > 5 ans après le traitement de première intention , dans sept ensembles de données d'essais indépendants et un registre national, impliquant au total plus de 7000 patients. [8], [82], [83], [85], [165], [199]. Des valeurs élevées du KELIM indiquent une élimination rapide du CA-125, qui reflèteraient une chimio-sensibilité plus grande. L'hypothèse est que la constante d'élimination à 100 jours est un facteur pronostique de la survie globale et de la survie sans progression.

Une méta-analyse s'imposait pour quantifier la valeur pronostique du score KELIM. Les objectifs de la présente étude réalisée avec MAOV étaient d'évaluer: (1) la valeur pronostique du KELIM pour la survie globale et la survie sans progression (2) l'impact de la chimio-sensibilité de la tumeur en ce qui concerne l'optimalité de la chirurgie de désobstruction par rapport au succès du traitement médico-chirurgical; (3) l'impact de la chimio sensibilité de la tumeur en ce qui concerne l'optimalité de la chirurgie d'exérèse par rapport à la réussite du traitement médico-chirurgical.

Cette étude a évalué le KELIM en tant que critère de substitution de la survie globale. Cependant, les questions soulevées sont trop nombreuses et cette dernière nécessite d'en faire une analyse plus approfondie, raison pour laquelle elle ne sera pas présentée.

9.2. Matériel et méthodes

Au sein de la méta-analyse MAOV, [134] toutes les patientes pour lesquelles des mesures individuelles du taux sérique de CA-125 étaient disponibles ont été incluses si au moins trois mesures étaient disponibles au cours des 100 premiers jours de traitement. Par tirage au sort avec stratification sur l'essai, les patients ont été répartis de manière aléatoire avec un rapport de 1 : 2 compte tenu de la taille importante de la base de données.[200]

9.2.1. Estimation de la cinétique individuelle KELIM

La cinétique longitudinale individuelle du CA-125 a été caractérisée à l'aide d'une approche basée sur la population, avec un modèle non linéaire à effet mixte et effet aléatoire pour

expliquer la variabilité interindividuelle. Le modèle structural qui décrit la cinétique du CA-125 est détaillé en section 3.8

L'ensemble de données d'apprentissage a été utilisé pour estimer les paramètres du modèle ; Le modèle a été stratifié à l'essai pour le KELIM, présenté en section 3. 8, et au niveau du CA-125 initial, afin de tenir en compte des études multiples. Les paramètres de population estimés sur la base d'apprentissage à partir des mesures du CA-125 des 100 jours premières jours ont été utilisés pour estimer les valeurs individuelles du KELIM dans la base de validation avec une méthode bayésienne[201]

9.2.2. Valeur pronostique du KELIM

Les associations du score KELIM avec la survie globale et la survie sans progression ont été estimées par des analyses de survie univariées puis par des analyses multivariées ajustées sur les principaux facteurs pronostiques mesurés. En particulier, la qualité de la chirurgie, telle que mesurée par la taille des résidus tumoraux, le stade FIGO ont été introduit dans le modèle. Une approche méta-analytique en deux étapes a été appliquée qui a conduit à d'une part estimer pour chaque étude l'hazard ratio quantifiant l'association entre KELIM et temps avant événement, puis à combiner les différents hazard ratios sous l'hypothèse d'indépendance [202]. La statistique I^2 a été utilisée pour évaluer l'hétérogénéité entre les essais [137]. La capacité de discrimination a été évaluée à l'aide de l'indice C de Harrell. A noter, c'est l'indice C sur toute la durée du suivi qui a été calculé, ce qui peut induire des biais, [139] notamment en présence de suivis différents selon les études ; l'AUC(t) aurait été recommandée. Pour faciliter l'interprétation clinique des analyses de survie, KELIM a été standardisé par la valeur médiane de l'ensemble des données, puis dichotomisé en un score KELIM : std KELIM < 1,0 a été considéré comme défavorable, tandis que std KELIM \geq 1,0 a été considéré comme favorable. Ce seuil a été précédemment considéré comme optimal [83], [85], [165]. Les analyses ont été réalisées avec un temps de landmark fixé à 100 jours après la randomisation, en excluant les patientes décédées ou ayant progressé au cours de cette période, afin d'éviter les biais potentiels liés aux événements précoces et à la cinétique du CA-125 [72].

Une approche méta-analytique en deux étapes a été appliquée pour évaluer la valeur de substitution de KELIM[94]. Cette partie ne sera pas présentée puisque ce ne sont pas des méthodes que nous avons étudiées dans nos travaux de recherche.

9.3 Résultats

9.3.1. Sélection des essais et des patients

Sur les 11 029 patientes issues des 17 essais inclus dans la base de données MAOV, 5884 patientes de huit essais randomisés étaient éligibles pour l'estimation du KELIM [26-33]. Parmi elles, 1962 et 3922 patientes ont été réparties dans les cohortes d'apprentissage et de validation, respectivement. Pour le calcul du score postopératoire était disponible pour 1912 (97,5 %) et 3802 (97,0 %) patientes, respectivement. Dans les deux cohortes, le KELIM médian était de 0,06 par jour avec pour intervalle interquartile (0,04-0,08). Par conséquent, le KELIM standardisé a été calculé comme étant le KELIM individuel/0,06.

Pour l'analyse de la survie globale, 1948 et 3894 patientes ont été incluses dans les ensembles d'apprentissage et de validation, respectivement. Pour la survie sans progression, 1888 et 3802 patientes ont été incluses dans ces deux ensembles.

9.3.2. Valeur pronostique

Survie globale

Les patients ayant un score KELIM favorable ($\geq 1,0$) ont connu une survie globale significativement plus longue que ceux ayant un score défavorable ($< 1,0$) : Sur l'échantillon d'apprentissage la survie globale médiane était de 81.8 mois (IC à 95 %, 68.7- non atteint (NA)) vs 31.1 mois (IC 95 %, 29.2-35.7) pour les sous-groupes favorable et défavorable respectivement. Sur l'échantillon de validation, la survie globale médiane était de 78.8 mois (IC 95 %, 72.9-89.1) contre 28.4 mois (IC 95 %, 26.7-30.8) ; avec un log-rank $P < 0,0001$.

Dans les analyses univariées de la survie globale, le C-index associé au KELIM était de 0,60 (IC à 95 %, 0,58-0,62) et de 0,62 (IC à 95 %, 0,60-0,64) dans les deux échantillons d'apprentissage et de validation, respectivement. Cela indique que dans 40% des cas, la survie de deux femmes prises au hasard sera plus longue pour la femme avec le moins bon pronostique.

Dans les modèles multivariés, ajustés sur le stade FIGO et la maladie résiduelle, un score KELIM favorable était indépendamment associé à une réduction de moitié du risque de décès (échantillon d'apprentissage, HR=0,51, [95% CI, 0,44-0,59] ; échantillon de validation, HR=0,46, [95% CI, 0,41-0,52]). L'inclusion des trois covariables, (FIGO, Résiduel de la maladie et KELIM) a amélioré l'indice C à 0,67 (IC à 95 %, 0,64-0,70) et à 0,68 (IC à 95 %, 0,65-0,71) dans les échantillons d'apprentissage et de validation, respectivement.

Lorsqu'on ajuste en fonction de la maladie résiduelle suite à la chirurgie, trois populations de pronostics différents ont été distinguées: (1) une population de bon pronostic, avec un score KELIM favorable et une chirurgie optimale (cohorte de validation, survie globale médiane, 105.1 mois ; IC à 95 %, 92,5-NA ; (2) une population de pronostic intermédiaire, avec un score KELIM favorable et une chirurgie sous-optimale (survie globale médiane, 48.0 mois ; IC à 95 %, 44.0-57.1), (3) soit un score KELIM défavorable et une chirurgie optimale (OS médiane, 45 mois ; IC à 95 %, 40-53); et (3) une population de mauvais pronostic, avec un score KELIM défavorable et une chirurgie sous-optimale (survie globale médiane, 22.1 mois ; IC à 95 %, 20.7-24.0).

Survie sans progression

Les résultats étaient cohérents avec ceux obtenus pour la survie globale. Les patientes ayant un score KELIM favorable avaient une survie sans progression (PFS) plus longue : échantillon d'apprentissage, PFS médiane, 26.8 mois (IC 95 %, 23.6-32.1) pour les patientes avec un KELIM standardisé < 1 contre 10.3 mois (IC 95 %, 9.6-11.7) pour les valeurs supérieures à 1; sur l'ensemble de validation, la PFS médiane était de 30.5 mois (IC 95 %, 28.0-34.3) contre 9.8 mois (IC 95 %, 9.4-10.3). Dans les analyses de Cox multivariées, en ajoutant le stade FIGO et le résidu de la maladie, un score KELIM favorable était associé à une meilleure PFS (ensemble d'apprentissage, HR 0.59 ; IC à 95 %, 0.52-0.66 ; ensemble de validation, HR de 0.49 ; IC à 95 %, 0.45-0.54). L'association de KELIM avec le stade FIGO et le résultat de la chirurgie dans le modèle multivarié a conduit à une amélioration de l'indice C à 0.66 (IC à 95 % 0.64-0.69), par

rapport à KELIM, FIGO ou le résultat de la chirurgie considérés seuls (0.59 [IC à 95 % 0.57-0.60], 0.58 [IC à 95 % 0.54-0.61], et 0.60 [IC à 95 % 0.57-0.62], respectivement). La valeur pronostique du KELIM est restée stable pour différents horizons de prédiction jusqu'à 5 ans (au-delà le recul de certains essais était insuffisant).

Les trois mêmes groupes pronostiques basés sur la combinaison du score KELIM et du résultat de la chirurgie ont été distingués, avec des différences médianes importantes en termes de PFS : (1) 58.1 mois pour la population de bon pronostic, (2) 15.1 mois en cas de score KELIM favorable et de chirurgie sous-optimale, 14.9 mois en cas de score KELIM défavorable et de chirurgie optimale et (3) 8.0 mois dans la population de mauvais pronostic. Le KELIM standard continu était également associé de manière indépendante à une probabilité plus élevée de rester longtemps sans maladie > 5 ans.

9.4. Conclusion

L'étude méta-analytique actuelle confirme les résultats de publications antérieures, à savoir l'association du score KELIM avec le pronostic et la survie des patients, indépendamment du fait que la résection chirurgicale ait été complète ou incomplète sur la base des données de plus de 5800 patients ayant participé à huit essais. En particulier, un score KELIM favorable était associé à une meilleure survie globale, avec un rapport de risque proche de 0,5 et des différences de survie globale médiane très significatives sur le plan clinique (environ 80 mois contre environ 30 mois).

La présente étude présente certaines limites communes à toutes les analyses que nous avons réalisées sur MAOV : l'optimalité de la chirurgie d'exérèse était basée sur des lésions résiduelles post-opératoires < ou ≥ 1 cm, comme critère de chirurgie standard alors sur l'exhaustivité de la chirurgie avec un résidu microscopique non visible est maintenant le critère de qualité recommandé. Des études plus récentes ont démontré que la valeur pronostique indépendante de KELIM était similaire lorsque la qualité de la chirurgie était évaluée selon des critères contemporains [83], [199]. En outre, les données individuelles concernant l'histologie et le grade de la tumeur n'ont pas pu être évaluées, en raison de l'absence d'examen pathologique central et des modifications des définitions pathologiques au fil du temps. Le statut mutationnel BRCA ou les déficiences de la recombinaison homologe (HRD) n'étaient pas disponibles, car ils n'étaient pas recherchés dans la pratique courante à l'époque.

Les résultats de cette méta-analyse confirment les publications précédentes sur ce sujet. Néanmoins, des analyses supplémentaires qui prennent en compte les pratiques actuelles sont à notre sens obligatoires pour la mise au point du suivi des patientes.

10. DISCUSSION

10.1 Discussion générale

L'objectif principal des travaux de cette thèse était d'évaluer les performances du biomarqueur CA-125 comme marqueur clé de la progression du cancer de l'ovaire et son rôle en tant que facteur pronostique de la survie pour le suivi des patientes avec un cancer de l'ovaire avancé et nouvellement diagnostiqué. En effet, comme cette maladie est caractérisée par un manque des symptômes, sa détection ne se fait qu'à un stade avancé. Par conséquent, le pronostic vital des patientes est mauvais avec un risque de progression rapide.

Identifier rapidement les patientes avec le plus mauvais pronostic pour adapter le traitement ou les orienter vers des essais cliniques est un objet de recherche important ; en particulier la détection d'une résistance au traitement systémique initial qui est associé à une forte toxicité pourrait justifier de rechercher des alternatives thérapeutiques. La majorité de ces patientes vont recevoir en première ligne de traitement une chimiothérapie en base de platine et taxane et ensuite continuer avec un traitement de maintenance avec Bevacizumab ou des thérapies ciblées. Très souvent les patientes vont présenter une résistance au traitement en base de platine qui est définie par une progression six mois après la fin du traitement de première ligne. Pour ces patientes les options thérapeutiques sont limitées et une priorité aux essais cliniques est proposée. [203]

Pour le suivi des patientes, deux stratégies sont principalement mises en place : (i) un suivi clinique avec titration du biomarqueur CA-125 et déclenchement de scanner si augmentation du CA-125 ou (ii) un suivi avec scanner régulier et/ou l'utilisation du CA-125. Malgré l'utilisation du biomarqueur CA-125 depuis les années 80, la capacité prédictive des meilleures statistiques pour décrire son déclin et son rôle pronostique sont mal quantifiés. L'utilisation des scanners tout au long du suivi des patientes remplace de plus en plus le suivi par le CA-125 et c'est la méthode privilégiée par le corps médical en recherche clinique en France ainsi qu'en routine dans plusieurs pays en Europe par exemple en Allemagne, aux Pays-Bas, en Italie et dans le monde. Par conséquent, au vu de la fragilité des patientes et du risque de rechute rapide, le suivi de ces patientes reste un enjeu important.

Le défi auquel nous souhaitons répondre via ces travaux de thèse était d'essayer de mettre un cadre sur l'analyse des données répétées de CA-125 au cours du suivi des patientes. A notre connaissance plusieurs essais cliniques ont étudié le rôle du CA-125 chez les patientes atteintes d'un cancer de l'ovaire mais ce dernier n'a jamais été étudié et validé avec la méthode de référence dite méta-analytique.

Dans ce travail de thèse nous avons implémenté une méthodologie pour évaluer la capacité pronostique de la dynamique du biomarqueur CA-125 dans le cadre de données de méta-analyse des données individuelles qui se caractérisent par une structure en grappe liée aux caractéristiques de chaque essai clinique inclus. Les méta-analyses permettent de synthétiser les résultats de ces essais cliniques, de nombreux pays, afin de répondre à la question initiale avec un niveau de preuve élevé. Pour obtenir le résumé de la trajectoire du biomarqueur au cours du suivi, nous avons utilisé des modèles statistiques adaptés à ce type des données. Plus

précisément, nous nous sommes intéressés à la meilleure façon d'analyser la décroissance du CA-125 durant le suivi, comment résumer les mesures du CA-125 durant ce suivi, la fenêtre du temps entre le délai de suivi par le CA-125 et le temps de prédiction d'une progression ou d'un décès.

Tout d'abord, nous avons défini la fenêtre de temps sur laquelle nous avons construit les prédictions individuelles. Ensuite, nous avons construit des résumés des mesures du CA-125 obtenues pendant la fenêtre du temps de suivi par 2 grandes approches : en prenant les valeurs observées sous traitement comme la valeur à 3 ou 6 mois, la pente calculée sur les observations disponibles etc. et en estimant les quantités à partir d'un modèle linéaire mixte hiérarchique, comme par exemple la valeur estimée à 3 et 6 mois, la décroissance du biomarqueur estimée à 3 et 6 mois, la pente de décroissance estimée à l'initiation du traitement. Du fait de la structure en grappe, des effets aléatoires emboîtés ont été utilisés, au niveau de l'étude ainsi qu'au niveau du patient pour prendre en compte respectivement les corrélations intra-études et intra-patients.

Nous avons ensuite introduit ces résumés sous la forme de variable indépendante du temps dans un modèle de régression de survie de type semi-paramétrique stratifiée à l'essai ; pour éviter le risque de biais d'immortalité, nous avons appliqué une analyse de landmark et considéré comme date d'origine pour le suivi du temps avant événement, la date clôturant la fenêtre de construction des prédictions individuelles (3 mois ou 6 mois). Nous avons comparé les différents résumés sur leur capacité prédictive.

