
Année 2013

Thèse n° : 2115

THÈSE

Pour l'obtention du grade de

Docteur de l'Université Bordeaux 2

Diplôme National - Arrêté du 7 août 2006

École doctorale Sociétés, Politique, Santé Publique

Spécialité : Santé Publique

Option : Biostatistiques

Présentée par

Mbéry SÈNE

DÉVELOPPEMENT D'OUTILS PRONOSTIQUES DYNAMIQUES DANS LE CANCER DE LA PROSTATE LOCALISÉ TRAITÉ PAR RADIOTHÉRAPIE

Soutenue le 13 Décembre 2013 devant la Commission d'examen :

Mr Daniel COMMENGES	Directeur de recherche	Président
Mr Roch GIORGI	Professeur d'Université	Rapporteur
Mr Xavier PAOLETTI	Biostatisticien	Rapporteur
Mme Catherine LEGRAND	Professeur d'Université	Examinatrice
Mme Simone MATHOULIN-PELISSIER	Professeur d'Université	Examinatrice
Mme Cécile PROUST-LIMA	Chargée de recherche	Directrice de thèse

Thèse préparée au sein de
l'équipe Biostatistique
de l'unité INSERM U897 ISPED
146 rue Léo Saignat 33 076 Bordeaux CEDEX

"La science est le tronc d'un baobab qu'une seule personne ne peut embrasser."

Proverbe africain.

A ma famille

Remerciements

A Monsieur Daniel Commenges pour son accueil au sein de cette équipe de Biostatistiques et qui me fait l'honneur de présider mon jury de thèse.

A Messieurs Roch Giorgi et Xavier Paoletti d'avoir accepté d'être les rapporteurs de ce travail.

A Madame Catherine Legrand de me faire le plaisir d'examiner ce travail de thèse.

A Madame Simone Mathoulin-Pelissier de me faire l'honneur de juger mon travail. Vous êtes une référence dans la recherche sur le cancer.

A Madame Cécile Proust-Lima de m'avoir confiée ce travail de thèse. Je te remercie profondément de ton aide qui a été si précieuse, de tes conseils, de ta disponibilité, de ta gentillesse et de ton humanité tout simplement. J'ai beaucoup appris pendant ces trois années et t'en serai reconnaissante pour toujours.

A l'INCA, l'Institut National du CAncer, pour avoir financé cette thèse dans le cadre du projet PREDYC.

A Hélène Jacqmin-Gadda, pour ton aide et tes conseils quand j'en ai eu besoin.

A Carine Bellera, pour tes conseils et tes encouragements. Merci également de m'avoir relue.

A toutes les personnes rencontrées ici à l'ISPED, qui sont encore là ou qui ne sont plus présentes. Mention spéciale à Célia et Jérémie, c'était un vrai plaisir d'avoir partagé le bureau avec vous pendant ces trois années.

A toutes ces personnes qui ont été là pour moi quand je suis arrivée dans ce pays il y a neuf ans maintenant, et sans qui, en partie, je n'en serai pas là.

Aux amis. Merci d'être toujours là.

A mes frères et sœurs, à toute ma famille. Merci à mes parents de m'avoir laissée partir en France pour continuer mes études alors que j'étais à peine majeure. Merci de m'avoir toujours encouragée et accompagnée malgré la distance qui nous séparait. Ces années ont été longues sans vous à mes côtés. Merci pour tout, du plus profond de mon cœur.

Et enfin à Guillaume. Merci pour tes relectures et tes corrections. Merci d'avoir été là dans les bons moments mais aussi dans les moments de doutes, d'inquiétude, de frustration. Merci pour ton soutien et tes encouragements inestimables. Du plus profond de mon cœur, merci d'être là.

Résumé : La prédiction d'un événement clinique à l'aide d'outils pronostiques est une question centrale en oncologie. L'émergence des biomarqueurs mesurés au cours du temps permet de proposer des outils incorporant les données répétées de ces biomarqueurs pour mieux guider le clinicien dans la prise en charge des patients. L'objectif de ce travail est de développer et valider des outils pronostiques dynamiques de rechute de cancer de la prostate, chez des patients traités initialement par radiothérapie externe, en prenant en compte les données répétées du PSA, l'antigène spécifique de la prostate, en plus des facteurs pronostiques standard. Ces outils sont dynamiques car ils peuvent être mis à jour à chaque nouvelle mesure disponible du biomarqueur. Ils sont construits à partir de modèles conjoints pour données longitudinales et de temps d'événement. Le principe de la modélisation conjointe est de décrire l'évolution du biomarqueur à travers un modèle linéaire mixte, décrire le risque d'événement à travers un modèle de survie et lier ces deux processus à travers une structure latente. Deux approches existent, les modèles conjoints à effets aléatoires partagés et les modèles conjoints à classes latentes. Dans un premier travail, nous avons tout d'abord comparé, en terme de qualité d'ajustement et de pouvoir prédictif, des modèles conjoints à effets aléatoires partagés différant par leur forme de dépendance entre le PSA et le risque de rechute clinique. Puis nous avons évalué et comparé ces deux approches de modélisation conjointe. Dans un deuxième travail, nous avons proposé un outil pronostique dynamique différentiel permettant d'évaluer le risque de rechute clinique suivant l'initiation ou non d'un second traitement (un traitement hormonal) au cours du suivi. Dans ces travaux, la validation de l'outil pronostique a été basée sur deux mesures de pouvoir prédictif : le score de Brier et l'entropie croisée pronostique. Dans un troisième travail, nous avons enfin décrit la dynamique des PSA après un second traitement de type hormonal chez des patients traités initialement par une radiothérapie seule.

Mots clés : Prédiction dynamique, précision prédictive, score de Brier, entropie croisée pronostique, modélisation conjointe, cancer de la prostate.