Leur capacité prédictive a été évaluée en termes de discrimination (AUC(t), Brier score) à un temps t , appelé horizon de prédiction. Nous avons utilisé la méthode '*leave-one-fraction-out*' pour cross-valider les capacités pronostiques du biomarqueur et se protéger du risque de sur-apprentissage. En effet, nous avons créé 10 sous échantillons représentatifs à partir de la population totale de la méta-analyse. Chaque échantillon comportait des patientes de toutes les études incluses dans la méta-analyse. L'échantillon d'apprentissage incluait les 9/10 de l'échantillon total et le 1/10 était l'échantillon de validation. Ceci a été répété 50 fois en créant des sous échantillons aléatoirement.

Dans le chapitre 7, nous avons évalué le résumé de la trajectoire du CA125 sur les premiers mois du suivi (temps de landmark : 3 & 6 mois) pour prédire la survie à un horizon de 24, 36, 48 et 60 mois. Afin de prendre en compte la nature non-linéaire du biomarqueur, nous avons modélisé cette trajectoire à l'aide de splines cubiques naturelles avec 3 ou 4 nœuds, dont un au début du suivi afin de détecter une possible décroissance rapide à l'initiation du traitement.

Ce travail a mis en évidence qu'un suivi de 3 mois avec le biomarqueur CA-125, nous donnait une relativement bonne qualité de prédiction pour la survie des patientes jusqu'à 48 mois. La valeur observée du CA-125 à 3 mois et la valeur estimée à 3 mois conduisaient à la meilleure capacité pronostique parmi les mesures observées et estimées. Les prédictions faites après 6 mois d'étude du CA-125 ne conduisaient pas à une capacité prédictive supérieure. Par conséquent, il n'y a pas besoin d'atteindre 6 mois pour décider un potentiel changement de traitement ou de suivi. L'ajout de la mesure du CA-125 à l'initiation du traitement n'a pas changé la performance de ces deux résumés.

Si on dichotomise la variable biomarqueur à 3 mois selon le seuil de 35 qui est considéré comme la valeur haute normale, on peut exprimer la capacité prédictive par les mesures classiques de valeurs prédictives positive et négative pour le risque de décès à un horizon donné : ainsi, une patiente qui a une valeur du CA-125 inférieure à 35 U/mL à 3 mois a un risque de décès de 20% à 24 mois et 66% à 60 mois. Malgré la normalisation de la valeur du CA-125 après trois mois de traitement, la patiente a un risque de 20% de décéder à 24 mois.

Dans la deuxième partie de cette thèse, nous avons adapté la même méthodologie afin d'évaluer les capacités pronostiques de la dynamique du CA-125 à 3 mois pour détecter une résistance des patientes à la chimiothérapie à base de platine. La résistance au platine est définie par une progression dans les 6 mois après la fin du traitement initial. Au vu de la rapidité de la progression de ce cancer et de la toxicité des traitements à base de sels de platine, l'identification précoce d'un échec très probable du traitement permettrait de stopper un traitement toxique et de proposer une alternative thérapeutique, possiblement un traitement expérimental dans le cadre d'un essai clinique, alors que le cancer ne s'est pas encore diffusé trop largement dans différents organes. Le marqueur CA-125 remplit un certain nombre de prérequis (facilité et standardisation de la mesure, données répétées, coût limité), considérant l'impact potentiel sur le traitement, l'évaluation précise de la capacité prédictive est essentielle.

Sur les 13 essais qui ont utilisé un traitement à base de platine et avec des données du CA-125 disponibles au cours du suivi, nous avons adapté la méthodologie utilisée pour la première publication pour prédire le risque de progression 6 mois après la fin de chimiothérapie par sels de platine.

La trajectoire du CA-125 pendant les 3 premiers mois de chimiothérapie a été représentée par les résumés observés et estimés comme décrit précédemment. Cette nouvelle analyse a permis de mettre en évidence que la valeur du CA-125 à 3 mois et la valeur estimée par le HLMM à 3 mois ont la meilleure performance prédictive parmi les résumés comparés.

Les résumés comme la pente du CA-125 estimée à l'initiation du traitement et à 3 mois, et la pente observée à 3 mois ont amélioré leur performance prédictive lorsque nous avons ajouté le stade FIGO et la maladie résiduelle dans une analyse multivariée. Ils ont montré des résultats équivalents avec la valeur du CA-125 observée à trois mois et la valeur du CA-125 estimée à trois mois, analysées en univarié.

Pour le seuil dit de normalisation du biomarqueur à 35 IU/ml, nous avons calculé la valeur prédictive positive et la valeur prédictive négative afin de pouvoir permettre la prise en décision de la part du clinicien. Par exemple, 19 % des patientes dont la valeur du CA-125 est inférieure à 35 à 3 mois sont résistantes au platine. Également, 56 % des patientes dont la valeur du CA-125 est supérieure à 35 sont résistantes au platine. On peut légitimement s'interroger si ces VPP et VPN sont suffisantes pour justifier un changement de traitement avant progression documentée de la maladie ; en effet, 44% des patientes considérées comme à risque de résistance sur la base du modèle de CA-125 n'auront en fait pas de progression sous platine

avant 6 mois et seront considérées comme sensible à cette chimiothérapie au bénéfice éprouvé.

La question qu'on peut se poser est de calculer la part des patients pour lesquelles on va estimer qu'il serait préférable d'effectuer un changement de traitement et qui pour autant n'en verrait pas le bénéfice sur son état de santé. Plus précisément, 21% des patientes avec une valeur du CA-125 à 3 mois <35 IU/ML vont quand même présenter une résistance au platine. Pour seulement 8% des patientes avec une valeur du CA-125 supérieure à 35 IU/ML on pourra envisager un changement de traitement mais qui ne vont finalement pas présenter une résistance à la chimiothérapie en base de platine. Pour ces patientes, on proposera un traitement probablement moins efficace alors que finalement ne sont pas résistantes au platine.

Un autre résumé de la trajectoire du CA-125 est le score KELIM. Ce dernier a été développé par l'équipe de Lyon 'CICLY' qui propose son utilisation pour le suivi des patientes en première ligne de traitement avec FIGO III, IV et d'histologie séreux. Le score KELIM en tant que facteur pronostique permet d'aider à l'orientation du traitement des patientes.[88], [204] Une application en ligne permet d'obtenir l'estimation de la constante KELIM à partir de 3 mesures du CA-125 durant les 100 premières jours de traitement.

En supposant des relations directes et indirectes entre le traitement et le biomarqueur, ils concluent à un système d'équations différentielles qui permet d'approcher numériquement le taux d'élimination du CA-125. Plusieurs essais cliniques ont été développés ces dernières années afin d'aider les cliniciens sur la prise de décision d'un changement de traitement. [81], [82], [83], [85] Des résultats intéressants ont été également mis en évidence lors de l'utilisation des données d'un sous-groupe de patientes de la méta-analyse MAOV.[183] Dans cette analyse 8 essais randomisés de la méta-analyse MAOV ont participé. L'échantillon total a été reparti en deux sous échantillons, un échantillon d'apprentissage de 1962 patientes et un échantillon de validation de 3922 patientes. Les patients ayant un score KELIM favorable ($\geq 1,0$) ont une survie globale significativement plus longue que ceux ayant un score défavorable ($< 1,0$) : Sur l'échantillon d'apprentissage la survie globale médiane était de 81.8 mois (IC à 95 %, 68.7- non atteint (NA)) vs 31.1 mois (IC 95 %, 29.2-35.7) pour les sous-groupes favorable et défavorable, respectivement. Dans les analyses univariées de la survie globale, le C-index associé au KELIM était de 0,60 (IC à 95 %, 0,58-0,62) et de 0,62 (IC à 95 %, 0,60-0,64) dans les deux échantillons d'apprentissage et de validation, respectivement.

Les résultats de cette étude ne sont pas directement comparables avec nos études pour deux raisons : (i) ils ne sont pas basés sur exactement la même population : Les patientes avec minimum deux mesures pendant les 3 premières mois de traitement ont été inclus à notre analyse alors que pour le calcul du KELIM chaque patiente devrait avoir trois mesures disponibles pendant les 100 premiers jours de traitement et (ii) la capacité discriminante du score KELIM a été évalué avec le C de Harrel alors que nous avons utilisé le AUC(t) et le Brier score à un temps de prédiction spécifique et non pas infini.

En effet, on voit que le CA-125 ou le KELIM sur ces études préliminaires ont montré des résultats prometteurs sur leur capacité d'identifier les patientes à risque plus élevée d'avoir une progression ou un décès. Néanmoins, le risque de prendre une mauvaise décision pour une patiente concernant le changement de son traitement est important si la décision est basée uniquement au niveau du CA-125. Le risque de toxicité est utile d'être étudié, lors d'un changement de traitement -non nécessaire, même si le nombre de faux positifs était inférieur à 10%. Pour ces raisons, nous conseillons la réalisation d'études plus approfondies avec une mise à jour des données de la méta-analyse et la mise en place d'un essai clinique qui va mesurer ce risque.

La mise en pratique clinique d'un marqueur pronostique pour le choix du traitement devrait reposer sur une étude randomisée comparant des stratégies, comme ce qui a été fait pour le CA-125 par Lindemann et al.[3] L'essai AURELIA a évalué la concordance entre la progression définie par le CA-125 et la progression définie par RECIST chez les patientes avec un cancer de l'ovaire résistant au platine. Dans cette étude seulement 43% des patientes ont eu un statut de progression concordant entre progression définie par le CA-125 et celle définie par RECIST.

Dans le cadre d'un changement de pratiques cliniques que ça soit dans le suivi ou le traitement des patientes, plusieurs éléments doivent être vérifiés et validés. Plus, pour mesurer la capacité prédictive d'un marqueur ou d'un résumé du marqueur durant une période, une recherche la plus exhaustive possible des papiers publiés sur le sujet doit être faite.

D'autre part, le cancer de l'ovaire est une maladie qui est lié non seulement à l'évolution du CA-125 mais aussi à d'autres marqueurs. Des analyses oncogénétiques sont aussi effectuées avant l'initiation du traitement afin d'identifier des anomalies génétiques et présence des syndromes comme par exemple le syndrome de Lynch ou des mutations comme BRCA1 et BRCA2 ainsi que les populations HRD. Ce sont des caractéristiques spécifiques à chaque patient qui peuvent avoir un impact sur le parcours du suivi et le choix du traitement mais ces informations n'ont pas été documentée dans les essais inclus dans la méta-analyse MAOV.

D'autres marqueurs sanguins sont aussi des candidats prometteurs pour le suivi des patientes avec un cancer des ovaires nouvellement diagnostiquées. Le développement des biotechnologies permet de détecter d'autres protéines présentes dans le sang qui sont produites en grande quantité par les cellules cancéreuses. Par exemple, en plus du marqueur CA-125, les marqueurs suivants peuvent être identifiés : HE4 (protéine épидидymaire humaine de type 4), CASA (Cancer Associated Serum Antigen) et l'ADN circulant. Pour le moment ces marqueurs ne sont pas systématiquement collectés dans les essais cliniques La FDA (Food and Drug Administration) aux Etats-Unis recommande le dosage de la protéine HE4 pour le suivi des patientes dans le cas d'un cancer de l'ovaire et on peut s'attendre à la disponibilité prochaine de ces données. L'utilisation du CA-125 avec le HE4 donne des résultats prometteurs pour la détection du cancer de l'ovaire. En effet, les deux marqueurs combinés présentent une amélioration de la sensibilité pour la détection du cancer de l'ovaire. [195], [205], [206]. Chez les patientes avec une maladie résiduelle suite à une première chirurgie, la protéine CASA

semble aussi améliorer la performance du CA-125 et permet d'optimiser la surveillance des patientes. [207]

La principale limite de cette étude est le fait que cette méta-analyse inclut des études conduites au début des années 2000. A la période à laquelle ces essais ont été lancés, les patientes ne pouvaient pas encore bénéficier ni de l'optimisation de leur parcours thérapeutique avec l'aide de leur statut BRCA ou HRD ni des thérapies ciblées incluant des inhibiteurs PARP qui font désormais partie des traitements de référence en maintenance des patientes HRD.

Les résultats de cette méta-analyse ont basés uniquement sur les mesures de discrimination et calibration qui ne sont pas des preuves suffisantes pour un changement de pratiques. Un essai randomisé permettra de mettre au point une stratégie de suivi et de changement de traitement. Par exemple, il serait intéressant chez les patientes qui ne normalisent pas le CA-125, les randomiser pour un changement de traitement immédiat versus attendre la progression radiologique.

Nous avons utilisé la méthode de « *leave-one-fraction-out* » pour cross-valider les résultats obtenus. La question de l'échantillonnage pour cross-validation dans une méta-analyse est un sujet insuffisamment exploré. Du fait de la réunion d'études multiples et indépendantes, une méta-analyse permet naturellement d'avoir un échantillon externe de validation en proposant la validation avec la méthode « *leave-one-study-out* ». Toutes les études à l'exclusion d'une servent à l'apprentissage tandis que l'étude restante correspond à l'échantillon de validation. La question qui se pose naturellement au vu de ce type de validation est la transportabilité des résultats d'une étude à une autre. C'est ce qui a été fait pour l'étude de validation de la PFS comme critère de substitution de l'OS[134] suite aux travaux conduit par Rotolo et al.[208]

En médecine personnalisée qui en plein développement, les biomarqueurs sont des éléments clefs pour construction du parcours thérapeutique. En effet, lorsqu'une nouvelle patiente arrive nous souhaitons construire son parcours de soins. Afin d'utiliser la méthodologie développée via cette thèse, on se demande pour cette nouvelle patiente quel sera l'effet essai, ou quel est l'effet lorsqu' on souhaite développer une nouvelle étude ? Pour calculer les prédictions individuelles de cette nouvelle patiente et en suivant la méthodologie développée, nous avons besoin de connaître l'effet aléatoire de « l'étude » de la nouvelle patiente ainsi que le risque instantané dit de base (pour des patients sans facteurs de pronostique) dans cette « étude ». On voit ici la limite induite par le modèle méta-analytique puisqu'une nouvelle patiente n'appartient pas à une étude, au moins clinique. Plusieurs possibilités sont envisageables. Idéalement, on pourrait imagine qu'un hôpital construise sa propre cohorte réunissant les patientes traitées dans le cadre du soin et pour lesquelles toutes les données de CA-125 seraient collectées, permettant de d'estimer l'effet centre spécifique. Cet effet peut être mis à jour régulièrement afin d'être calibré au plus juste. L'inférence bayésienne est bien adaptée à cette approche d'apprentissage continue. La modification des pratiques, des outils de

dépistage ou de diagnostic seraient naturellement pris en compte par les mises à jour successives. Si une telle option n'est pas disponible les solutions suivantes sont proposées par Debray et al 2013 [117]: a) calculer un intercept moyen, c'est-à-dire fixer un effet aléatoire centre égal à 0, b) sélectionner une valeur pour l'effet aléatoire parmi les estimations disponibles pour les différentes études de la méta-analyse celle qui serait la plus similaire à la population dont la nouvelle patiente serait tirée (par exemple l'étude française pour prédire la valeur d'une patiente en métropole), c) construire un effet étude virtuel sous certaines hypothèses et pour des caractéristiques similaires.



L'application de ces méthodes sur la mise en place de la prédiction individuelle reste difficile à appliquer surtout à cause des hypothèses fortes qui sont indispensables. Afin de pouvoir évaluer l'impact que de l'erreur que ces hypothèses peuvent inciter, il serait intéressant d'appliquer la méthode sur une étude avec une population un peu différente et quantifier l'impact d'une mauvaise calibration sur la classification des patients.

10.2 Conclusion

Ce travail de thèse repose sur le plus grand échantillon de données de CA-125 collectées, nettoyées, standardisées dans le cadre de recherches cliniques disponible à ce jour. C'est, à notre connaissance, le premier à comparer la capacité pronostique de plusieurs résumés du CA-125 au sein d'une méta-analyse. Des résultats prometteurs ont été révélés lors de ces travaux mais qui nécessitent des preuves supplémentaires afin de permettre le changement des pratiques de suivi et du traitement d'une patiente.

11. ANNEXES


Figure 1 : Article publié



Article

CA-125 Early Dynamics to Predict Overall Survival in Women with Newly Diagnosed Advanced Ovarian Cancer Based on Meta-Analysis Data


Eleni Karamouza ^{1,2,*}, Rosalind M. Glasspool ³, Caroline Kelly ⁴, Liz-Anne Lewsley ⁴, Karen Carty ⁴, Gunnar B. Kristensen ⁵, Josee-Lyne Ethier ⁶, Tatsuo Kagimura ⁷, Nozomu Yanaihara ⁸, Sabrina Chiara Cecere ⁹, Benoit You ^{10,11}, Ingrid A. Boere ¹², Eric Pujade-Lauraine ¹³, Isabelle Ray-Coquard ¹⁴, Cécile Proust-Lima ¹⁵ and Xavier Paoletti ^{16,17}

 **check for updates**

Citation: Karamouza, E.; Glasspool, R.M.; Kelly, C.; Lewsley, L.-A.; Carty, K.; Kristensen, G.B.; Ethier, J.-L.; Kagimura, T.; Yanaihara, N.; Cecere, S.C.; et al. CA-125 Early Dynamics to Predict Overall Survival in Women with Newly Diagnosed Advanced Ovarian Cancer Based on Meta-Analysis Data. *Cancers* 2023, 15, 1823. <https://doi.org/10.3390/cancers15061823>

Academic Editor: Caroline Bascoul-Mollevi

Received: 4 February 2023
Revised: 10 March 2023
Accepted: 13 March 2023
Published: 17 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Cancer antigen 125 (CA-125) is a protein found at a high concentration in the blood of patients with specific types of cancer, mainly ovarian cancer. In 2004, the Gynecologic Cancer Intergroup (GCIIG) proposed criteria defining response to treatment, as well as disease progression, based on the CA-125 concentration. Ever since, for the follow-up of ovarian cancer patients, the CA-125 concentration and/or CT-scans are used. This paper aims to compare different summaries of CA-125 evolution in the 3 to 6 months following treatment initiation in newly diagnosed advanced ovarian cancer and explore their prognostic capacity to predict overall survival. Based on individual patient data from the GCIIG meta-analysis, we propose the most appropriate timeframe between follow-up and the prediction horizon in order to obtain robust, dynamic, individual predictions.

Abstract: (1) Background: Cancer antigen 125 (CA-125) is a protein produced by ovarian cancer cells that is used for patients' monitoring. However, the best ways to analyze its decline and prognostic role are poorly quantified. (2) Methods: We leveraged individual patient data from the Gynecologic Cancer Intergroup (GCIIG) meta-analysis (N = 5573) to compare different approaches summarizing the early trajectory of CA-125 before the prediction time (called the landmark time) at 3 or 6 months after treatment initiation in order to predict overall survival. These summaries included observed and estimated measures obtained by a linear mixed model (LMM). Their performances were evaluated

Cancers 2023, 15, 1823. <https://doi.org/10.3390/cancers15061823>

<https://www.mdpi.com/journal/cancers>

Table 1 : Classification FIGO & TNM pour le cancer ovarien. [37]

Stade I : Les cellules cancéreuses sont présentes seulement aux ovaires.

		T1-N0-M0
IA	Croissance limitée à un ovaire ; pas d'ascite contenant des cellules malignes. Aucune tumeur à la surface externe ; capsule intacte	
		T1a-N0-M0
IB	Croissance limitée aux deux ovaires ; pas d'ascite contenant des cellules malignes. Aucune tumeur sur les surfaces externes ; capsules intactes	
		T1b-N0-M0
IC*	Tumeur limitée à un ou aux deux ovaires ou trompes de Fallope, avec l'un des éléments suivants	
IC1	Déversement chirurgical	
		T1c1-N0-M0
IC2	Capsule rompue avant une intervention chirurgicale ou tumeur à la surface de l'ovaire ou de la trompe de Fallope	
		T1c2-N0-M0
IC3	Cellules malignes dans les ascites ou les lavages péritonéaux	
		T1c3-N0-M0

Stade II : La tumeur implique 1 ou les deux ovaires ou trompes de Fallope avec extension pelvienne (sous le bord pelvien) ou cancer péritonéal

		T2-N0-M0
IIA	Extension et/ou implants sur utérus et/ou trompes de Fallope et/ou ovaires	
		T2a-N0-M0
IIB	Extension à d'autres tissus intrapéritonéaux pelviens	
		T2b-N0-M0

Stade III : La tumeur implique un ou les deux ovaires ou trompes de Fallope, ou un cancer péritonéal, avec propagation confirmée histologiquement ou par la cytologie, au péritoine en dehors du bassin et/ou métastase aux ganglions lymphatiques rétropéritonéaux

		T1/T2-N0-M0
IIIA1	Ganglions lymphatiques rétropéritonéaux positifs uniquement prouvés histologiquement ou par la cytologie :	
IIIA1(i)	Métastase jusqu'à 10 mm dans la plus grande dimension	
IIIA1(ii)	Métastase de plus de 10 mm dans la plus grande dimension	

IIIA2	Atteinte péritonéale microscopique extrapelvienne (au-dessus du bord pelvien) avec ou sans ganglions lymphatiques rétropéritonéaux positifs	T3a2-N0/N1-M0
IIIB	Métastases péritonéales macroscopiques au-delà du bassin jusqu'à 2 cm dans la plus grande dimension, avec ou sans métastases aux ganglions lymphatiques rétropéritonéaux	T3b-N0/N1-M0
IIIC	Métastases péritonéales macroscopiques au-delà du bassin de plus de 2 cm dans leur plus grande dimension, avec ou sans métastases aux ganglions lymphatiques rétropéritonéaux (comprend l'extension de la tumeur à la capsule du foie et de la rate sans atteinte parenchymateuse de l'un ou l'autre organe)	T3c-N0/N1-M0

Stade IV : Métastases à distance excluant les métastases péritonéales

IVA	Épanchement pleural avec cytologie positive	
IVB	Métastases parenchymateuses et métastases aux organes extra-abdominaux (y compris les ganglions lymphatiques inguinaux et les ganglions lymphatiques situés en dehors de la cavité abdominale)	Tout T, Tout N, M1

Table 2 : Les cinq plus fréquents types histologiques du cancer de l'ovaire [34]

	Séreux bas grade	Séreux haut grade	Mucineux	Endométrioïde	Cellules claires
Fréquence	<5%	70%	2 - 3%	10%	5 - 10%
Immunophénotypage	CK7+, WT1+, ER+	CK7+, CK20, PAX8+, WT1+	CK7+, CK20-, ER-, PR-, WT1-	CK7+, PAX8+, CK20-, WT1-	napsin A+, WT1-, p53-, ER-
Lésion précurseur	Lésions potentielles malignes de bas grade	Carcinome intraépithélial séreux des trompes	Début de lésion mucineuse	Endométriose	Endométriose
Anomalies moléculaires	KRAS, BRAF, NRAS, PIK3CA	TP53, BRCA1/2, HRR	KRAS, HER2, PIK3CA	ARID1A, PTEN, PIK3CA, POLE, MMR	ARID1A, PIK3CA, PTEN
Pronostic	Intermédiaire	Faible	Bon	Favorable	Intermédiaire

Table 3 : Recommandations pour la validation d'un modèle pronostique dans un cadre méta-analytique [125]

Fournir des détails sur les données individuelles	Décrire le nombre de patients et d'événement pour chaque étude en fonction de son utilisation pour l'apprentissage ou la validation
	Décrire les données manquantes, notamment sur chaque prédicteur et les possibles exclusions de patients faute de données suffisantes
	Rapporter les études non incluses ou non obtenues, et les effectifs associés
Méthodes statistiques pour le développement du modèle	Prendre en compte la structure en grappes des études en autorisant des risques de base différents entre les études
	Détailler les critères de sélection et les procédures pour sélectionner des variables incluses dans le modèle
	S'il existe une forte hétérogénéité pour certains prédicteurs, tenter de la réduire par un choix d'autres prédicteurs
	Modéliser les variables continue sans tenter de les catégoriser
	Fournir les estimations et erreurs standards des paramètres du modèle final
	Préciser la gestion des données manquantes
Validation du modèle	Valider le modèle sur un échantillon externe et par cross-validation
	Détailler le choix du risque de base lors de l'application du modèle pour la partie validation
	Décrire s'il y a des éléments montrant que la validation varie entre les différentes études / échantillons utilisés
	Fournir une estimation de la capacité prédictive du modèle finale et précisant bien quel risque de base a été et sera utilisé

Table 4 :TRIPOD-SRMA* Liste pour le 'reporting' des revues systématiques d'études avec modèles de prédiction

Section et thème	Numéro d'item	Item de la liste
Titre		
Titre	1	Identifier le rapport comme une revue systématique ou une méta-analyse (ou les deux) d'études de modèles diagnostiques ou pronostiques. Préciser la population cible

Section et thème	Numéro d'item	Item de la liste
		et le(s) résultat(s) prévu(s) en fonction de la question de la revue.
Résumé		
Résumé	2	Voir la liste TRIPOD-SRMA pour les résumés*
Introduction		
Motivation	3	Décrire la raison d'être de la revue dans le contexte des connaissances existantes.
Objectives	4	Fournir une déclaration explicite de l'objectif ou des objectifs visés en se référant à la population cible, à l'indice et aux modèles de comparaison (le cas échéant), au(x) résultat(s), au temps (horizon de prédiction et moment prévu pour l'utilisation du modèle) et au contexte.
Méthodes		
Critères d'éligibilité de l'étude	5	Préciser les caractéristiques de l'étude utilisées comme critères d'éligibilité, y compris les modèles de prédiction présentant un intérêt particulier, et préciser si les études de développement ou de validation (ou les deux) étaient éligibles.
Sources d'information	6	Préciser toutes les bases de données, registres, sites web, organisations, listes de référence et autres sources recherchées ou consultées pour identifier les études. Préciser la date à laquelle chaque source a été consultée pour la dernière fois.
Stratégie de recherche	7	Présenter les stratégies de recherche complètes pour toutes les bases de données, tous les registres et tous les sites web, y compris les filtres et les limites utilisés.
Processus de sélection des études	8	Préciser les méthodes utilisées pour décider si une étude répond aux critères d'inclusion de la revue, y compris le nombre de réviseurs qui ont examiné chaque enregistrement et chaque rapport récupéré, s'ils ont travaillé indépendamment et, le cas échéant, les détails des outils d'automatisation utilisés dans le processus.
Processus de collecte des données	9	Précisez les méthodes utilisées pour collecter les données des rapports d'étude, y compris le nombre de réviseurs qui

Section et thème	Numéro d'item	Item de la liste
		ont collecté les données de chaque rapport, s'ils ont travaillé de manière indépendante, tout processus d'obtention ou de confirmation des données auprès des investigateurs de l'étude et, le cas échéant, les détails des outils d'automatisation utilisés dans le processus.
Tems des données	10a	Énumérer et définir tous les éléments pour lesquels des données ont été recherchées dans chaque étude.
	10b	Indiquer les mesures de performance du modèle recherchées (par exemple, mesures de calibration, de discrimination, d'adéquation globale du modèle, d'utilité clinique).
	10c	Décrivez comment les données souhaitées mais non rapportées (points 10a, 10b) ont été traitées (par exemple, en contactant les auteurs, en les calculant à partir d'autres informations rapportées).
Risque de biais et évaluation de l'applicabilité	11	Préciser les méthodes utilisées pour évaluer le risque de biais dans les études incluses et leur applicabilité à la question de l'examen. Cela doit être fait séparément pour chaque développement et validation de modèle. Inclure les détails de tout outil utilisé, le nombre de réviseurs qui ont évalué chaque étude et s'ils ont travaillé de manière indépendante.
Méthodes de synthèse	12a	Décrivez les méthodes utilisées pour synthétiser les estimations des mesures de performance pour chaque modèle. Si une méta-analyse a été réalisée, décrivez les méthodes utilisées, y compris toute transformation des données avant la mise en commun, la manière dont toute hétérogénéité dans la performance du modèle a été quantifiée et traitée, et le(s) logiciel(s) utilisé(s)..
	12b	Décrire les méthodes utilisées pour explorer les causes possibles de l'hétérogénéité des performances des modèles (par exemple, analyse de sous-groupes, méta-régression), en précisant si elles ont été planifiées ou non.
	12c	Décrire les analyses de sensibilité effectuées pour évaluer la robustesse des résultats synthétisés.

Section et thème	Numéro d'item	Item de la liste
Évaluation de la certitude	13	Décrire les méthodes utilisées pour évaluer la certitude (ou la confiance) dans l'ensemble des preuves d'un modèle de prédiction.
Résultats		
Sélection de l'étude	14	Décrire les résultats du processus de recherche et de sélection, du nombre d'enregistrements identifiés lors de la recherche au nombre d'études et de modèles inclus dans l'examen, de préférence à l'aide d'un diagramme de flux.
Caractéristiques de l'étude et du modèle	15	Présenter les caractéristiques de l'étude et les détails du modèle extraits (conformément au point 10a) et citer les rapports d'étude.
Risque de biais et applicabilité	16	Présenter les résultats de l'évaluation du risque de biais et de l'applicabilité. Cela doit être fait séparément pour chaque développement et validation de modèle dans chaque étude incluse.
Résultats des performances du modèle dans les études individuelles	17	Présenter les estimations de performance et les intervalles de confiance pour chaque modèle et toutes les évaluations, en précisant s'ils se rapportent à la performance de la validation interne ou externe. S'il s'agit d'une validation interne, donnez des détails sur la méthode.
Résultats de synthèses	18a	Présenter les résultats de toute synthèse des performances du modèle, ainsi que les détails des estimations de l'étude qui y ont contribué. Si une méta-analyse a été réalisée, présenter, pour chaque modèle et chaque mesure de performance, les résultats sommaires, les intervalles de confiance/crédibles et les mesures de l'hétérogénéité. Des diagrammes de forêt peuvent être utiles.
	18b	Pour chaque modèle, présenter les résultats de toutes les recherches sur les causes possibles de l'hétérogénéité des performances du modèle.
	18c	Présenter les résultats de toutes les analyses de sensibilité effectuées pour évaluer la robustesse des résultats synthétisés.

Section et thème	Numéro d'item	Item de la liste
Certitude des preuves	19	Présenter toute évaluation de la certitude (ou de la confiance) dans l'ensemble des preuves pour chaque modèle de prédiction d'intérêt.
Discussion		
Résumé des éléments de preuve	20	Résumer les principaux résultats, y compris les points forts et les limites des données.
Limitations	21	Discuter des points forts et des limites du processus de révision.
Conséquences	22	Discuter des implications des résultats dans le contexte d'autres preuves et pour la pratique, la politique et la recherche future.
Autres informations		
Enregistrement et protocole	23a	Fournir les informations relatives à l'enregistrement de l'examen, y compris le nom du registre et le numéro d'enregistrement, ou indiquer que l'examen n'a pas été enregistré.
	23b	Indiquer où le protocole d'examen peut être consulté ou préciser qu'aucun protocole n'a été préparé.
	23c	Décrire et expliquer toute modification apportée aux informations fournies lors de l'enregistrement ou dans le protocole.
Soutien	24	Décrire les sources de soutien financier ou non financier pour l'examen et le rôle des bailleurs de fonds ou des sponsors dans l'examen.
Intérêts concurrents	25	Déclarer les intérêts concurrents des auteurs de la revue.
Disponibilité des données, du code et d'autres matériels	26	Indiquer quels sont les éléments suivants qui sont accessibles au public et où ils peuvent être trouvés : modèles de formulaires de collecte de données ; données extraites des études incluses ; données utilisées pour toutes les analyses ; code analytique ; tout autre matériel utilisé dans le cadre de l'examen.

*TRIPOD-SRMA= Transparent reporting of multivariable prediction models for individual prognosis or diagnosis tailored for systematic reviews and meta-analyses.