Abstract : The prediction of a clinical event with prognostic tools is a central issue in oncology. The emergence of biomarkers measured over time can provide tools incorporating repeated data of these biomarkers to better guide the clinician in the management of patients. The objective of this work is to develop and validate dynamic prognostic tools of recurrence of prostate cancer in patients initially treated by external beam radiation therapy, taking into account the repeated data of PSA, the Prostate-Specific Antigen, in addition to standard prognostic factors. These tools are dynamic because they can be updated at each available new measurement of the biomarker. They are built from joint models for longitudinal and time-to-event data. The principle of joint modelling is to describe the evolution of the biomarker through a linear mixed model, describe the risk of event through a survival model and link these two processes through a latent structure. Two approaches exist, shared random-effect models and joint latent class models. In a first study, we first compared in terms of goodness-of-fit and predictive accuracy shared random-effect models differing in the form of dependency between the PSA and the risk of clinical recurrence. Then we have evaluated and compared these two approaches of joint modelling. In a second study, we proposed a differential dynamic prognostic tool to evaluate the risk of clinical recurrence according to the initiation or not of a second treatment (an hormonal treatment) during the follow-up. In these works, validation of the prognostic tool was based on two measures of predictive accuracy : the Brier score and the prognostic cross-entropy. In a third study, we have described the PSA dynamics after a second treatment (hormonal) in patients initially treated by a radiation therapy alone.

Key words : Dynamic prediction, predictive accuracy, Brier score, prognostic cross-entropy, joint modelling, prostate cancer.

Table des matières

Table des figures	xiii
Liste des tableaux	xix
I Contexte épidémiologique	1
1 Cancer de la prostate	1
1.1 La prostate	1
1.2 Le cancer	2
1.2.1 L'épidémiologie	2
1.2.2 Le diagnostic	3
1.2.3 Les facteurs de risques	5
1.2.4 Les stades	6
1.2.5 Les traitements	8
2 Le biomarqueur Prostate-Specific-Antigen (PSA)	11
2.1 PSA et diagnostic de cancer	11
2.2 PSA et progression du cancer	12
2.3 PSA et rechute biochimique post-radiothérapie	14
3 Rechute clinique de cancer de la prostate	16

3.1	PSA et rechute clinique	16
3.2	Prédictions individuelles de rechute	16
3.3	L'initiation d'hormonothérapie au cours du suivi	17
4	Objectifs de la thèse	19
II	Etat de l'art	21
1	Analyse de données longitudinales	21
1.1	Modèle linéaire mixte	21
1.2	Estimation des paramètres	23
1.3	Prédictions et résidus	24
1.4	Données manquantes dans les études longitudinales	25
2	Analyse de données de survie	27
2.1	Quelques définitions	28
2.2	Modèle à risques proportionnels	29
2.3	Estimation	30
2.4	Résidus de martingale et de Cox-Snell	31
2.5	Inclusion de covariables dépendant du temps	32
3	Modélisation conjointe pour données longitudinales et de temps d'événement	34
3.1	Approche "Two-stage" : les prémisses de la modélisation conjointe .	34
3.2	Principe de la modélisation conjointe	36
3.3	Modèles conjoints à effets aléatoires partagés	38
3.3.1	Spécification du sous-modèle longitudinal	39
3.3.2	Spécification du sous-modèle de survie	40
3.3.3	Estimation du modèle	41
3.3.4	Évaluation du modèle	45
3.3.5	Extensions du modèle	45
3.3.6	Domaines d'application du modèle	46
3.3.7	Limites du modèle	47

3.4	Modèles conjoints à classes latentes	48
3.4.1	Spécification du modèle	49
3.4.2	Estimation du modèle	51
3.4.3	Classification a posteriori	53
3.4.4	Évaluation du modèle	54
3.4.5	Extensions et limites du modèle	55
4	Outils pronostiques dynamiques	56
4.1	Prédictions individuelles dynamiques	57
4.2	Évaluation de l'outil	60
4.2.1	Erreur de prédiction quadratique ou score de Brier	61
4.2.2	L'entropie croisée pronostique	63
4.2.3	Validation interne/externe	65

III Prédiction dynamique du risque de rechute clinique de cancer de la prostate : comparaison d'approches

		67
1	Introduction	67
2	Article	69
2.1	Introduction	73
2.2	Shared random-effect models	76
2.2.1	Longitudinal submodel	77
2.2.2	Survival submodel	78
2.2.3	Maximum likelihood estimation	79
2.3	Dynamic predictions	81
2.3.1	Individual prediction	81
2.3.2	Dynamic prognostic tool	81
2.3.3	Evaluation of the predictive accuracy	82
2.4	Application to prostate cancer data	84
2.4.1	Description of the dataset	85

2.4.2	Specification of the joint models	87
2.4.3	Joint models estimations	89
2.4.4	Goodness-of-fit assessment	91
2.4.5	Predictive accuracy assessment	96
2.4.6	Individual predictions	98
2.5	Conclusion	98
3	Comparaison entre les modèles à effets aléatoires partagés et les modèles à classes latentes	102
4	Conclusion	107

IV Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l’initiation d’un second traitement au cours du suivi **109**

1	Introduction	109
2	Article	111
2.1	Introduction	115
2.2	Joint models	117
2.2.1	Notation	118
2.2.2	Longitudinal submodel	118
2.2.3	Survival submodel	119
2.2.4	Maximum likelihood estimation	119
2.3	Individual dynamic predictions	120
2.4	Evaluation of predictive accuracy	121
2.4.1	Measures of predictive accuracy	122
2.4.2	Approximated cross-validated estimators	123
2.4.3	Averaged predictive accuracy	124
2.4.4	Confidence interval	125
2.5	Application to the prediction of prostate cancer recurrence	126

2.5.1	Datasets	126
2.5.2	Specification of the joint models	128
2.5.3	Estimation and goodness-of-fit of the joint models	130
2.5.4	Predictive accuracy of the joint models	134
2.6	Discussion	140
3	Analyse supplémentaire	145
4	Exemples de prédiction individuelle dynamique avec tous les modèles conjoints	149
5	Conclusion	150
V Dynamique des PSA après l'initiation d'une hormonothérapie chez des patients traités initialement par radiothérapie pour un cancer de la prostate localisé		153
1	Introduction	153
2	Analyses statistiques	155
3	Sélection des données	158
4	Résultats	159
4.1	Description de l'échantillon	159
4.2	Dynamique globale du PSA	163
4.3	Facteurs associés à la dynamique du PSA	165
4.4	Analyse de sensibilité	168
4.5	Facteurs prédictifs de rechute clinique après une hormonothérapie	169
5	Conclusion	172
VI Discussion, conclusion et perspectives		175
1	Discussion	175
1.1	Prédiction dynamique : comparaison d'approches	175
1.2	Prédiction dynamique différentielle	178
1.3	Dynamique des PSA après l'hormonothérapie	180

Table des matières

2	Conclusion	181
3	Perspectives	182
	Références bibliographiques	185
	Annexe	203
	Liste des publications et communications scientifiques	221

Table des figures

I.1	Évolution caractéristique du PSA post-radiothérapie.	14
I.2	Changement de dynamique du PSA après initiation d'un nouveau traitement.	18
II.1	Graphe orienté structurel de la modélisation conjointe de données longitudinales et de temps d'événement	38
II.2	Graphe orienté du modèle conjoint à effets aléatoires partagés	42
II.3	Graphe orienté du modèle conjoint à classes latentes	50
II.4	Graphe orienté du modèle conjoint avec une dépendance entre le biomarqueur longitudinal et le temps d'événement à la fois à travers les effets aléatoires et la classe latente	55
II.5	Outil pronostique dynamique	57
III.1	Individual observed trajectories of $\log(PSA + 0.1)$ over time for (A) censored patients and (B) patients who clinically recurred.	86
III.2	Comparison of the subject-specific predictions (A) and the marginal predictions (B) with the observed PSA values averaged by intervals of times having the same amount of PSA measures.	92

III.3	Mean predicted subject-specific survival curves from four joint models and observed Kaplan-Meier survival curve with 95% confidence bands.	93
III.4	Kaplan-Meier estimates of the Cox-Snell residuals from four joint models ((b), (c), (f) and (g)) in plain black line, their 95% confidence bands in dashed black line, and the theoretical survival curve of a unit exponential random variable in plain grey line.	94
III.5	Martingale residuals <i>versus</i> predicted covariates summarizing the dependency between the PSA dynamics and the risk of recurrence in four joint models ((c), (f), (b) and (g)).	95
III.6	(A) $CVPOL_a$ computed from the 7 shared random-effect models from times of prediction from 1 to 6 years after the end of radiation therapy. For each time of prediction, the number of subjects still at risk (and the number of clinical recurrence) are indicated. (B) Differences in EPOCE estimate with 95% tracking interval (T.I.) between 4 joint models from times of prediction from 1 to 6 years after the end of radiation therapy.	97
III.7	Individual dynamic predicted probability of clinical recurrence (●) within 3 years and updated every 6 months computed from model (f) and its 95% confidence bands (bold plain line) obtained by a Monte Carlo method with 2000 draws ; repeated PSA measures are denoted by ×. The patient had a T-stage of 2, a Gleason score of 7, an initial PSA of 9.7ng/mL and had a clinical recurrence after 3.8 years following the end of EBRT.	99

-
- IV.1 Individual observed trajectories of $\log(\text{PSA} + 0.1)$ after the end of EBRT until the observed survival time (at the vertical black dash line) : (a) for two patients who received HT (at vertical grey dashed line) and who subsequently recurred, (b) for two patients who received HT and were subsequently censored, (c) for two patients who recurred without initiating any HT and (d) for two patients who were censored without initiating any HT. Black dots represent observed values of PSA before HT and the black curve represents the subject-specific PSA predictions from the linear mixed model. Grey dots are the observed PSA values after HT and the grey curve represents the extrapolated subject-specific PSA predictions from the mixed model based only on pre-HT data. It gives the expected PSA trajectory assuming the patient did not receive any HT. 127
- IV.2 Predictive accuracy measures after an immediate initiation of ST averaged over the times of ST initiation for 6 joint models : with (a) POL estimate, (b) difference in POL and 95% C.I., (c) BS estimate, (d) difference in BS and 95% C.I. Negative (respectively positive) differences indicate the first model has a better (respectively worse) predictive ability. 135
- IV.3 Predictive accuracy measures in absence of ST for 6 joint models at times from 1 to 6 after EBRT with (a) EPOCE estimate, (b) difference in EPOCE and 95% C.I., (c) BS estimate, (d) difference in BS and 95% C.I., (e) IBS estimate and (f) difference in IBS and 95% C.I. Negative (positive) differences indicate the first model has a better (worse) predictive ability. 137

IV.4	Individual predicted probabilities of clinical recurrence within 3 years assuming either immediate initiation of HT (grey) or no initiation of HT in the next 3 years (black). Repeated PSA measurement are denoted \times . After each new measurement, the prediction within 3 years is computed at the vector of parameters estimates $\hat{\theta}$ (unfilled black triangle and unfilled grey circle), and the distribution is approximated by a Monte Carlo method (solid black triangle and solid grey circle indicate the median and the intervals indicate the 95% bands).	140
IV.5	Mesures de pouvoir prédictif si l'initiation de l'hormonothérapie est considérée comme une censure.	147
IV.6	Mesures de pouvoir prédictif si l'initiation de l'hormonothérapie est considérée comme une rechute clinique.	148
IV.7	Probabilités prédites de rechute clinique avec les différents modèles conjoints si le patient a une initiation immédiate de l'hormonothérapie (a) et s'il ne reçoit aucun second traitement dans les 3 prochaines années (b). . .	150
V.1	Exemples de deux patients de BM dont les profils d'évolution du PSA ressemblent à ceux de patients ayant reçu une seconde hormonothérapie. . .	159
V.2	Trajectoires individuelles du $\log(\text{PSA} + 0,1)$ après l'hormonothérapie . .	162
V.3	Évolutions moyennes prédites du $\log(\text{PSA} + 0,1)$ pour les trajectoires en 2 phases avec $\{f((t; -2, 5), t)\}$ et $\{f((t; -2, 5), t/\log(t + 1))\}$ et des erreurs indépendantes et pour les trajectoires avec des erreurs indépendantes en plus du processus d'auto-corrélation pour les autres courbes.	165
V.4	Evolution moyenne prédite du $\log(\text{PSA} + 0,1)$ après l'HT suivant le niveau de PSA initial, au diagnostic, pour un patient avec un niveau de PSA précédant l'initiation de l'HT égal à 3 ng/ml (en log) et une pente égale à 0,69.	167