Table 5 : Caractéristiques des essais

Essai Investigateur Principal	Traitement Investigationnel (# patientes)	Traitement Standard (# patientes)	Progression ^a	Std trt	Inv trt	Date incl usion	Mediane suivi (années) (q1-q3)	Mediane OS, (années)	Mediane PFS, (années)
Maintenance									
Vergote et al, 2014 (EORTC-55041)[209]	Erlotinib (420)	Observation (415)	CA-125-Clinical(confir mation with CT)	412	419	2005	4.3 (3.8 - 4.8)	4.6	1.0
Hirte et al, 2006 (CCTG-OV.12)[210]	Tanomastat (122)	Placebo (121)	CA-125-Clinical(confir mation with CT)	121	122	1998	0.9 (0.6 - 1.3)	NA	0.9
Reyners et al,2012 (DoCaCel)[211]	Docetaxel + Carboplatin + Celecoxib (97)	Docetaxel + Carboplatin (99)	GCIG criteria	99	97	2003	4.1 (2.6 - 5.7)	2.9	1.2
Oza et al, 2015 (MRC-ICON7)[212]	Bevacizumab (764)	Standard chemotherapy (764)	GCIG criteria	764	764	2006	4.6 (4.2 - 5.1)	4.8	1.6
Mannel et al, 2011 (GOG-0175)[213]	Low Dose Paclitaxel (274)	Observation (268)	GCIG criteria	268	274	1998	11.6 (8.5 - 13.7)	NA	NA
No maintenance									
Aravantinos et al, 2008 (HECOG-4A99)[214]	Cisplatin+Paclitax el+Doxorubicin (236)	Paclitaxel + Carboplatin (233)	CT-scan	221	225	1999	13.7 (5.4 - 16.1)	3.2	1.3
Pignata et al, 2011 (MITO-2)[215]	Carboplatin + Liposomal doxorubicin(410)	Carboplatin + Paclitaxel (410)	Mixed ^b	392	396	2003	6.0 (5.0 - 7.1)	4.7	1.5
Vasey et al, 2004 (SCOTROC-1)[216]	Docetaxel+Carbop latin (539)	Paclitaxel + Carboplatin (538)	CA-125-Clinical(confir mation with CT)	537	538	1998	2.0 (1.6 - 2.4)	2.9	1.2
Sugiyama et al, 2016 (JGOG-3017)[217]	Irinotecan + Cisplatin (332)	Carboplatin + Paclitaxel (335)	CT-scan	332	329	2009	3.7 (2.82 - 4.8)	NA	NA
Hoskins et al, 2010 (CCTG-OV.16)[218]	Cisplatin + Topotecan followed by Paclitaxel + Carboplatin (409)	Paclitaxel + Carboplatin (410)	GCIG criteria	410	409	2002	8.2 (7.5 - 8.9)	3.7	1.3

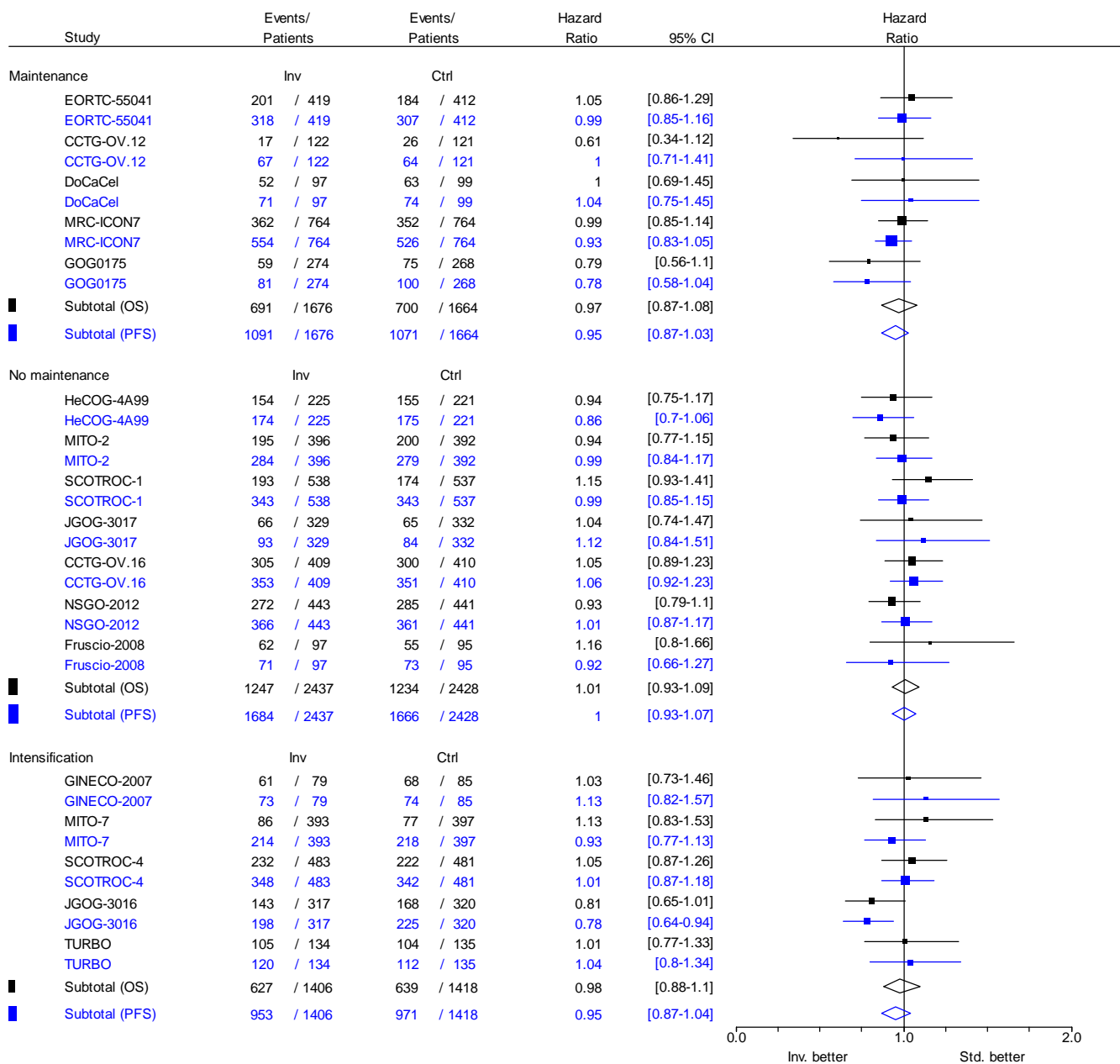
Lindemann et al, 2012 (NSGO-2012)[219]	Paclitaxel+Carboplatin+Epirubicin (445)	Paclitaxel + Carboplatin(442)	CA-125-Clinical(confir- mation with CT)	441	443	1999	5.3 (4.3 - 5.9)	3.4	1.4
Fruscio et al, 2008[220]	Cisplatin + Isosfamide + Paclitaxel (106)	Cisplatin + Epirubicin + Paclitaxel (103)	CA-125-Clinical(confir- mation with CT)	95	97	1997	6.8 (6.2 - 7.3)	4.7	1.9
Intensification therapy									
Ray-Coquard et al, 2007 (GINECO-2007)[221]	1800 Cyclophosphamide + 50 Erubicin + 75 Cisplatin + 5 Filgr (79)	500 Cyclophosphamide + 50 Erubicin + 75 Cisplatin (85)	CA-125-Clinical(confir- mation with CT)	85	79	1994	8.6 (6.2 - 9.9)	2.7	1.2
Pignata et al, 2014 (MITO-7)[222]	Weekly Carboplatin + Paclitaxel (406)	Every 3 weeks Carboplatin+Paclitaxel (404)	GCIg criteria	397	393	2008	1.9 (1.4 - 2.6)	4.0	1.5
Banerjee et al, 2013 (SCOTROC-4)[223]	Carboplatin Dose escalated (483)	Carboplatin Flat Dose (481)	CA-125-Clinical(confir- mation with CT)	481	483	2005	2.7 (1.7 - 3.6)	2.7	1.0
Katsumata et al, 2013 (JGOG-3016)[224]	Dose dense Carboplatin (317)	Conventional Carboplatin (320)	CA-125-Clinical(confir- mation with CT)	320	317	2004	6.5 (5.9 - 7.2)	6.2	2.3
Van der Burg et al, 2014 (TURBO)[225]	Weekly Paclitaxel + Carboplatin (134)	3-weekly Paclitaxel + Carboplatin (136)	CA-125-Clinical(confir- mation with CT)	135	134	1998	9.4 (8.4 - 11.4)	3.6	1.5

Abréviations : CA-125, antigène 125 du cancer ; CT, tomographie assistée par ordinateur ; GCIg, Gynecologic Cancer InterGroup ; IQR, intervalle interquartile ; NR, non atteint ; OS, survie globale ; PFS, survie sans progression.

^a Les « critères GCIg » indiquent que les patientes ont été suivies à la fois par des mesures sérielles des taux de CA125 et par des mesures radiologiques.

^b La progression des patientes du Groupe d'investigateurs national des Etudes des Cancers Ovariens (GINECO) a été évaluée par le taux de CA125 et confirmée par tomodensitométrie, tandis que les patientes des Multicenter Italian Trials in Ovarian Cancer and Gynecologic Malignancies (MITO) ont été évaluées selon les lignes directrices du GCIg.

Figure 2: Effet de traitement (HR) global et essai par essai sur la OS (noir) et la PFS (bleu)



Cochran Heterogeneity test p-value=0.7, I²=0%
 p-value=0.6, I²=0%

Table 6 : AUC (A, B) et score de Brier (C, D) validés par validation croisée 5 fois pour la prédiction de la survie globale à 24, 36, 48 et 60 mois à partir de l'historique du CA125 jusqu'à 3 mois en considérant différents résumés du CA125. Demi-population comprenant uniquement le bras « régime standard »

Null model: model de survie stratifié à l'essai, ORD: Décroissance relative Observée, OVLT: Valeur observée du CA-125 au temps de landmark, EVLT: Valeur estimée au temps de landmark, ESB: Pente estimée à l'initiation du traitement, ESLT: Pente estimée au temps de landmark

A. AUC	Time	T=24	T=36	T=48	T=60
	Null model		0.592 [0.51 ; 0.667]	0.6 [0.522 ; 0.667]	0.588 [0.512 ; 0.656]
ORD		0.585 [0.503 ; 0.672]	0.604 [0.53 ; 0.67]	0.592 [0.505 ; 0.671]	0.544 [0.461 ; 0.624]
OVLT		0.714 [0.618 ; 0.793]	0.702 [0.613 ; 0.77]	0.7 [0.603 ; 0.769]	0.663 [0.575 ; 0.742]
EVLT		0.712 [0.618 ; 0.791]	0.701 [0.616 ; 0.769]	0.701 [0.608 ; 0.773]	0.667 [0.584 ; 0.745]
ESB		0.642 [0.548 ; 0.739]	0.636 [0.554 ; 0.706]	0.629 [0.548 ; 0.699]	0.582 [0.501 ; 0.657]
ESLT		0.634 [0.547 ; 0.73]	0.645 [0.565 ; 0.716]	0.638 [0.558 ; 0.712]	0.595 [0.511 ; 0.671]
B. AUC	Null model*	0.67 [0.587 ; 0.752]	0.687 [0.608 ; 0.761]	0.691 [0.599 ; 0.777]	0.662 [0.584 ; 0.74]
	ORD*	0.695 [0.598 ; 0.779]	0.706 [0.621 ; 0.778]	0.715 [0.628 ; 0.792]	0.687 [0.608 ; 0.764]
	OVLT*	0.713 [0.619 ; 0.792]	0.713 [0.628 ; 0.782]	0.716 [0.625 ; 0.792]	0.685 [0.599 ; 0.762]
	EVLT*	0.711 [0.618 ; 0.793]	0.712 [0.629 ; 0.78]	0.716 [0.625 ; 0.792]	0.686 [0.604 ; 0.759]
	ESB*	0.683 [0.595 ; 0.765]	0.696 [0.612 ; 0.768]	0.699 [0.609 ; 0.774]	0.671 [0.592 ; 0.749]
	ESLT*	0.67 [0.587 ; 0.751]	0.686 [0.607 ; 0.761]	0.69 [0.596 ; 0.771]	0.661 [0.584 ; 0.741]
C. Brier	Null Model	0.162 [0.131 ; 0.196]	0.216 [0.193 ; 0.24]	0.243 [0.225 ; 0.265]	0.263 [0.246 ; 0.29]
	ORD	0.163 [0.132 ; 0.196]	0.217 [0.195 ; 0.24]	0.244 [0.225 ; 0.266]	0.263 [0.246 ; 0.289]
	OVLT	0.153 [0.124 ; 0.185]	0.197 [0.17 ; 0.229]	0.213 [0.19 ; 0.244]	0.232 [0.21 ; 0.262]
	EVLT	0.152 [0.124 ; 0.185]	0.197 [0.169 ; 0.229]	0.212 [0.189 ; 0.244]	0.231 [0.21 ; 0.259]
	ESB	0.16 [0.13 ; 0.193]	0.212 [0.188 ; 0.236]	0.236 [0.217 ; 0.261]	0.256 [0.236 ; 0.281]
	ESLT	0.161 [0.13 ; 0.194]	0.211 [0.187 ; 0.236]	0.235 [0.216 ; 0.258]	0.254 [0.233 ; 0.277]
D. Brier	Null model*	0.158 [0.127 ; 0.191]	0.203 [0.177 ; 0.229]	0.22 [0.194 ; 0.249]	0.235 [0.21 ; 0.265]
	ORD*	0.153 [0.126 ; 0.188]	0.196 [0.169 ; 0.224]	0.209 [0.184 ; 0.239]	0.226 [0.201 ; 0.253]
	OVLT*	0.153 [0.125 ; 0.187]	0.196 [0.168 ; 0.227]	0.21 [0.186 ; 0.242]	0.227 [0.204 ; 0.257]
	EVLT*	0.153 [0.125 ; 0.187]	0.195 [0.167 ; 0.226]	0.209 [0.185 ; 0.243]	0.226 [0.203 ; 0.255]
	ESB*	0.156 [0.126 ; 0.191]	0.201 [0.175 ; 0.227]	0.217 [0.193 ; 0.245]	0.233 [0.209 ; 0.261]
	ESLT*	0.158 [0.127 ; 0.191]	0.203 [0.177 ; 0.229]	0.22 [0.194 ; 0.25]	0.236 [0.21 ; 0.265]

* tous les modèles sont ajustés au CA-125 à l'initiation du traitement

Table 7 : AUC (A,B) et score de Brier (C,D) validés par validation croisée 5 fois pour la prédiction de la survie globale à 24, 36, 48 et 60 mois à partir de l'historique du CA125 jusqu'à 6 mois en considérant différents résumés du CA125. Demi-population comprenant uniquement le bras « régime standard »

A. AUC	Time	T=24	T=36	T=48	T=60
	Null model		0.588 [0.53 ; 0.649]	0.589 [0.539 ; 0.644]	0.58 [0.532 ; 0.628]
ORD		0.595 [0.529 ; 0.664]	0.607 [0.549 ; 0.667]	0.604 [0.549 ; 0.654]	0.573 [0.515 ; 0.626]
OVLT		0.727 [0.667 ; 0.777]	0.704 [0.65 ; 0.755]	0.686 [0.632 ; 0.735]	0.641 [0.58 ; 0.697]
EVLT		0.721 [0.659 ; 0.774]	0.704 [0.649 ; 0.751]	0.687 [0.631 ; 0.737]	0.638 [0.572 ; 0.691]
ESB		0.634 [0.575 ; 0.694]	0.641 [0.594 ; 0.695]	0.635 [0.578 ; 0.683]	0.598 [0.542 ; 0.66]
ESLT		0.592 [0.532 ; 0.657]	0.598 [0.543 ; 0.651]	0.597 [0.541 ; 0.648]	0.562 [0.505 ; 0.619]
B. AUC	Null model*	0.683 [0.625 ; 0.738]	0.687 [0.632 ; 0.743]	0.683 [0.627 ; 0.732]	0.66 [0.603 ; 0.713]
	ORD*	0.705 [0.646 ; 0.757]	0.703 [0.646 ; 0.754]	0.702 [0.649 ; 0.752]	0.674 [0.621 ; 0.724]
	OVLT*	0.737 [0.677 ; 0.785]	0.727 [0.673 ; 0.777]	0.715 [0.657 ; 0.769]	0.685 [0.627 ; 0.733]
	EVLT*	0.733 [0.67 ; 0.782]	0.725 [0.667 ; 0.776]	0.715 [0.656 ; 0.771]	0.681 [0.623 ; 0.732]
	ESB*	0.696 [0.632 ; 0.752]	0.701 [0.645 ; 0.755]	0.697 [0.643 ; 0.748]	0.676 [0.619 ; 0.729]
	ESLT*	0.682 [0.621 ; 0.737]	0.686 [0.629 ; 0.743]	0.682 [0.627 ; 0.732]	0.659 [0.603 ; 0.714]
C. Brier	Null Model	0.167 [0.148 ; 0.187]	0.214 [0.198 ; 0.231]	0.24 [0.228 ; 0.255]	0.264 [0.25 ; 0.28]
	ORD	0.168 [0.149 ; 0.189]	0.215 [0.198 ; 0.232]	0.241 [0.227 ; 0.256]	0.263 [0.249 ; 0.279]
	OVLT	0.148 [0.129 ; 0.168]	0.191 [0.175 ; 0.214]	0.215 [0.2 ; 0.234]	0.238 [0.221 ; 0.258]
	EVLT	0.148 [0.129 ; 0.169]	0.191 [0.174 ; 0.213]	0.215 [0.199 ; 0.235]	0.239 [0.222 ; 0.259]
	ESB	0.164 [0.145 ; 0.184]	0.208 [0.192 ; 0.226]	0.232 [0.218 ; 0.249]	0.253 [0.235 ; 0.271]
	ESLT	0.167 [0.147 ; 0.187]	0.214 [0.197 ; 0.231]	0.239 [0.225 ; 0.254]	0.261 [0.247 ; 0.277]
D. Brier	Null model*	0.158 [0.142 ; 0.181]	0.199 [0.182 ; 0.217]	0.218 [0.201 ; 0.235]	0.235 [0.216 ; 0.253]
	ORD*	0.153 [0.136 ; 0.175]	0.194 [0.177 ; 0.215]	0.212 [0.196 ; 0.23]	0.231 [0.213 ; 0.25]
	OVLT*	0.147 [0.131 ; 0.168]	0.188 [0.171 ; 0.209]	0.208 [0.192 ; 0.227]	0.227 [0.209 ; 0.248]
	EVLT*	0.148 [0.131 ; 0.168]	0.188 [0.171 ; 0.209]	0.208 [0.191 ; 0.229]	0.228 [0.211 ; 0.248]
	ESB*	0.157 [0.141 ; 0.179]	0.196 [0.178 ; 0.216]	0.214 [0.196 ; 0.234]	0.23 [0.212 ; 0.248]
	ESLT*	0.159 [0.142 ; 0.181]	0.199 [0.182 ; 0.218]	0.219 [0.201 ; 0.236]	0.235 [0.216 ; 0.253]