-
- V.5 Evolution moyenne prédite du $\log(\text{PSA} + 0,1)$ après l'HT suivant le niveau de PSA précédant l'initiation de l'HT pour un patient avec un niveau de PSA au diagnostic égal à 2,6 ng/ml (en log) et une pente de PSA précédant l'initiation de l'HT égale à 0,69. 167
- V.6 Evolution moyenne prédite du $\log(\text{PSA} + 0,1)$ suivant la pente de PSA précédant l'initiation de l'HT pour un patient avec un niveau de PSA au diagnostic égal à 2,6 ng/ml (en log) et un niveau de PSA précédant l'initiation de l'HT égal à 2,3 ng/ml (en log). Une pente de 0,69 correspond à un temps de doublement de PSA de 1 année et 0,17 correspond à 4 années. 168

Liste des tableaux

III.1	Goodness-of-fit statistics and parameter estimates (standard-deviation) of the association between PSA dynamics and risk of recurrence adjusted for other prognostic factors from joint models (a) to (g).	90
IV.1	Goodness-of-fit statistics.	132
IV.2	Parameters estimates (and standard error (se)) of the HT and the association between the PSA dynamics and the risk of clinical recurrence adjusted on the prognostic factors.	133
V.1	Description des données sélectionnées des différents cohortes.	161
V.2	Qualité d'ajustement des différents modèles.	164
V.3	Facteurs pronostiques sur les différentes phases d'évolution pour la modèle en 2 phases $f(t, 5, 5)$ et une tendance linéaire sur le long-terme ajusté sur les trois cohortes (UM, RTOG, BM).	166
V.4	Estimation des paramètres du modèle final réalisée sur des ensembles de données différents.	170

Chapitre I

Contexte épidémiologique

Ce premier chapitre présente le contexte épidémiologique de la thèse qui porte sur le cancer de la prostate. On y présente le cancer et les différentes questions de recherche qu'il soulève.

1 Cancer de la prostate

1.1 La prostate

La prostate est une glande de l'appareil génital masculin située sous la vessie, en avant du rectum. Elle est contenue dans une capsule qui l'isole quelque peu des organes voisins. Son rôle principal est de sécréter et stocker le liquide séminal qui est le principal constituant du sperme. Elle est sous la dépendance de la testostérone, sécrétée essentiellement par les testicules et en petite partie par les glandes surrénales. La prostate contient également un muscle qui aide à expulser le sperme pendant l'éjaculation. La prostate saine est approximativement de la taille d'une noix, de 15 à 25 g à l'âge adulte. Pour fonctionner correctement, la prostate a besoin d'androgènes (hormones masculines) dont le plus important est la testostérone. La prostate peut être le site de plusieurs infections telles que l'adénome de la prostate qui est une tumeur bénigne, la prostatite qui est une inflammation de la prostate ou le *cancer de la prostate*.

1.2 Le cancer

Le cancer est le résultat d'un dysfonctionnement au niveau de certaines cellules de l'organisme. Celles-ci se mettent à se multiplier de manière anarchique et à se métastaser en migrant de la prostate jusqu'à d'autres parties du corps (surtout les os et ganglions lymphatiques). Le cancer de la prostate est un cancer qui se développe habituellement dans les cellules glandulaires de la prostate. Le cancer de la prostate survient indépendamment de l'adénome de la prostate appelé également hypertrophie bénigne de la prostate. C'est dans l'immense majorité des cas un adénocarcinome.

1.2.1 L'épidémiologie

Le cancer de la prostate est le deuxième cancer le plus diagnostiqué au monde chez l'homme avec près de 903500 nouveaux cas par an, ce qui représente 14% parmi tous les types de cancer en 2008 (Globocan, 2008; Jemal et al., 2011). C'est le premier cancer chez l'homme de plus de 50 ans en Europe et aux États-Unis. On constate de larges différences régionales (Haas et al., 2008) avec les taux d'incidence les plus élevés enregistrés dans les pays développés et les taux de mortalité les plus élevés dans les pays en voie de développement (Jemal et al., 2011, Center et al., 2012). Ces différences s'expliquent par le fait que, dans les pays développés, les dépistages sont plus courants et l'accès aux traitements plus facile.

En 2010 en France, le nombre de nouveaux cas était estimé à 71500 avec des taux d'incidence et de mortalité respectivement de 128,8 et 11,2 pour 100000 [Institut National du Cancer] alors que ces taux étaient respectivement de 118,3 et 12,7 en 2008 (Globocan, 2008).

1.2.2 Le diagnostic

Des examens diagnostiques permettent de détecter la présence d'un cancer, d'identifier le type du cancer, la taille de la tumeur ainsi que le stade de la maladie dans le but de proposer un traitement adéquat. On distingue trois méthodes de dépistage du cancer de la prostate : le toucher rectal, le dosage du PSA (antigène spécifique de la prostate) et la biopsie.

Le toucher rectal est une méthode qui consiste à palper la prostate à travers la paroi du rectum. Cette méthode permet de détecter d'éventuelles anomalies dans la zone périphérique de la prostate. [Okotie et al. \(2007\)](#) et [Gosselaar et al. \(2008\)](#) ont montré qu'un toucher rectal suspect est souvent associé à un cancer plus agressif. L'utilisation du toucher rectal est souvent recommandée en association avec un dosage du PSA total sérique.

En effet, le PSA, antigène spécifique de la prostate, est une protéine naturellement sécrétée par les cellules prostatiques et par les cellules hormonales chez les sujets atteints d'un cancer de la prostate. Elle a pour fonction de liquéfier le liquide séminal et ainsi permettre une meilleure circulation des spermatozoïdes. Cette protéine se retrouve dans le sang si bien qu'elle peut être facilement dosée. En présence de cellules cancéreuses, la sécrétion de PSA est automatiquement augmentée.

Généralement, un taux de PSA inférieur à 4 ng/ml est considéré normal ([Carter, 2000](#)), suspect lorsqu'il est compris entre 4 et 10 ng/ml et fortement corrélé à la présence d'un cancer au-delà des 10 ng/ml. Cette corrélation fût notamment observée dans une étude de 1653 hommes par [Catalona et al. \(1991\)](#). Parmi les 30 sujets ayant un taux de PSA supérieur à 10 ng/ml, 67.9% étaient atteints d'un cancer de la prostate. [Bradford et al. \(2006\)](#) ont également montré que des valeurs très élevées de PSA témoignent le plus souvent de formes évoluées du cancer. Pour des valeurs de PSA dites suspectes, c'est-à-dire comprises entre 4 et 10 ng/ml, il peut être nécessaire de prendre en compte d'autres

indicateurs : le ratio PSA libre/PSA total et la vélocité ou le temps de doublement du PSA ([Heidenreich et al., 2013](#)).

Le PSA qui circule dans le sang est de deux natures, l'un lié aux protéines et l'autre qui est dit libre. La mesure du pourcentage de PSA libre (PSAL) sur le PSA total (PSAT) dans le sang diminue en cas de cancer de la prostate. Il a été constaté que le taux de PSAL est plus bas en présence d'une tumeur maligne par rapport à une tumeur bénigne. C'est la raison pour laquelle un rapport PSAL/PSAT inférieur à 15% est généralement en faveur d'un diagnostic de cancer de la prostate ([Salomon et al., 2010](#)).

La vélocité (PSAV) et le temps de doublement (PSADT) sont quant à eux des indicateurs temporels. La vélocité est utilisée en cas d'évolution linéaire et s'exprime en ng/ml/année tandis que le temps de doublement s'applique en cas d'évolution exponentielle. Ce dernier correspond au temps mis par le taux sérique de PSA pour doubler ; c'est aussi l'inverse de la demi-vie du PSA ([Ruffion et al., 2005](#)). Cependant, même si cette cinétique du PSA est un facteur pronostique après une prostatectomie ou une radiothérapie ([D'Amico et al., 2004a, 2005](#)), il n'existe pas de justifications particulières à utiliser la cinétique plutôt que la PSA total dans le cadre d'un diagnostic ([O'Brien et al., 2009](#) ; [Vickers et al., 2009](#)).

[Richie et al. \(1994\)](#) montrent, à partir de données d'essais cliniques multicentriques chez des hommes âgés de 50 à 91 ans et sans aucun antécédent de cancer, que le dosage du PSA est significativement plus efficace que le toucher rectal (sensibilité et spécificité de 75% et 87% contre 75% et 69% respectivement), pour diagnostiquer le cancer de la prostate ([Richie et al., 1993](#)). Ils montrent également que la combinaison de ces deux méthodes permet de détecter un plus grand nombre de cas de cancer.

Nous venons de voir qu'une anomalie suspecte au toucher rectal et/ou une valeur élevée du PSA sont des signes potentiels de présence de cancer. Néanmoins d'autres facteurs peuvent engendrer une augmentation du taux de PSA tels que l'hypertrophie bénigne de

la prostate, une inflammation, une infection ou simplement l'âge.

Seul un troisième examen, la biopsie, permet d'avoir un diagnostic définitif du cancer. En effet, la preuve définitive de la présence d'un cancer ne peut être établie qu'en analysant du tissu prostatique. La biopsie consiste à prélever plusieurs fragments de tissu prostatique à travers le rectum. Un examen microscopique des fragments prélevés, complété au besoin par des examens radiographiques spécialisés (scanner, scintigraphie osseuse, IRM), peut alors permettre d'infirmier ou de confirmer le diagnostic.

1.2.3 Les facteurs de risques

Même si les causes du cancer de la prostate restent inconnues, quelques facteurs de risques ont été identifiés. Le facteur de risque le plus important est l'âge. [Leitzmann and Rohrmann \(2012\)](#) ont montré qu'entre 2000 et 2008, le taux d'incidence chez les hommes âgés de 40 à 44 ans est de 9.2 pour 100.000 alors qu'il est de 984.8 chez les hommes âgés entre 70 et 74 ans. [Richie et al. \(1993\)](#) ont aussi montré, à partir de biopsies réalisées chez des sujets avec un taux de PSA supérieur à 4 ng/ml et un toucher rectal suspect, que le taux de détection du cancer augmentait de manière significative avec l'âge passant de 3% chez les hommes entre 50 et 59 ans à 14% chez ceux ayant plus de 80 ans.

[Sakr et al. \(1993\)](#) ont également observé, dans une population d'hommes de moins de 49 ans, que certains sujets âgés seulement de 20 ans présentaient déjà l'existence de cellules cancéreuses dans la prostate (10%). Des études faites sur des autopsies ont montré que ce taux augmente jusqu'à 30% chez des hommes âgés de 50 ans et à plus de 50% chez ceux ayant 80 ans ([Chodak, 2006](#)). D'autres autopsies réalisées aux États-Unis ont montré que la prévalence du cancer de la prostate augmente avec l'âge de manière significative (2% de cas de cancer chez les moins de 60 ans et 19% chez les plus de 60 ans, p -value < 0.001) ([Haas et al., 2007](#)). La survenue du cancer est rare avant 40 ans et la plupart des cas de cancer se déclarent entre 60 et 90 ans ([Haas et al., 2008](#)). Cependant, le cancer

de la prostate a une évolution très lente, en particulier lorsqu'il est localisé, et ne permet pas d'être décelé rapidement. Certains sujets peuvent vivre longtemps sans aucun signe de la maladie. Dans la plupart des cas, il est découvert après 70 ans ([Drouin and Rouprêt, 2009](#)).

Les autres facteurs de risque établis sont les antécédents familiaux et les origines ethnique et géographique ([Jemal et al., 2011](#)). A partir de la base de données de l'agence internationale pour la recherche sur le cancer, [Haas et al. \(2008\)](#) montrent ainsi que le taux d'incidence du cancer de la prostate standardisé sur l'âge est plus élevé aux États-Unis, notamment dans la population noire américaine (185,4 pour 100 000), tandis que la Chine présente le plus faible taux (1,7 pour 100 000).

1.2.4 Les stades

On distingue essentiellement quatre stades différents relatifs au cancer de la prostate. Chacun d'eux dépend de l'évolution et de la propagation des cellules cancéreuses dans l'organisme et se différencie également selon certains facteurs tels que la taille et la localisation de la tumeur. Le stade du cancer est défini en utilisant la classification TNM proposée par le chirurgien français Pierre Denoix et qui prend en compte trois critères ([Sobin and Wittekind, 2009](#)) :

- T (Tumor (tumeur)) suivi d'un nombre de 0 à 4 décrit la taille de la tumeur et la quantité de tissu voisin qu'elle a envahi. T0 correspond à l'absence de tumeur, T1 à une tumeur non palpable au toucher rectal ou non visible en imagerie, T2 à une tumeur localisée à la prostate, T3 à une tumeur étendue au delà de la capsule prostatique et T4 à une tumeur fixée à la paroi pelvienne ou ayant atteint d'autres organes adjacents.