Null model: model de survie stratifié à l'essai, ORD: Décroissance relative Observée, OVLT: Valeur observée du CA-125 au temps de landmark, EVLT: Valeur estimée au temps de landmark, ESB: Pente estimée à l'initiation du traitement, ESLT: Pente estimée au temps de landmark

* tous les modèles sont ajustés au CA-125 à l'initiation du traitement

12. BIBLIOGRAPHIE

- [1] F. Bray *et al.*, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 74, no. 3, pp. 229–263, 2024, doi: 10.3322/caac.21834.
- [2] G. C. B. Trétarre, M. Mounier, P. Arveux, C. Lecoffre, L. Lafay, "Survie des personnes atteintes de cancer en France 1989-2018," [Http://Www.E-Cancer.Fr/Publications](http://Www.E-Cancer.Fr/Publications), 2020.
- [3] K. Lindemann *et al.*, "Poor concordance between CA-125 and RECIST at the time of disease progression in patients with platinum-resistant ovarian cancer: Analysis of the AURELIA trial," *Ann. Oncol.*, vol. 27, no. 8, pp. 1505–1510, 2016, doi: 10.1093/annonc/mdw238.
- [4] G. Funston *et al.*, "CA125 test result, test-to-diagnosis interval, and stage in ovarian cancer at diagnosis: A retrospective cohort study using electronic health records," *Br. J. Gen. Pract.*, vol. 71, no. 707, pp. E465–E472, 2021, doi: 10.3399/BJGP.2020.0859.
- [5] R. L. PRENTICE, "Covariate measurement errors and parameter estimation in a failure time regression model," *Biometrika*, vol. 69, no. 2, p. 331, Aug. 1982, [Online]. Available: <http://10.0.4.69/biomet/69.2.331>
- [6] C. Proust-Lima and J. M. G. Taylor, "Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: A joint modeling approach," *Biostatistics*, vol. 10, no. 3, pp. 535–549, 2009, doi: 10.1093/biostatistics/kxp009.
- [7] J. E. Overall, S. Tonidandel, and R. R. Starbuck, "Last-observation-carried-forward (LOCF) and tests for difference in mean rates of change in controlled repeated measurements designs with dropouts," *Soc. Sci. Res.*, vol. 38, no. 2, pp. 492–503, 2009, doi: 10.1016/j.ssresearch.2009.01.004.
- [8] O. Colombari *et al.*, "Early Modeled Longitudinal CA-125 Kinetics and Survival of Ovarian Cancer Patients: A GINECO AGO MRC CTU Study.," *Clin. Cancer Res.*, p. clincanres.3335.2018, 2019, doi: 10.1158/1078-0432.CCR-18-3335.
- [9] D. Rizopoulos, *Joint models for longitudinal and time-to-event data : with applications in R*. CRC Press, 2012. Accessed: Jun. 25, 2019. [Online]. Available: <https://www.crcpress.com/Joint-Models-for-Longitudinal-and-Time-to-Event-Data-With-Applications/Rizopoulos/p/book/9781439872864>
- [10] P. J. Heagerty, T. Lumley, and M. S. Pepe, "Time-dependent ROC curves for censored survival data and a diagnostic marker," *Biometrics*, vol. 56, no. 2, pp. 337–344, 2000, doi: 10.1111/j.0006-341X.2000.00337.x.
- [11] T. A. Gerds and M. Schumacher, "Consistent estimation of the expected brier score in general survival models with right-censored event times," *Biometrical J.*, vol. 48, no. 6, pp. 1029–1040, 2006, doi: 10.1002/bimj.200610301.
- [12] P. Blanche, C. Proust-Lima, L. Loubère, C. Berr, J. F. Dartigues, and H. Jacqmin-Gadda, "Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks," *Biometrics*, vol. 71, no. 1, pp. 102–113, 2015, doi: 10.1111/biom.12232.
- [13] T. A. Gerds, T. Cai, and M. Schumacher, "The performance of risk prediction models," *Biometrical J.*, vol. 50, no. 4, pp. 457–479, 2008, doi: 10.1002/bimj.200810443.
- [14] T. P. A. Debray, R. D. Riley, M. M. Rovers, J. B. Reitsma, and K. G. M. Moons, "Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance

- on Their Use," *PLoS Med.*, vol. 12, no. 10, pp. 1–12, 2015, doi: 10.1371/journal.pmed.1001886.
- [15] T. P. A. Debray *et al.*, "A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes," *Stat. Methods Med. Res.*, vol. 28, no. 9, pp. 2768–2786, 2019, doi: 10.1177/0962280218785504.
- [16] J. Ensor *et al.*, "Individual participant data meta-analysis for external validation, recalibration, and updating of a flexible parametric prognostic model," *Stat. Med.*, vol. 40, no. 13, pp. 3066–3084, 2021, doi: 10.1002/sim.8959.
- [17] I. national du cancer (INCa), "Panorama des cancers, Édition 2023," in *Cancer*, 2023. [Online]. Available: <https://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Panorama-des-cancers-en-France-edition-2023>
- [18] E. Shirin Dason, O. Koshkina, C. Chan, and M. Sobel, "Diagnosis and management of polycystic ovarian syndrome," *C. Can. Med. Assoc. J.*, vol. 196, no. 3, pp. E85–E94, 2024, doi: 10.1503/cmaj.231251.
- [19] B. S. Deweerdt, "Ovarian cancer outlook," *Nature*, vol. 600, 2021.
- [20] J. Prat, A. Ribé, and A. Gallardo, "Hereditary ovarian cancer," *Hum. Pathol.*, vol. 36, no. 8, pp. 861–870, 2005, doi: 10.1016/j.humpath.2005.06.006.
- [21] D. Ford *et al.*, "Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families," *Am. J. Hum. Genet.*, vol. 62, no. 3, pp. 676–689, 1998, doi: 10.1086/301749.
- [22] A. Antoniou *et al.*, "Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: A combined analysis of 22 studies," *Am. J. Hum. Genet.*, vol. 72, no. 5, pp. 1117–1130, 2003, doi: 10.1086/375033.
- [23] J. R. McLaughlin *et al.*, "Reproductive risk factors for ovarian cancer in carriers of BRCA1 or BRCA2 mutations: a case-control study," *Lancet Oncol.*, vol. 8, no. 1, pp. 26–34, 2007, doi: 10.1016/S1470-2045(06)70983-4.
- [24] H. T. Lynch, P. M. Lynch, S. J. Lanspa, C. L. Snyder, J. F. Lynch, and C. R. Boland, "Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medicolegal ramifications," *Clin. Genet.*, vol. 76, no. 1, pp. 1–18, 2009, doi: 10.1111/j.1399-0004.2009.01230.x.
- [25] L. Andrews and D. G. Mutch, "Hereditary Ovarian Cancer and Risk Reduction," *Best Pract. Res. Clin. Obstet. Gynaecol.*, vol. 41, pp. 31–48, 2017, doi: 10.1016/j.bpobgyn.2016.10.017.
- [26] N. Wentzensen *et al.*, "Ovarian cancer risk factors by histologic subtype: An analysis from the Ovarian Cancer Cohort Consortium," *J. Clin. Oncol.*, vol. 34, no. 24, pp. 2888–2898, 2016, doi: 10.1200/JCO.2016.66.8178.
- [27] T. Y. Sia *et al.*, "The effect of older age on treatment outcomes in women with advanced ovarian cancer receiving chemotherapy: An NRG-Oncology/Gynecologic Oncology Group (GOG-0182-ICON5) ancillary study," *Gynecol. Oncol.*, vol. 173, pp. 130–137, 2023, doi: 10.1016/j.ygyno.2023.03.018.
- [28] M. C. Pike, C. L. Pearce, R. Peters, W. Cozen, P. Wan, and A. H. Wu, "Hormonal factors and the risk of invasive ovarian cancer: A population-based case-control study," *Fertil. Steril.*, vol. 82, no. 1, pp. 186–195, 2004, doi: 10.1016/j.fertnstert.2004.03.013.
- [29] C. Group and O. Cancer, "Menopausal hormone use and ovarian cancer risk: Individual

- participant meta-analysis of 52 epidemiological studies,” *Lancet*, vol. 385, no. 9980, pp. 1835–1842, 2015, doi: 10.1016/S0140-6736(14)61687-1.
- [30] G. Chene *et al.*, “Is endometriosis a precancerous lesion? Perspectives and clinical implications,” *Gynecol. Obstet. Fertil.*, vol. 44, no. 2, pp. 106–112, 2016, doi: 10.1016/j.gyobfe.2016.01.001.
- [31] S. J. Kim *et al.*, “Weight gain and the risk of ovarian cancer in BRCA1 and BRCA2 mutation carriers,” *Cancer Epidemiol. Biomarkers Prev.*, vol. 30, no. 11, pp. 2038–2043, 2021, doi: 10.1158/1055-9965.EPI-21-0296.
- [32] M. Constanza Camargo *et al.*, “Occupational exposure to asbestos and ovarian cancer: A meta-analysis,” *Environ. Health Perspect.*, vol. 119, no. 9, pp. 1211–1217, 2011, doi: 10.1289/ehp.1003283.
- [33] C. Main, X. Chen, L. W. Chamley, M. Zhao, and Q. Chen, “Understanding How Pregnancy Protects Against Ovarian and Endometrial Cancer Development: Fetal Antigens May Be Involved,” *Endocrinol. (United States)*, vol. 163, no. 11, pp. 1–6, 2022, doi: 10.1210/endocr/bqac141.
- [34] M. Kossai, A. Leary, J. Y. Scoazec, and C. Genestie, “Ovarian Cancer: A Heterogeneous Disease,” *Pathobiology*, vol. 85, no. 1–2, pp. 41–49, 2018, doi: 10.1159/000479006.
- [35] S. Lheureux, C. Gourley, I. Vergote, and A. M. Oza, “Epithelial ovarian cancer,” *Lancet*, vol. 393, no. 10177, pp. 1240–1253, 2019, doi: 10.1016/S0140-6736(18)32552-2.
- [36] S. Pecorelli, J. L. Benedet, W. T. Creasman, and J. H. Shepherd, “FIGO staging of gynecologic cancer,” *Int. J. Gynecol. Obstet.*, vol. 64, no. 1, pp. 5–10, 1999, doi: 10.1016/S0020-7292(98)00234-3.
- [37] J. S. Berek, S. T. Kehoe, L. Kumar, and M. Friedlander, “Cancer of the ovary, fallopian tube, and peritoneum,” *Int. J. Gynecol. Obstet.*, vol. 143, pp. 59–78, 2018, doi: 10.1002/ijgo.12614.
- [38] N. Bhatla and L. Denny, “FIGO Cancer Report 2018,” *Int. J. Gynecol. Obstet.*, vol. 143, pp. 2–3, 2018, doi: 10.1002/ijgo.12608.
- [39] V. Tuninetti *et al.*, “Ki67 as a Predictor of Response to PARP Inhibitors in Platinum Sensitive BRCA Wild Type Ovarian Cancer: The MITO 37 Retrospective Study,” *Cancers (Basel)*, vol. 15, no. 4, pp. 1–15, 2023, doi: 10.3390/cancers15041032.
- [40] A. Embaby *et al.*, “WEE1 inhibitor adavosertib in combination with carboplatin in advanced TP53 mutated ovarian cancer: A biomarker-enriched phase II study,” *Gynecol. Oncol.*, vol. 174, no. May, pp. 239–246, 2023, doi: 10.1016/j.ygyno.2023.05.063.
- [41] H. Luo, X. Xu, M. Ye, B. Sheng, and X. Zhu, “The prognostic value of HER2 in ovarian cancer: A meta-analysis of observational studies,” *PLoS One*, vol. 13, no. 1, pp. 1–16, 2018, doi: 10.1371/journal.pone.0191972.
- [42] N. Colombo *et al.*, “ESMO-ESGO consensus conference recommendations on ovarian cancer: Pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease,” *Ann. Oncol.*, vol. 30, no. 5, pp. 672–705, 2019, doi: 10.1093/annonc/mdz062.
- [43] M. F. Vine, R. B. Ness, B. Calingaert, J. M. Schildkraut, and A. Berchuck, “Types and duration of symptoms prior to diagnosis of invasive or borderline ovarian tumor,” *Gynecol. Oncol.*, vol. 83, no. 3, pp. 466–471, 2001, doi: 10.1006/gyno.2001.6411.
- [44] B. A. Goff, L. S. Mandel, C. H. Melancon, and H. G. Muntz, “Frequency of symptoms of ovarian

- cancer in women presenting to primary care clinics,” *Jama*, vol. 291, no. 22, pp. 2705–2712, 2004, doi: 10.1001/jama.291.22.2705.
- [45] J. Dilley *et al.*, “Ovarian cancer symptoms, routes to diagnosis and survival – Population cohort study in the ‘no screen’ arm of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS),” *Gynecol. Oncol.*, vol. 158, no. 2, pp. 316–322, 2020, doi: 10.1016/j.ygyno.2020.05.002.
- [46] B. A. Goff *et al.*, “Development of an ovarian cancer symptom index: Possibilities for earlier detection,” *Cancer*, vol. 109, no. 2, pp. 221–227, 2007, doi: 10.1002/cncr.22371.
- [47] R. C. Bast *et al.*, “A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer,” *Obstet. Gynecol. Surv.*, vol. 39, no. 3, pp. 169–171, 1983, doi: 10.1097/00006254-198403000-00020.
- [48] R. P. Woolas *et al.*, “Elevation of multiple serum markers in patients with stage I ovarian cancer,” *J. Natl. Cancer Inst.*, vol. 85, no. 21, pp. 1748–1751, 1993, doi: 10.1093/jnci/85.21.1748.
- [49] R. J. Kurman and I. M. Shih, “The dualistic model of ovarian carcinogenesis revisited, revised, and expanded,” *Am. J. Pathol.*, vol. 186, no. 4, pp. 733–747, 2016, doi: 10.1016/j.ajpath.2015.11.011.
- [50] G. Funston, W. Hamilton, G. Abel, E. J. Crosbie, B. Rous, and F. M. Walter, “The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study,” *PLoS Med.*, vol. 17, no. 10, pp. 1–18, 2020, doi: 10.1371/journal.pmed.1003295.
- [51] S. S. Buys *et al.*, “Effect of screening on ovarian cancer mortality: The Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial,” *Jama*, vol. 305, no. 22, pp. 2295–2302, 2011, doi: 10.1001/jama.2011.766.
- [52] H. Kobayashi *et al.*, “A randomized study of screening for ovarian cancer: A multicenter study in Japan,” *Int. J. Gynecol. Cancer*, vol. 18, no. 3, pp. 414–420, 2008, doi: 10.1111/j.1525-1438.2007.01035.x.
- [53] U. Menon *et al.*, “Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial,” *Lancet*, vol. 397, no. 10290, pp. 2182–2193, 2021, doi: 10.1016/S0140-6736(21)00731-5.
- [54] J. L. Tanyi *et al.*, “A Phase III Study of Pafolacianine Injection (OTL38) for Intraoperative Imaging of Folate Receptor-Positive Ovarian Cancer (Study 006),” *J. Clin. Oncol.*, vol. 41, no. 2, pp. 276–284, 2023, doi: 10.1200/JCO.22.00291.
- [55] Y. Tazi, P. Pautier, A. Leary, and C. Lhomme, “Chimiothérapie des cancers épithéliaux de l’ovaire,” *Gynecol. Obstet. Fertil.*, vol. 41, no. 10, pp. 611–616, 2013, doi: 10.1016/j.gyobfe.2013.08.008.
- [56] T. Foo, A. George, and S. Banerjee, “PARP inhibitors in ovarian cancer: An overview of the practice-changing trials,” *Genes Chromosom. Cancer*, vol. 60, no. 5, pp. 385–397, 2020, doi: 10.1002/gcc.22935.
- [57] C. Grimm, C. Cropet, and I. Ray-Coquard, “Maintenance olaparib plus bevacizumab (bev) after platinum-based chemotherapy plus bev in patients (pts) with newly diagnosed advanced high-grade ovarian cancer (HGOC): Efficacy by timing of surgery and residual tumor status in the Phase III PAOLA-1 trial,” *Gynecol. Oncol.*, vol. 159, p. 19, Oct. 2020, doi:

10.1016/J.YGYNO.2020.06.037.