- N (Nodes (ganglions)) suivi d'un nombre de 0 à 1 définit le degré de propagation dans les ganglions lymphatiques. N0 représente une absence de métastases ganglionnaires régionales et N1 une atteinte ganglionnaire régionale.
- M (Metastases (métastases)) suivi du nombre 0 ou 1 spécifie la propagation du cancer à d'autres parties du corps. M0 correspond à une absence de métastases à distance ; à l'inverse, M1 représente la présence de métastases à distance.

Lorsque la tumeur se développe uniquement dans la glande prostatique, on parle de *cancer localisé* (T1/T2, N0, M0). On distingue 3 sous catégories pour le cancer localisé : le *cancer localisé à risque faible*, à *risque intermédiaire* ou à *haut risque*. Cette sous classification porte le nom de classification de D'Amico ([D'Amico et al., 1998](#)) et est établie selon le stade TNM, le score de Gleason et la valeur du PSA.

Le score de Gleason, qui mesure l'agressivité des cellules cancéreuses, se définit comme la somme des deux grades de cellules (chacun allant de 1 à 5) les plus représentés dans la tumeur analysée par biopsie. Ainsi, un score égal à 2 correspond à une tumeur très proche d'un tissu bénin tandis qu'un score égal à 10 représente une tumeur très agressive. Parmi les valeurs intermédiaires, l'ordre indiqué à son importance, le premier grade correspondant au grade de cellules le plus représenté. Il existe donc une distinction selon que le score de Gleason égal à 7 correspond à 3+4 ou 4+3. Un sujet ayant un score de Gleason égal à 3+4 sera classé dans le groupe à risque intermédiaire alors qu'un sujet ayant un score de Gleason égal à 4+3 serait plutôt classé dans le groupe à haut risque ([D'Amico et al., 1998](#)).

Dans le cas où la tumeur s'étend au delà de la capsule prostatique (T3/T4, N0, M0), on parle de *cancer localement avancé*, dans celui où les cellules cancéreuses gagnent la circulation lymphatique et se fixent sur les ganglions (T, N1, M0), on parle de *cancer avec atteinte des ganglions voisins* et enfin lorsque le cancer présente une ou plusieurs métastases à distance (T, N, M1), on parle de *cancer métastatique* .

L'importance de connaître le stade d'avancement du cancer permettra de proposer au patient le meilleur traitement à adopter contre la maladie.

1.2.5 Les traitements

Il existe plusieurs types de traitements qui dépendent du stade du cancer, de l'âge et de l'état de santé général du patient. Par exemple, pour un patient ayant un cancer localisé à risque faible et une espérance de vie inférieure à 10 ans, une simple surveillance régulière est suffisante. Cependant, si la surveillance montrait l'apparition de signes de progression locale ou générale, il pourrait alors être mis sous traitement palliatif ou hormonal. Cette stratégie est appelée la surveillance avec traitement différé ([Steinberg et al., 1998](#) ; [Holmberg et al., 2002](#) ; [Heidenreich et al., 2013](#)).

On distingue principalement deux types de traitement : des traitements locaux et systémiques. Les traitements locaux sont réservés au cancer localisé et ont pour but d'enlever ou détruire la tumeur. Parmi eux, la *prostatectomie totale* est une opération chirurgicale visant à enlever complètement la prostate. Chez les hommes présentant un cancer localisé et une espérance de vie supérieure à 10 ans, l'objectif de cette chirurgie est d'éradiquer la maladie tout en préservant au maximum la qualité de vie des patients, notamment la continence et la puissance sexuelle ([Bianco et al., 2005](#)). La prostatectomie a montré un bénéfice pour la survie globale et pour la survie spécifique au cancer de la prostate en comparaison à une surveillance attentive ([Bill-Axelsson et al., 2011](#)).

La *radiothérapie externe (RT)* est une alternative à la chirurgie et consiste à utiliser des composants radioactifs (des rayonnements) dans le but de détruire les cellules cancéreuses en les empêchant de se multiplier tout en préservant le plus possible les tissus sains et les organes voisins. Elle peut être également utilisée en complément d'une chirurgie pour limiter les risques de rechute locale. La RT est fractionnée sur plusieurs séances et les doses

administrées varient suivant le type de cancer. Une dose de radiation totale d'au moins 74 Gy (Gray) est recommandée chez des patients ayant un cancer localisé à risque faible. [Kupelian et al. \(2005\)](#) ont démontré que ces patients avaient une augmentation significative de leur survie sans progression biochimique par rapport à des patients ayant reçu des doses de radiothérapie inférieures à 72 Gy (69% contre 63% avec une $p\text{-value}=0.046$). Pour les patients ayant un cancer localisé à risque intermédiaire, il a été montré que l'augmentation de la dose totale (76-81 Gy) avait un impact bénéfique sur la survie à 5 ans sans progression ([Peeters et al., 2006](#)). Cette augmentation de dose améliore également la survie à 5 ans sans progression biochimique chez des patients ayant un cancer localisé à haut risque, mais ne semble pas suffire à éradiquer le risque de rechute ([D'Amico et al., 2008](#)). Un autre type de radiothérapie dite *interne* consiste à implanter directement dans la prostate des grains d'iode radioactifs. Cette thérapie est également appelée *curiethérapie*. D'autres traitements alternatifs existent comme le traitement par ultrasons focalisés de haute intensité (HiFU) ou la *cryothérapie* appelée aussi *cryochirurgie* ([Salomon et al., 2010](#); [Heidenreich et al., 2013](#)). L'HiFU entraîne la destruction des cellules cancéreuses par le biais d'ultrasons délivrés au contact de la prostate à l'aide d'une sonde placée dans le rectum tandis que la cryothérapie est une technique récente basée sur l'utilisation du froid pour détruire la tumeur.