- [58] I. Vergote *et al.*, "Population-adjusted indirect treatment comparison of the SOLO1 and PAOLA-1/ENGOT-ov25 trials evaluating maintenance olaparib or bevacizumab or the combination of both in newly diagnosed, advanced BRCA-mutated ovarian cancer," *Eur. J. Cancer*, vol. 157, pp. 415–423, 2021, doi: 10.1016/j.ejca.2021.08.023.
- [59] A. González-Martín *et al.*, "Niraparib in Patients with Newly Diagnosed Advanced Ovarian Cancer," *N. Engl. J. Med.*, vol. 381, no. 25, pp. 2391–2402, 2019, doi: 10.1056/nejmoa1910962.
- [60] M. R. Mirza *et al.*, "Prospective evaluation of the tolerability and efficacy of niraparib dosing based on baseline body weight and platelet count: Results from the PRIMA/ENGOT-OV26/GOG-3012 trial," *Cancer*, vol. 129, no. 12, pp. 1846–1855, 2022, doi: 10.1002/cncr.34706.
- [61] E. C. K. W. P. Tew, C. Lacchetti, A. Ellis, K. Maxian, S. Banerjee, M. Bookman, M.B. Jones, JM Lee, S. Lheureux, J. F. Liu, K. N. Moore, C. Muller, P. Rodriguez, C. Walsh, S. N. Wes, "PARP Inhibitors in the Management of Ovarian Cancer: ASCO Guideline," *J Clin Oncol. A*, vol. 18, no. 10, pp. 647–655, 2020, doi: 10.1200/JCO.20.01924.
- [62] D. S. Vinay *et al.*, "Immune evasion in cancer: Mechanistic basis and therapeutic strategies," *Semin. Cancer Biol.*, vol. 35, pp. S185–S198, 2015, doi: 10.1016/j.semcancer.2015.03.004.
- [63] D. M. Pardoll, "The blockade of immune checkpoints in cancer immunotherapy," *Nat. Rev. Cancer*, vol. 12, no. 4, pp. 252–264, 2012, doi: 10.1038/nrc3239.
- [64] Y. Iwai, M. Ishida, Y. Tanaka, T. Okazaki, T. Honjo, and N. Minato, "Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 19, pp. 12293–12297, 2002, doi: 10.1073/pnas.192461099.
- [65] A. Varga *et al.*, "Pembrolizumab in patients with programmed death ligand 1–positive advanced ovarian cancer: Analysis of KEYNOTE-028," *Gynecol. Oncol.*, vol. 152, no. 2, pp. 243–250, 2018, doi: 10.1016/j.ygyno.2018.11.017.
- [66] G. J. S. Rustin, "Follow-up with CA125 after primary therapy of advanced ovarian cancer has major implications for treatment outcome and trial performances and should not be routinely performed," *Ann. Oncol.*, vol. 22, no. SUPPL.8, pp. 45–48, 2011, doi: 10.1093/annonc/mdr471.
- [67] G. J. S. Rustin *et al.*, "Definitions for response and progression in ovarian cancer clinical trials incorporating recist 1.1 and CA 125 agreed by the gynecological cancer intergroup (GCIG)," *Int. J. Gynecol. Cancer*, vol. 21, no. 2, pp. 419–423, 2011, doi: 10.1097/IGC.0b013e3182070f17.
- [68] R. J. Morgan *et al.*, "Ovarian cancer, version 2.2013: Featured updates to the NCCN guidelines," *JNCCN J. Natl. Compr. Cancer Netw.*, vol. 11, no. 10, pp. 1199–1209, 2013, doi: 10.6004/jnccn.2013.0142.
- [69] G. J. S. Rustin *et al.*, "Early versus delayed treatment of relapsed ovarian cancer (MRC OV05/EORTC 55955): a randomised trial," *Lancet*, vol. 376, no. 9747, pp. 1155–1163, Oct. 2010, doi: 10.1016/S0140-6736(10)61268-8.
- [70] G. J. S. Rustin and M. R. Hall, "Is CA125 useful in monitoring patients with platinum-resistant ovarian cancer?," *Ann. Oncol.*, vol. 27, no. 8, pp. 1365–1366, 2016, doi: 10.1093/annonc/mdw253.

- [71] J. R. Anderson, K. C. Cain, and R. D. Gelber, "Analysis of survival by tumor response and other comparisons of time-to-event by outcome variables," *J. Clin. Oncol.*, vol. 26, no. 24, pp. 3913–3915, 2008, doi: 10.1200/JCO.2008.16.1000.
- [72] A. Gleiss, R. Oberbauer, and G. Heinze, "An unjustified benefit: immortal time bias in the analysis of time-dependent events," *Transpl. Int.*, vol. 31, no. 2, pp. 125–130, 2017, doi: 10.1111/tri.13081.
- [73] N. M. Laird and J. H. Ware, "Random-Effects Models for Longitudinal Data Author (s): Nan M . Laird and James H . Ware Published by : International Biometric Society Stable URL : <http://www.jstor.org/stable/2529876> REFERENCES Linked references are available on JSTOR for this article," *Biometrics*, vol. 38, no. 4, pp. 963–974, 1982.
- [74] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*. Springer New York, 2009. [Online]. Available: <https://books.google.fr/books?id=jmPkX4VU7h0C>
- [75] R. D. Riley *et al.*, "External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges," *BMJ*, vol. 353, pp. 27–30, 2016, doi: 10.1136/bmj.i3140.
- [76] A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, and M. Schmid, "A review of spline function procedures in R," *BMC Med. Res. Methodol.*, vol. 19, no. 1, pp. 1–16, 2019, doi: 10.1186/s12874-019-0666-3.
- [77] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976, doi: 10.1093/biomet/63.3.581.
- [78] J. G. Ibrahim and G. Molenberghs, "Missing data methods in longitudinal studies: A review," *Test*, vol. 18, no. 1, pp. 1–43, 2009, doi: 10.1007/s11749-009-0138-x.
- [79] R. Bruno *et al.*, "Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models," *Clin. Cancer Res.*, vol. 26, no. 8, pp. 1787–1795, 2020, doi: 10.1158/1078-0432.CCR-19-0287.
- [80] M. Cerou, E. Comets, and M. Chenel, *Application of Item Response Theory to Model Disease Major Depressive Disorder Ac ce pt us cr ip t*. 2020.
- [81] S. Ducoulombier *et al.*, "Modeling CA-125 during neoadjuvant chemotherapy for predicting optimal cytoreduction and relapse risk in ovarian cancer," *Anticancer Res.*, vol. 37, no. 12, pp. 6879–6886, 2017, doi: 10.21873/anticanres.12150.
- [82] B. You *et al.*, "The strong prognostic value of KELIM , a model-based parameter from CA 125 kinetics in ovarian cancer : Data from CALYPSO trial (a GINECO-GCIG study) ☆," *Gynecol. Oncol.*, vol. 130, no. 2, pp. 289–294, 2013, doi: 10.1016/j.ygyno.2013.05.013.
- [83] B. You *et al.*, "CA-125 ELImination rate constant K (KELIM) is a marker of chemosensitivity in patients with ovarian cancer: Results from the phase II CHIVA trial," *Clin. Cancer Res.*, vol. 26, no. 17, pp. 4625–4632, 2020, doi: 10.1158/1078-0432.CCR-20-0054.
- [84] B. You *et al.*, "The role of the tumor primary chemosensitivity relative to the success of the medical-surgical management in patients with advanced ovarian carcinomas," *Cancer Treat. Rev.*, vol. 100, no. September, p. 102294, 2021, doi: 10.1016/j.ctrv.2021.102294.
- [85] B. You *et al.*, "Low probability of disease cure in advanced ovarian carcinomas before the PARP inhibitor era," *Br. J. Cancer*, vol. 127, no. 1, pp. 79–83, 2022, doi: 10.1038/s41416-022-01732-7.
- [86] A. Lauby *et al.*, "The Increasing Prognostic and Predictive Roles of the Tumor Primary

Chemosensitivity Assessed by CA-125 Elimination Rate Constant K (KELIM) in Ovarian Cancer : A Narrative Review,” *Cancers (Basel)*, 2022.

- [87] A. Carrot *et al.*, “Modeled Early Longitudinal PSA Kinetics Prognostic Value in Rising PSA Prostate Cancer Patients after Local Therapy Treated with ADT +/- Docetaxel,” *Cancers (Basel)*, vol. 14, no. 3, 2022, doi: 10.3390/cancers14030815.
- [88] O. Colombari *et al.*, “Mathematical modeling of the early modeled CA-125 longitudinal kinetics (KELIM-PARP) as a pragmatic indicator of rucaparib efficacy in patients with recurrent ovarian carcinoma in ARIEL2 & STUDY 10,” *eBioMedicine*, vol. 89, no. February, p. 104477, 2023, doi: 10.1016/j.ebiom.2023.104477.
- [89] P. M. E.L. Kaplan, “Nonparametric Estimation from Incomplete Observations Author (s): E . L . Kaplan and Paul Meier Source : Journal of the American Statistical Association , Vol . 53 , No . 282 (Jun . , 1958) , pp . 457- Published by : American Statistical Association Sta,” *Am. Stat. Assoc.*, vol. 53, no. 282, pp. 457–481, 1958.
- [90] D. R. Cox, “Regression Models and Life-Tables,” *J. R. Stat. Soc. Ser. B*, vol. 34, no. 2, pp. 187–202, Jan. 1972, doi: 10.1111/j.2517-6161.1972.tb00899.x.
- [91] J. R. Anderson, K. C. Cain, and R. D. Gelber, “Analysis of survival by tumor response,” *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.*, vol. 1, no. 11, pp. 710–719, Nov. 1983, doi: 10.1200/JCO.1983.1.11.710.
- [92] H. C. Van Houwelingen, “Dynamic prediction by landmarking in event history analysis,” *Scand. J. Stat.*, vol. 34, no. 1, pp. 70–85, 2007, doi: 10.1111/j.1467-9469.2006.00529.x.
- [93] H. Putter and H. C. van Houwelingen, “Understanding Landmarking and Its Relation with Time-Dependent Cox Regression,” *Stat. Biosci.*, vol. 9, no. 2, pp. 489–503, 2017, doi: 10.1007/s12561-016-9157-9.
- [94] M. Buyse, P. Thirion, R. W. Carlson, T. Burzykowski, G. Molenberghs, and P. Piedbois, “Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: A meta-analysis,” *Lancet*, vol. 356, no. 9227, pp. 373–378, 2000, doi: 10.1016/S0140-6736(00)02528-9.
- [95] J. Xu, R. O’QUIGLEY, “Estimating average regression effect under non-proportional hazards,” *Biostatistics*, vol. 1, no. 4, pp. 423–439, 2000, doi: 10.1093/biostatistics/1.4.423.
- [96] D. Y. Lin and L. J. Wei, “The Robust Inference for the Cox Proportional Hazards Model,” *J. Am. Stat. Assoc.*, vol. 84, no. 408, pp. 1074–1078, Oct. 1989, doi: 10.2307/2290085.
- [97] K. Mauff, E. W. Steyerberg, G. Nijpels, A. A. W. A. van der Heijden, and D. Rizopoulos, “Extension of the association structure in joint models to include weighted cumulative effects,” *Stat. Med.*, vol. 36, no. 23, pp. 3746–3759, 2017, doi: 10.1002/sim.7385.
- [98] E. R. Brown and J. G. Ibrahim, “Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials,” *Biometrics*, vol. 59, no. 3, pp. 686–693, 2003, doi: 10.1111/1541-0420.00079.
- [99] D. Rizopoulos and P. Ghosh, “A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event,” *Stat. Med.*, vol. 30, no. 12, pp. 1366–1380, 2011, doi: 10.1002/sim.4205.
- [100] M. Murawska, D. Rizopoulos, and E. Lesaffre, “A two-stage joint model for nonlinear longitudinal response and a time-to-event with application in transplantation studies,” *J. Probab. Stat.*, vol. 2012, 2012, doi: 10.1155/2012/194194.

- [101] E. R. Andrinopoulou, D. Rizopoulos, J. J. M. Takkenberg, and E. Lesaffre, "Joint modeling of two longitudinal outcomes and competing risk data," *Stat. Med.*, vol. 33, no. 18, pp. 3167–3178, 2014, doi: 10.1002/sim.6158.
- [102] R. M. Elashoff, G. Li, and N. Li, "A joint model for longitudinal measurements and survival data in the presence of multiple failure types," *Biometrics*, vol. 64, no. 3, pp. 762–771, 2008, doi: 10.1111/j.1541-0420.2007.00952.x.
- [103] M. Delporte *et al.*, "A joint normal-binary (probit) model," *Int. Stat. Rev.*, vol. 90, no. S1, pp. S37–S51, 2022, doi: 10.1111/insr.12532.
- [104] J. M. G. Taylor *et al.*, "Using Joint Models," vol. 69, no. 1, pp. 206–213, 2013, doi: 10.1111/j.1541-0420.2012.01823.x.Real-time.
- [105] L. M. McCrink, A. H. Marshall, and K. J. Cairns, "Advances in Joint Modelling: A Review of Recent Developments with Application to the Survival of End Stage Renal Disease Patients.," *Int. Stat. Rev.*, vol. 81, no. 2, pp. 249–269, 2013, [Online]. Available: <http://proxy.insermbiblio.inist.fr/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid&db=asx&AN=89981825&lang=fr&site=eds-live>
- [106] G. Papageorgiou, K. Mauff, A. Tomer, and D. Rizopoulos, "An overview of joint modeling of time-to-event and longitudinal outcomes," *Annu. Rev. Stat. Its Appl.*, vol. 6, pp. 223–240, 2019, doi: 10.1146/annurev-statistics-030718-105048.
- [107] T. Saulnier *et al.*, "Joint models for the longitudinal analysis of measurement scales in the presence of informative dropout," *Methods*, vol. 203, pp. 142–151, 2022, doi: <https://doi.org/10.1016/j.ymeth.2022.03.003>.
- [108] J. P. Pignon and C. Hill, "Meta-analyses of randomised clinical trials in oncology," *Lancet Oncol.*, vol. 2, no. 8, pp. 475–482, 2001, doi: 10.1016/S1470-2045(01)00453-3.
- [109] L. A. Stewart and M. K. B. Parmar, "Meta-analysis of the literature or of individual patient data: is there a difference?," *Lancet*, vol. 341, no. 8842, pp. 418–422, 1993, doi: 10.1016/0140-6736(93)93004-K.
- [110] HAS, "Niveau de preuve et gradation des recommandations de bonne pratique," 2013. https://www.has-sante.fr/jcms/c_1600564/fr/niveau-de-preuve-et-gradation-des-recommandations-de-bonne-pratique-etat-des-lieux
- [111] M. H. Murad, N. Asi, M. Alsawas, and F. Alahdab, "New evidence pyramid," *Evid. Based. Med.*, vol. 21, no. 4, pp. 125–127, 2016, doi: 10.1136/ebmed-2016-110401.
- [112] B. Hutton, F. Catalá-López, and D. Moher, "The PRISMA statement extension for systematic reviews incorporating network meta-analysis: PRISMA-NMA," *Med. Clínica (English Ed.)*, vol. 147, no. 6, pp. 262–266, 2016, doi: 10.1016/j.medcle.2016.10.003.
- [113] B. Levis, M. Hattle, and R. D. Riley, "PRIME-IPD SERIES Part 2. Retrieving, checking, and harmonizing data are underappreciated challenges in individual participant data meta-analyses," *J. Clin. Epidemiol.*, vol. 136, pp. 221–223, 2021, doi: 10.1016/j.jclinepi.2021.05.006.
- [114] S. Yusuf, R. Peto, J. Lewis, R. Collins, and P. Sleight, "Beta blockade during and after myocardial infarction: An overview of the randomized trials," *Prog. Cardiovasc. Dis.*, vol. 27, no. 5, pp. 335–371, 1985, doi: [https://doi.org/10.1016/S0033-0620\(85\)80003-7](https://doi.org/10.1016/S0033-0620(85)80003-7).
- [115] D. G. A. JulianPT Higgins, Simon G Thompson, Jonathan J Deeks, "Measuring inconsistency in knowledgebases," *Bmj*, vol. 27, no. 2, pp. 159–184, 2003, doi: 10.1007/s10844-006-2974-4.

- [116] J. P. T. Higgins, S. G. Thompson, and D. J. Spiegelhalter, "A re-evaluation of random-effects meta-analysis," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 172, no. 1, pp. 137–159, 2009, doi: 10.1111/j.1467-985X.2008.00552.x.
- [117] T. P. A. Debray, K. G. M. Moons, I. Ahmed, H. Koffijberg, and R. D. Riley, "A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis," *Stat. Med.*, vol. 32, no. 18, pp. 3158–3180, 2013, doi: 10.1002/sim.5732.
- [118] R. D. Riley, E. G. Elia, G. Malin, K. Hemming, and M. P. Price, "Multivariate meta-analysis of prognostic factor studies with multiple cut-points and/or methods of measurement," *Stat. Med.*, vol. 34, no. 17, pp. 2481–2496, 2015, doi: 10.1002/sim.6493.
- [119] A. J. Vickers and D. G. Altman, "Analysing controlled trials with baseline and follow up measurements," *Br. Med. J.*, vol. 323, no. 7321, pp. 1123–1124, 2001.
- [120] R. D. Riley *et al.*, "A guide to systematic review and meta-analysis of prognostic factor studies," *BMJ*, vol. 364, 2019, doi: 10.1136/bmj.k4597.
- [121] K. I. E. Snell *et al.*, "Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA)," *Bmj*, pp. 1–8, 2023, doi: 10.1136/bmj-2022-073538.
- [122] T. P. A. Debray, Y. Vergouwe, H. Koffijberg, D. Nieboer, E. W. Steyerberg, and K. G. M. Moons, "A new framework to enhance the interpretation of external validation studies of clinical prediction models," *J. Clin. Epidemiol.*, vol. 68, no. 3, pp. 279–289, 2014, doi: 10.1016/j.jclinepi.2014.06.018.
- [123] W. G. Cochran, "The Combination of Estimates from Different Experiments Published by : International Biometric Society Stable URL : <http://www.jstor.org/stable/3001666> REFERENCES Linked references are available on JSTOR for this article : You may need to log in to JSTOR," vol. 10, no. 1, pp. 101–129, 1954.
- [124] M. Buyse *et al.*, "Statistical evaluation of surrogate endpoints with examples from cancer clinical trials," *Biom. J.*, vol. 58, no. 1, pp. 104–132, Jan. 2016, doi: 10.1002/bimj.201400049.
- [125] I. Ahmed, T. P. Debray, K. G. Moons, and R. D. Riley, "Developing and validating risk prediction models in an individual participant data meta-analysis," *BMC Med. Res. Methodol.*, vol. 14, no. 1, 2014, doi: 10.1186/1471-2288-14-3.
- [126] L. A. Stewart *et al.*, "Preferred reporting items for a systematic review and meta-analysis of individual participant data: The PRISMA-IPD statement," *JAMA - J. Am. Med. Assoc.*, vol. 313, no. 16, pp. 1657–1665, 2015, doi: 10.1001/jama.2015.3656.
- [127] G. Abo-Zaid *et al.*, "Individual participant data meta-analyses should not ignore clustering," *J. Clin. Epidemiol.*, vol. 66, no. 8, pp. 865-873.e4, 2013, doi: 10.1016/j.jclinepi.2012.12.017.
- [128] P. Hougaard, *Frailty models for survival data*. 1995.
- [129] J. O'Quigley and J. Stare, "Proportional hazards models with frailties and random effects," *Stat. Med.*, vol. 21, no. 21, pp. 3219–3233, 2002, doi: 10.1002/sim.1259.
- [130] F. Dudbridge, "Empirical Bayes factors for common hypothesis tests," *PLoS One*, vol. 19, no. 2 February, pp. 1–24, 2024, doi: 10.1371/journal.pone.0297874.
- [131] P. Royston, M. K. B. Parmar, and R. Sylvester, "Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer," *Stat. Med.*, vol. 23, no. 6, pp. 907–926, 2004, doi: 10.1002/sim.1691.

- [132] E. W. Steyerberg, M. J. C. Eijkemans, J. C. Van Houwelingen, K. L. Lee, and J. D. F. Habbema, "Prognostic models based on literature and individual patient data in logistic regression analysis," *Stat. Med.*, vol. 19, no. 2, pp. 141–160, 2000, doi: 10.1002/(SICI)1097-0258(20000130)19:2<141::AID-SIM334>3.0.CO;2-O.
- [133] D. B. Toll, K. J. M. Janssen, Y. Vergouwe, and K. G. M. Moons, "Validation, updating and impact of clinical prediction rules: A review," *J. Clin. Epidemiol.*, vol. 61, no. 11, pp. 1085–1094, 2008, doi: 10.1016/j.jclinepi.2008.04.008.
- [134] X. Paoletti *et al.*, "Assessment of Progression-Free Survival as a Surrogate End Point of Overall Survival in First-Line Treatment of Ovarian Cancer: A Systematic Review and Meta-analysis," *JAMA Netw. Open*, vol. 3, no. 1, 2020, doi: 10.1001/jamanetworkopen.2019.18939.
- [135] M. Buyse *et al.*, "The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials," *Stat. Med.*, vol. 18, no. 24, pp. 3435–3451, Dec. 1999, doi: [https://doi.org/10.1002/\(SICI\)1097-0258\(19991230\)18:24<3435::AID-SIM365>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0258(19991230)18:24<3435::AID-SIM365>3.0.CO;2-O).
- [136] L. A. Stewart and J. C. Michael, "Practical methodology of meta-analyses (overviews) using updated individual patient data," *Stat. Med.*, vol. 14, no. 19, pp. 2057–2079, 1995, doi: 10.1002/sim.4780141902.
- [137] J. P. T. Higgins and S. G. Thompson, "Quantifying heterogeneity in a meta-analysis.," *Stat. Med.*, vol. 21, no. 11, pp. 1539–1558, Jun. 2002, doi: 10.1002/sim.1186.
- [138] K. Albain *et al.*, "Comparisons between different polychemotherapy regimens for early breast cancer: Meta-analyses of long-term outcome among 100 000 women in 123 randomised trials," *Lancet*, vol. 379, no. 9814, pp. 432–444, 2012, doi: 10.1016/S0140-6736(11)61625-5.
- [139] P. Blanche, A. Latouche, and V. Viallon, "Time-dependent AUC with right-censored data: a survey study," pp. 1–12, 2012, [Online]. Available: <http://arxiv.org/abs/1210.6805>
- [140] P. Blanche, J. F. Dartigues, and H. Jacqmin-Gadda, "Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks," *Stat. Med.*, vol. 32, no. 30, pp. 5381–5397, 2013, doi: 10.1002/sim.5958.
- [141] P. J. Heagerty and Y. Zheng, "Survival model predictive accuracy and ROC curves," *Biometrics*, vol. 61, no. 1, pp. 92–105, 2005, doi: 10.1111/j.0006-341X.2005.030814.x.
- [142] P. Blanche, J. F. Dartigues, and H. Jacqmin-Gadda, "Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring," *Biometrical J.*, vol. 55, no. 5, pp. 687–704, 2013, doi: 10.1002/bimj.201200045.
- [143] L. E. Chambless and G. Diao, "Estimation of time-dependent area under the ROC curve for long-term risk prediction," *Stat. Med.*, vol. 25, no. 20, pp. 3474–3486, 2006, doi: 10.1002/sim.2299.
- [144] U. Hajime, C. Tianxi, T. Lu, and L. J. Wei, "Evaluating Prediction Rules for t-Year Survivors With Censored Regression Models.," *J. Am. Stat. Assoc.*, vol. 102, no. 478, pp. 527–537, 2007, [Online]. Available: <http://proxy.insermbiblio.inist.fr/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid&db=asx&AN=25292095&lang=fr&site=eds-live>
- [145] Hung HUNG and Chin-Tsang CHIANG, "Estimation methods for time-dependent AUC models with survival data," *Can. J. of Statistics*, vol. 38, no. 1, pp. 8–26, 2010, doi: 10.1002/cjs.
- [146] G. W. BRIER, "VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY," *US*

- Weather Bur. Washington, D. C.*, vol. 78, no. 6, pp. 1–8, 1950, doi: 10.1016/0016-0032(94)90228-3.
- [147] A. H. MURPHY, “A New Vector Partition of the Probability Score,” *Natl. Cent. Atmospheric Res.*, vol. 11, no. 1, pp. 1–14, 1973, [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI
- [148] E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher, “Assessment and comparison of prognostic classification schemes for survival data,” *Stat. Med.*, vol. 18, no. 17–18, pp. 2529–2545, 1999, doi: 10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5.
- [149] D. G. Altman and P. Royston, “What do we mean by validating a prognostic model?,” *Stat. Med.*, vol. 19, no. 4, pp. 453–473, 2000, doi: 10.1002/(SICI)1097-0258(20000229)19:4<453::AID-SIM350>3.0.CO;2-5.
- [150] T. A. Gerds and M. Schumacher, “Efron-type measures of prediction error for survival analysis,” *Biometrics*, vol. 63, no. 4, pp. 1283–1287, 2007, doi: 10.1111/j.1541-0420.2007.00832.x.
- [151] E. W. Steyerberg, S. E. Bleeker, H. A. Moll, D. E. Grobbee, and K. G. M. Moons, “Internal and external validation of predictive models: A simulation study of bias and precision in small samples,” *J. Clin. Epidemiol.*, vol. 56, no. 5, pp. 441–447, 2003, doi: 10.1016/S0895-4356(03)00047-7.
- [152] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*. Springer New York, 1992. [Online]. Available: <http://proxy.insermbiblio.inist.fr/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid&db=edssjb&AN=edssjb.978.1.4612.4380.9.41&lang=fr&site=eds-live>
- [153] P. J. M. Verweij, H. C. V. A. N. Houwelingen, H. C. Van Houwelingen, and H. C. V. A. N. Houwelingen, “Cross-Validation I N Survival Analysis,” *Stat. Med.*, vol. 12, pp. 2305–2314, 1993.
- [154] M. Schumacher, N. Holländer, and W. Sauerbrei, “Resampling and cross-validation techniques: A tool to reduce bias caused by model building?,” *Stat. Med.*, vol. 16, no. 24, pp. 2813–2827, 1997, doi: 10.1002/(SICI)1097-0258(19971230)16:24<2813::AID-SIM701>3.0.CO;2-Z.
- [155] R. Simon, “Development and validation of therapeutically relevant multi-gene biomarker classifiers,” *J. Natl. Cancer Inst.*, vol. 97, no. 12, pp. 866–867, 2005, doi: 10.1093/jnci/dji168.
- [156] K. G. M. Moons *et al.*, “Risk prediction models: II. External validation, model updating, and impact assessment,” *Heart*, vol. 98, no. 9, pp. 691–698, 2012, doi: 10.1136/heartjnl-2011-301247.
- [157] K. I. E. Snell *et al.*, “Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model,” *J. Clin. Epidemiol.*, vol. 69, no. 2016, pp. 40–50, 2015, doi: 10.1016/j.jclinepi.2015.05.009.
- [158] E. W. Steyerberg *et al.*, “Predicting outcome after traumatic brain injury: Development and international validation of prognostic scores based on admission characteristics,” *PLoS Med.*,

- vol. 5, no. 8, pp. 1251–1261, 2008, doi: 10.1371/journal.pmed.0050165.
- [159] J. A. Doherty, L. C. Peres, C. Wang, P. Way, C. S. Greene, and J. M. Schildkraut, “Challenges and Opportunities in Studying the Epidemiology of Ovarian Cancer Subtypes,” *Curr Epidemiol Rep*, vol. 4, no. 3, pp. 211–220, 2017, doi: 10.1007/s40471-017-0115-y.Challenges.
- [160] I. Ray-Coquard *et al.*, “Olaparib plus Bevacizumab as First-Line Maintenance in Ovarian Cancer,” *N. Engl. J. Med.*, vol. 381, no. 25, pp. 2416–2428, 2019, doi: 10.1056/nejmoa1911361.
- [161] S. Banerjee *et al.*, “Maintenance olaparib for patients with newly diagnosed advanced ovarian cancer and a BRCA mutation (SOLO1/GOG 3004): 5-year follow-up of a randomised, double-blind, placebo-controlled, phase 3 trial,” *Lancet Oncol.*, vol. 22, no. 12, 2021, doi: 10.1016/S1470-2045(21)00531-3.
- [162] P. T. Taylor, D. M. Haverstick, and G. J. S. Rustin, “Re: New guidelines to evaluate the response to treatment in solid tumors (ovarian cancer) (multiple letters) [1],” *J. Natl. Cancer Inst.*, vol. 97, no. 2, pp. 151–152, 2005, doi: 10.1093/jnci/dji027.
- [163] G. J. S. Rustin *et al.*, “Re: New Guidelines to Evaluate the Response to Treatment in Solid Tumors (Ovarian Cancer),” *J. Natl. Cancer Inst.*, vol. 96, no. 6, pp. 487–488, 2000, doi: 10.1093/jnci/djh081.
- [164] K. M. Esselen *et al.*, “Use of CA-125 tests and computed tomographic scans for surveillance in ovarian cancer,” *JAMA Oncol.*, vol. 2, no. 11, pp. 1427–1433, 2016, doi: 10.1001/jamaoncol.2016.1842.
- [165] O. Colombari *et al.*, “Early modeled longitudinal CA-125 kinetics and survival of ovarian cancer patients: A GINECO AGO MRC CTU Study,” *Clin. Cancer Res.*, vol. 25, no. 17, pp. 5342–5350, 2019, doi: 10.1158/1078-0432.CCR-18-3335.
- [166] L. Ferrer, H. Putter, and C. Proust-Lima, “Individual dynamic predictions using landmarking and joint modelling: Validation of estimators and robustness assessment,” *Stat. Methods Med. Res.*, vol. 28, no. 12, pp. 3649–3666, 2019, doi: 10.1177/0962280218811837.
- [167] T. A. Gerds, T. Cai, and M. Schumacher, “The performance of risk prediction models,” *Biometrical J.*, vol. 50, no. 4, pp. 457–479, 2008, doi: 10.1002/bimj.200810443.
- [168] C. Tudur Smith, P. R. Williamson, and A. G. Marson, “Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes,” *Stat. Med.*, vol. 24, no. 9, pp. 1307–1319, 2005, doi: 10.1002/sim.2050.
- [169] T. A. Gerds, T. H. Scheike, and P. K. Andersen, “Absolute risk regression for competing risks: Interpretation, link functions, and prediction,” *Stat. Med.*, vol. 31, no. 29, pp. 3921–3930, 2012, doi: 10.1002/sim.5459.
- [170] C. T. Smith, P. R. Williamson, and A. G. Marson, “An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes,” *J. Eval. Clin. Pract.*, vol. 11, no. 5, pp. 468–478, 2005, doi: 10.1111/j.1365-2753.2005.00559.x.
- [171] S. I. Kim *et al.*, “Effect of BRCA mutational status on survival outcome in advanced-stage high-grade serous ovarian cancer,” *J. Ovarian Res.*, vol. 12, no. 1, pp. 1–10, 2019, doi: 10.1186/s13048-019-0511-7.
- [172] P. Giampaolino, V. Foreste, L. Della Corte, C. Di Filippo, G. Iorio, and G. Bifulco, “Role of biomarkers for early detection of ovarian cancer recurrence,” *Gland Surg.*, vol. 9, no. 4, pp.

1102–1111, 2020, doi: 10.21037/gs-20-544.