Les traitements systémiques s'adressent, quant à eux, aux patients dont le cancer est étendu en dehors de la prostate, à ceux ayant eu une récurrence après un traitement initial ou encore, à titre palliatif, chez des patients très âgés. Parmi eux, le *traitement hormonal* ou *l'hormonothérapie (HT)* est basé sur le caractère hormono-dépendant du cancer de la prostate et consiste en l'administration de produits empêchant la sécrétion d'androgènes (testostérone), bloquant ainsi le développement du cancer. La chimiothérapie fait également partie des traitements généraux et peut être une alternative dans le cas d'un cancer de la prostate résistant à la castration (chimique ou physique) ([Salomon et al., 2010](#)).

Il existe différentes recommandations thérapeutiques suivant le degré de risque de la tumeur. Elles peuvent se caractériser par l'utilisation d'un traitement unique (local ou général) ou d'une combinaison de traitements. Les sujets ayant une tumeur à risque faible auront généralement comme recommandation des traitements locaux suivis d'une possible surveillance active. Les sujets à risque intermédiaire et à haut risque auront, quant à eux, la possibilité de recevoir la combinaison d'un traitement local et d'un traitement général. Dans le traitement des tumeurs à risques intermédiaires, un traitement par radiothérapie avec augmentation de dose ou l'association de celui-ci avec une hormonothérapie courte (6 mois) peuvent être proposés (Salomon et al., 2010). Il a été montré par exemple dans une étude randomisée publiée en 2004 composée en majorité de patients ayant des tumeurs à risque intermédiaire, que la survie globale à 5 ans était significativement meilleure dans le groupe ayant eu une combinaison radiothérapie/hormonothérapie que celle du groupe ayant eu seulement une radiothérapie (88% contre 78%, p-value=0,04) (D'Amico et al., 2004a). Un traitement exclusif par hormonothérapie peut aussi être une option chez des patients atteints d'un cancer de la prostate localisé et n'étant pas éligibles à un traitement local (Salomon et al., 2010).

Une fois le traitement terminé, un suivi régulier du patient est mis en place, via des examens cliniques et des dosages du PSA, afin de surveiller comment le patient réagit après la fin du traitement, d'identifier l'apparition d'effets indésirables voire éventuellement la réapparition du cancer (*une rechute*). Ce suivi est plus intensif dans les cinq premières années où le risque de rechute est le plus important.

2 Le biomarqueur Prostate-Specific-Antigen (PSA)

Dans le domaine médical, un biomarqueur est défini comme une molécule ou une protéine anormalement manquant ou sur-exprimée en présence d'une maladie. Il peut être utilisé pour le dépistage, le diagnostic, l'évaluation de la réponse ou de la tolérance à un traitement. Un biomarqueur est une caractéristique mesurable, indicatrice de processus biologiques normaux ou pathologiques, ou de réponses pharmacologiques à une intervention thérapeutique ([Atkinson et al., 2001](#)). Il permet, avec l'arrivée des premiers traitements de thérapie ciblée et de nouveaux tests de dépistage ou de diagnostic plus sensibles et plus rapides, d'envisager une médecine plus efficace, une médecine dite "personnalisée".

Dans le cancer de la prostate, comme il a déjà été évoqué en [1.2.2](#), le biomarqueur utilisé est le PSA. A la fin des années 80, l'utilisation des mesures du PSA comme biomarqueur de diagnostic ou de progression du cancer a connu une grande ascension.

2.1 PSA et diagnostic de cancer

Même si le dosage du PSA est très utilisé pour le diagnostic du cancer comme nous l'avons vu dans la section [1.2.2](#), son utilisation est de plus en plus controversée. Tout d'abord, une augmentation du taux de PSA peut être due à plusieurs facteurs autres que le cancer de la prostate tels que les infections ou une hypertrophie bénigne. Ensuite, comme le souligne [Barry \(2001\)](#), il peut y avoir de nombreuses sources de variations entre les mesures de PSA : des variations suivant les kits de tests ou d'un laboratoire à un autre. L'existence de telles variations peut entraîner de faux diagnostics engendrant une estimation erronée de l'incidence du cancer de la prostate. De plus, la sensibilité et la spécificité du test de PSA et du seuil auquel une biopsie est faite ne sont pas non plus clairement définis ([Barry, 2001](#)).

Une étude européenne randomisée faite en 2009 ([Schröder et al., 2009](#)) a montré, chez les hommes de 55 à 69 ans, que le dépistage par le PSA réduit le taux de mortalité de 20%. Toutefois, ce dépistage est aussi associé à un risque élevé de sur-diagnostic (défini comme le diagnostic chez des patients asymptomatiques) de 50% dans le groupe de dépistage, ou de sur-traitement pour un cancer peu évolutif.

Cette controverse a été également soulevée par une étude américaine similaire qui a étudié l'effet du dépistage par un dosage de PSA et un toucher rectal sur le taux de mortalité par cancer de la prostate. Il a été démontré qu'il n'y avait pas de différence significative sur le taux de mortalité en ayant fait un dépistage ou non. ([Andriole et al., 2009](#)).

Enfin, un nouveau biomarqueur découvert à la fin des années 1990, le PCA3, est de plus en plus utilisé pour le diagnostic du cancer de la prostate. C'est un gène spécifique à ce cancer et, contrairement au PSA, le PCA3 n'est produit que par les cellules cancéreuses et n'est pas influencé par la taille de la prostate ([Deras et al., 2008](#)). Il permet donc de mieux différencier le cancer de la prostate des maladies bénignes et non cancéreuses telles que les infections. Le PCA3 est dosé dans les urines après un massage prostatique. Des études ont montré que l'utilisation du PCA3 permet d'obtenir des résultats prometteurs dans le diagnostic du cancer de la prostate en terme de sensibilité et de spécificité, ils semblent être meilleurs que ceux obtenus avec le PSA total ([Deras et al., 2008](#)).

2.2 PSA et progression du cancer

Si diverses études remettent en question la pertinence d'une prise en compte du dosage du PSA lors du diagnostic, il y a un consensus sur l'utilité de ce biomarqueur pour le suivi des patients après un traitement, que ce soit après une prostatectomie ([Pound et al., 1999](#)) ou une radiothérapie ([Roach III et al., 2006](#)).