- [173] A. Davis, A. V. Tinker, and M. Friedlander, “‘platinum resistant’ ovarian cancer: What is it, who to treat and how to measure benefit?,” *Gynecol. Oncol.*, vol. 133, no. 3, pp. 624–631, 2014, doi: 10.1016/j.ygyno.2014.02.038.
- [174] K. A. Lawson-Michod *et al.*, “Pathways to ovarian cancer diagnosis: a qualitative study,” *BMC Womens. Health*, vol. 22, no. 1, pp. 1–16, 2022, doi: 10.1186/s12905-022-02016-1.
- [175] M. J. Flynn and J. A. Ledermann, “Ovarian cancer recurrence: is the definition of platinum resistance modified by PARPi and other intervening treatments? The evolving landscape in the management of platinum-resistant ovarian cancer,” *Cancer Drug Resist.*, pp. 424–435, 2022, doi: 10.20517/cdr.2022.13.
- [176] B. Van Zyl, D. Tang, and N. A. Bowden, “Biomarkers of platinum resistance in ovarian cancer: What can we use to improve treatment,” *Endocr. Relat. Cancer*, vol. 25, no. 5, pp. R303–R318, 2018, doi: 10.1530/ERC-17-0336.
- [177] G. J. S. Rustin and M. R. Hall, “Is CA125 useful in monitoring patients with platinum-resistant ovarian cancer?,” *Ann. Oncol.*, vol. 27, no. 8, pp. 1365–1366, 2016, doi: 10.1093/annonc/mdw253.
- [178] C. K. Lee *et al.*, “Prognostic nomogram to predict progression-free survival in patients with platinum-sensitive recurrent ovarian cancer,” *Br. J. Cancer*, vol. 105, no. 8, pp. 1144–1150, 2011, doi: 10.1038/bjc.2011.364.
- [179] M. Markman, M. Federico, P. Y. Liu, E. Hannigan, and D. Alberts, “Significance of early changes in the serum CA-125 antigen level on overall survival in advanced ovarian cancer,” *Gynecol. Oncol.*, vol. 103, no. 1, pp. 195–198, Oct. 2006, doi: 10.1016/j.ygyno.2006.02.024.
- [180] M. E. Skaznik-Wikiel *et al.*, “Possible use of CA-125 level normalization after the third chemotherapy cycle in deciding on chemotherapy regimen in patients with epithelial ovarian cancer,” *Int. J. Gynecol. Cancer*, vol. 21, no. 6, pp. 1013–1017, Aug. 2011, doi: 10.1097/IGC.0b013e31821ce903.
- [181] A. Pelissier *et al.*, “Dynamic analysis of CA125 decline during neoadjuvant chemotherapy in patients with epithelial ovarian cancer as a predictor for platinum sensitivity,” *Anticancer Res.*, vol. 36, no. 4, pp. 1865–1871, 2016.
- [182] E. Karamouza *et al.*, “CA-125 Early Dynamics to Predict Overall Survival in Women with Newly Diagnosed Advanced Ovarian Cancer Based on,” *Cancers (Basel)*, vol. 15, no. 1823, 2023, doi: <https://doi.org/10.3390/cancers15061823>.
- [183] P. Corbaux *et al.*, “Survival and modelled cancer antigen-125 ELIMination rate constant K score in ovarian cancer patients in first-line before poly(ADP-ribose) polymerase inhibitor era: A Gynaecologic Cancer Intergroup meta-analysis,” *Eur. J. Cancer*, vol. 191, p. 112966, 2023, doi: 10.1016/j.ejca.2023.112966.
- [184] J. M. T. and H. J.-G. Cécile Proust-Lima, Mbéry Séne, “Joint latent class models for longitudinal and time-to-event data: A review,” *Stat. Methods Med. Res.*, vol. 176, no. 23, pp. 74–90, 2012, doi: <https://doi.org/10.1177/0962280212445839>.
- [185] A. Devaux, R. Genuer, K. Peres, and C. Proust-Lima, “Individual dynamic prediction of clinical endpoint from large dimensional longitudinal biomarker history: a landmark approach,” *BMC Med. Res. Methodol.*, vol. 22, no. 1, pp. 1–14, 2022, doi: 10.1186/s12874-022-01660-3.
- [186] S. J. E. Stock *et al.*, “Development and validation of a risk prediction model of preterm birth

- for women with preterm labour symptoms (the QUIDS study): A prospective cohort study and individual participant data meta-analysis," *PLoS Med.*, vol. 18, no. 7, pp. 1–22, 2021, doi: 10.1371/journal.pmed.1003686.
- [187] H. H. and C.-T. CHIANG, "Estimation methods for time-dependent AUC models with survival data," *Can. J. Stat.*, vol. 38, no. 1, pp. 8–26, 2010, doi: 10.1002/cjs.
- [188] M. . Gerds, T.A., & Kattan, *Medical Risk Prediction Models: With Ties to Machine Learning*, 1st ed. 2021. doi: <https://doi.org/10.1201/9781138384484>.
- [189] D. Bates, M. Mächler, B. M. Bolker, and S. C. Walker, "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.*, vol. 67, no. 1, 2015, doi: 10.18637/jss.v067.i01.
- [190] R. C. Arend *et al.*, "Ofranergene Obadenovec (Ofra-Vec, VB-111) With Weekly Paclitaxel for Platinum-Resistant Ovarian Cancer: Randomized Controlled Phase III Trial (OVAL Study/GOG 3018)," *J. Clin. Oncol.*, vol. 42, no. 2, pp. 170–179, 2024, doi: 10.1200/JCO.22.02915.
- [191] D. R. Nebgen, K. H. Lu, and R. C. Bast, "Novel Approaches to Ovarian Cancer Screening," *Curr. Oncol. Rep.*, vol. 21, no. 8, 2020, doi: 10.1007/s11912-019-0816-0.
- [192] E. Potenza, G. Parpinel, M. E. Laudani, C. Macchi, L. Fuso, and P. Zola, "Prognostic and predictive value of combined HE-4 and CA-125 biomarkers during chemotherapy in patients with epithelial ovarian cancer," *Int. J. Biol. Markers*, vol. 35, no. 4, pp. 20–27, 2020, doi: 10.1177/1724600820955195.
- [193] M. P. Furrer Daniela, Jean Grégoire, Stéphane Turcotte, P. D. Dimcho Bachvarov, Dominique Trudel, Bernard Têtu, and B. Isabelle, "Performance of preoperative plasma HE4 and CA-125 levels in predicting ovarian cancer mortality in women with epithelial ovarian cancer (EOC).," *PLoS One*, vol. 35, no. 15_suppl, pp. e17076–e17076, 2019, doi: 10.1200/jco.2017.35.15_suppl.e17076.
- [194] Y. Rong and L. Li, "Early clearance of serum HE4 and CA125 in predicting platinum sensitivity and prognosis in epithelial ovarian cancer," *J. Ovarian Res.*, vol. 14, no. 1, pp. 1–9, 2021, doi: 10.1186/s13048-020-00759-9.
- [195] A. Samborski *et al.*, "HE4 and CA125 serum biomarker monitoring in women with epithelial ovarian cancer," *Tumor Biol.*, vol. 44, no. 1, pp. 205–213, 2022, doi: 10.3233/TUB-220016.
- [196] M. Zhang, S. Cheng, Y. Jin, Y. Zhao, and Y. Wang, "Roles of CA125 in diagnosis, prediction, and oncogenesis of ovarian cancer," *Biochim. Biophys. Acta - Rev. Cancer*, vol. 1875, no. 2, p. 188503, 2021, doi: 10.1016/j.bbcan.2021.188503.
- [197] H. Y. S. N. and K. K. L. C. Ruiqian Zhang , Michelle K. Y. Siu, "Molecular Biomarkers for the Early Detection of Ovarian Cancer," *Int. J. Mol. Sci.*, vol. 23, no. 12041, 2022, doi: <https://doi.org/10.3390/ijms231912041>.
- [198] N. Colombo *et al.*, "ESMO-ESGO consensus conference recommendations on ovarian cancer: Pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease," *Int. J. Gynecol. Cancer*, vol. 29, no. 4, pp. 728–760, 2019, doi: 10.1136/ijgc-2019-000308.
- [199] O. Colombari, A. Clamp, A. Cook, I. A. McNeish, and B. You, "Benefit From Fractionated Dose-Dense Chemotherapy in Patients With Poor Prognostic Ovarian Cancer: ICON-8 Trial," *JCO Clin. Cancer Informatics*, no. 7, pp. 1–11, 2023, doi: 10.1200/cci.22.00188.
- [200] D. G. Altman, Y. Vergouwe, P. Royston, and K. G. M. Moons, "Prognosis and prognostic research: validating a prognostic model," *BMJ*, vol. 338, 2009, doi: 10.1136/bmj.b605.

- [201] R. J. Bauer, "NONMEM Tutorial Part II: Estimation Methods and Advanced Examples," *CPT Pharmacometrics Syst. Pharmacol.*, vol. 8, no. 8, pp. 538–556, 2019, doi: 10.1002/psp4.12422.
- [202] M. C. Simmonds, J. P. T. Higginsa, L. A. Stewartb, J. F. Tierneyb, M. J. Clarke, and S. G. Thompson, "Meta-analysis of individual patient data from randomized trials: a review of methods used in practice," *Clin. Trials*, vol. 2, no. 3, pp. 209–217, Jun. 2005, doi: 10.1191/1740774505cn087oa.
- [203] P. Pautier *et al.*, "Prise en charge médicale de la récurrence du cancer épithélial de l'ovaire: Medical management of recurrent epithelial ovarian cancer," *Bull. Cancer*, vol. 108, no. 9, pp. S22–S32, 2021, doi: 10.1016/S0007-4551(21)00584-1.
- [204] L. Pr, B. You, and C. De Lyon, "Cancer de l'ovaire : les HCL révolutionnent la prise en charge des patientes grâce à un algorithme permettant de mieux anticiper la réponse au traitement," pp. 72–74, 2023.
- [205] R. G. Moore *et al.*, "The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass," *Gynecol. Oncol.*, vol. 108, no. 2, pp. 402–408, 2008, doi: 10.1016/j.ygyno.2007.10.017.
- [206] J. J. B. and S. J. S. Richard G. Moore, b, M. Craig Miller, Paul Disilvestro, Lisa M. Landrum, Walter Gajewski, "Evaluation of the Diagnostic Accuracy of the Risk of Ovarian Malignancy Algorithm in Women With a Pelvic Mass," *Obs. Gynecol.*, vol. 118, no. 401, pp. 280–288, 2011, doi: 10.1097/AOG.0b013e318224fce2.Evaluation.
- [207] J. Riedinger *et al.*, "Evaluation clinique du cancer associated cerum antigen (CASA) comme marqueur tumoral des cancers ovariens: comparaison avec le CA 125," *Immuno anal Biol*, 2000.
- [208] F. Rotolo, X. Paoletti, and S. Michiels, "surrosurv: An R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials," *Comput. Methods Programs Biomed.*, vol. 155, pp. 189–198, Mar. 2018, doi: 10.1016/J.CMPB.2017.12.005.
- [209] I. B. Vergote *et al.*, "Randomized phase III study of erlotinib versus observation in patients with no evidence of disease progression after first-line platin-based chemotherapy for ovarian carcinoma: A European Organisation for Research and Treatment of Cancer-Gynaecological Ca," *J. Clin. Oncol.*, vol. 32, no. 4, pp. 320–326, 2014, doi: 10.1200/JCO.2013.50.5669.
- [210] H. Hirte *et al.*, "A phase III randomized trial of BAY 12-9566 (tanomastat) as maintenance therapy in patients with advanced ovarian cancer responsive to primary surgery and paclitaxel/platinum containing chemotherapy: A National Cancer Institute of Canada Clinical Trials G," *Gynecol. Oncol.*, vol. 102, no. 2, pp. 300–308, 2006, doi: 10.1016/j.ygyno.2005.12.020.
- [211] A. K. L. Reyners *et al.*, "A randomized phase ii study investigating the addition of the specific COX-2 inhibitor celecoxib to docetaxel plus carboplatin as first-line chemotherapy for stage IC to IV epithelial ovarian cancer, fallopian tube or primary peritoneal carcinomas: The Do," *Ann. Oncol.*, vol. 23, no. 11, pp. 2896–2902, 2012, doi: 10.1093/annonc/mds107.
- [212] A. M. Oza *et al.*, "Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): Overall survival results of a phase 3 randomised trial," *Lancet Oncol.*, vol. 16, no. 8, pp. 928–936, 2015, doi: 10.1016/S1470-2045(15)00086-8.
- [213] R. Mannel *et al.*, "A Randomized Phase III Trial of IV Carboplatin and Paclitaxel x 3 Courses Followed by Observation Versus Weekly Maintenance Low Dose Paclitaxel in Patients with

- Early Stage Ovarian Carcinoma: a Gynecologic Oncology Group Study," *Gynecol Oncol*, vol. 122, no. 1, pp. 89–94, 2011, doi: 10.1016/j.ygyno.2011.03.013.A.
- [214] G. Aravantinos *et al.*, "Carboplatin and Paclitaxel versus Cisplatin, Paclitaxel and Doxorubicin for first-line chemotherapy of Advanced Ovarian Cancer: A Hellenic Cooperative Oncology Group (HeCOG) study," *Eur. J. Cancer*, vol. 44, no. 15, pp. 2169–2177, 2008, doi: 10.1016/j.ejca.2008.06.035.
- [215] S. Pignata *et al.*, "Carboplatin plus paclitaxel versus carboplatin plus pegylated liposomal doxorubicin as first-line treatment for patients with ovarian cancer: The MITO-2 randomized phase III trial," *J. Clin. Oncol.*, vol. 29, no. 27, pp. 3628–3635, 2011, doi: 10.1200/JCO.2010.33.8566.
- [216] P. A. Vasey *et al.*, "Phase III randomized trial of docetaxel-carboplatin versus paclitaxel-carboplatin as first-line chemotherapy for ovarian carcinoma," *J. Natl. Cancer Inst.*, vol. 96, no. 22, pp. 1682–1691, 2004, doi: 10.1093/jnci/djh323.
- [217] T. Sugiyama *et al.*, "Randomized phase III trial of irinotecan plus cisplatin compared with paclitaxel plus carboplatin as first-line chemotherapy for ovarian clear cell carcinoma: JGOG3017/GCIG trial," *J. Clin. Oncol.*, vol. 34, no. 24, pp. 2881–2887, 2016, doi: 10.1200/JCO.2016.66.9010.
- [218] P. Hoskins *et al.*, "Advanced ovarian cancer: Phase III randomized study of sequential cisplatin-topotecan and carboplatin-paclitaxel vs carboplatin-paclitaxel," *J. Natl. Cancer Inst.*, vol. 102, no. 20, pp. 1547–1556, 2010, doi: 10.1093/jnci/djq362.
- [219] K. Lindemann *et al.*, "First-line treatment of advanced ovarian cancer with paclitaxel/carboplatin with or without epirubicin (TEC versus TC)-a gynecologic cancer intergroup study of the NSGO, EORTC GCG and NCIC CTG," *Ann. Oncol.*, vol. 23, no. 10, pp. 2613–2619, 2012, doi: 10.1093/annonc/mds060.
- [220] R. Fruscio *et al.*, "A phase II randomised clinical trial comparing cisplatin, paclitaxel and ifosfamide with cisplatin, paclitaxel and epirubicin in newly diagnosed advanced epithelial ovarian cancer: Long-term survival analysis," *Br. J. Cancer*, vol. 98, no. 4, pp. 720–727, 2008, doi: 10.1038/sj.bjc.6604231.
- [221] I. Ray-Coquard *et al.*, "Intensified dose of cyclophosphamide with G-CSF support versus standard dose combined with platinum in first-line treatment of advanced ovarian cancer a randomised study from the GINECO group," *Br. J. Cancer*, vol. 97, no. 9, pp. 1200–1205, 2007, doi: 10.1038/sj.bjc.6604026.
- [222] S. Pignata *et al.*, "Carboplatin plus paclitaxel once a week versus every 3 weeks in patients with advanced ovarian cancer (MITO-7): A randomised, multicentre, open-label, phase 3 trial," *Lancet Oncol.*, vol. 15, no. 4, pp. 396–405, 2014, doi: 10.1016/S1470-2045(14)70049-X.
- [223] S. Banerjee *et al.*, "A multicenter, randomized trial of flat dosing versus inpatient dose escalation of single-agent carboplatin as first-line chemotherapy for advanced ovarian cancer: An SGCTG (SCOTROC 4) and ANZGOG study on behalf of GCIG," *Ann. Oncol.*, vol. 24, no. 3, pp. 679–687, 2013, doi: 10.1093/annonc/mds494.
- [224] N. Katsumata *et al.*, "Long-term results of dose-dense paclitaxel and carboplatin versus conventional paclitaxel and carboplatin for treatment of advanced epithelial ovarian, fallopian tube, or primary peritoneal cancer (JGOG 3016): A randomised, controlled, open-label trial," *Lancet Oncol.*, vol. 14, no. 10, pp. 1020–1026, 2013, doi: 10.1016/S1470-2045(13)70363-2.
- [225] M. E. L. Van Der Burg *et al.*, "Long-term results of a randomised phase III trial of weekly versus

three-weekly paclitaxel/platinum induction therapy followed by standard or extended three-weekly paclitaxel/platinum in European patients with advanced epithelial ovarian cancer,” *Eur. J. Cancer*, vol. 50, no. 15, pp. 2592–2601, 2014, doi: 10.1016/j.ejca.2014.07.015.