I.2 Le biomarqueur Prostate-Specific-Antigen (PSA)

Après un traitement par radiothérapie, on observe une évolution biphasique, avec dès la fin du traitement, une chute du taux de PSA qui traduit le contrôle de la maladie suivie d'une phase de stabilisation ou de croissance plus ou moins lente à long-terme (Figure I.1). Cette croissance peut être liée à une résurgence de la maladie, c'est à dire à une *rechute clinique locale, distante ou métastatique* (Shipley et al., 1999; Bellera et al., 2008b). Le suivi du taux de PSA après la fin d'un traitement permet donc de visualiser la progression de la maladie au cours du temps, d'apprécier l'efficacité du traitement dans les premiers mois et d'anticiper une éventuelle rechute.

Cette évolution biphasique du PSA après une radiothérapie a été décrite par plusieurs auteurs. Pauler and Finkelstein (2002) et Bellera et al. (2008a) ont décrit la dynamique de PSA à travers un modèle mixte en supposant une trajectoire linéaire sur la phase de déclin, un changement de pente spécifique au sujet (point de nadir) et une stabilisation ou croissance du PSA linéaire à long terme. Proust-Lima et al. (2008) ont, quant à eux, considéré une trajectoire plus lisse permettant de décrire au plus près l'évolution du PSA dans les premiers mois après la fin du traitement.

Proust-Lima et al. (2008) ont aussi trouvé, en décrivant la dynamique du PSA à travers ce modèle linéaire mixte, que les facteurs pronostiques standard tels que le niveau de PSA initial (pré-radiothérapie), le stade du cancer ainsi que le score de Gleason étaient significativement associés à la dynamique du PSA. Ils ont ainsi montré que seul le PSA au diagnostic était associé au niveau du PSA post-radiothérapie, que le PSA au diagnostic et le stage du cancer étaient associés à la phase de déclin à court-terme et que le PSA au diagnostic, le stade du cancer et le score de Gleason étaient associés à la phase de stabilisation ou de croissance sur le long-terme.

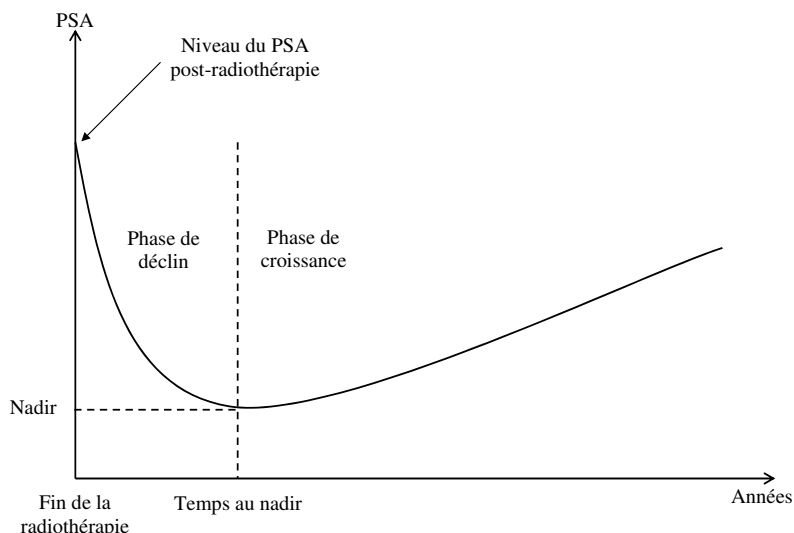


Figure I.1 – Évolution caractéristique du PSA post-radiothérapie.

2.3 PSA et rechute biochimique post-radiothérapie

En 1996, la société américaine thérapeutique de radiologie et d'oncologie ([ASTRO, 1997](#)) a établi un consensus pour une définition commune de la rechute dite biochimique basée sur le taux de PSA. Ils définissent la rechute biochimique, après une radiothérapie externe, comme *trois augmentations successives* du taux de PSA après le nadir ; le nadir étant le plus bas taux de PSA atteint.

La sensibilité et la spécificité de cette règle, appelée règle ASTRO, par rapport à une rechute clinique (locale, distante, décès par cancer, etc) ont été évaluées et ont donné des estimations allant de 55% à 92% et de 68% à 80% pour la sensibilité et la spécificité respectivement ([Thames et al., 2003](#) ; [Kuban et al., 2005](#) ; [Horwitz et al., 2005](#) ; [Horwitz et al., 2006](#)).

I.2 Le biomarqueur Prostate-Specific-Antigen (PSA)

Dix ans plus tard, une nouvelle conférence à Phoenix a abouti à une nouvelle définition de la rechute biochimique qui présenterait une meilleure corrélation avec les résultats cliniques. Elle correspond à *une élévation de 2ng/ml* du taux de PSA après le nadir, chez les patients ayant reçu ou non un traitement hormonal après avoir été traités par radiothérapie externe (Roach III et al., 2006). Bellera et al. (2009) montrent, à travers le calendrier des mesures et le temps de doublement du PSA, que cette nouvelle définition de rechute biochimique (PSA nadir + 2ng/ml), appelée définition de Phoenix, est meilleure que celle établie par ASTRO. En effet, elle permet de prendre en compte les fluctuations habituelles du taux de PSA à la suite d'un traitement par radiothérapie et, en particulier, les *phénomènes de rebond* fréquents n'ayant pas de valeur pronostique et pouvant entraîner la suspicion voire le traitement d'une rechute biochimique qui est en réalité inexistante (Critz et al., 2000 ; Rosser et al., 2002 ; Sengoz et al., 2003 ; Toledano et al., 2008).

Cette nouvelle définition permet d'obtenir une sensibilité et une spécificité plus robustes que celles de la définition d'ASTRO, 72% et 83% contre 51% et 81% (Kuban et al., 2006). Elle semble également favorable lors de suivis à court-terme contrairement à ASTRO qui serait meilleure dans le cas de suivis à long-terme (Kuban et al., 2003).

Le niveau de PSA initial, le stade du cancer et le score de Gleason ont été identifiés comme étant prédictifs d'une rechute biochimique (Hancock et al., 1995 ; Shipley et al., 1999 ; Hanks et al., 2002 ; Hanlon et al., 2002 ; Roach III et al., 2007). D'autres facteurs ont également été associés à une rechute biochimique, c'est le cas du temps de doublement de PSA (Pollack et al., 1994) et du nadir de PSA (Hanks et al., 2002). Une rechute biochimique entraîne souvent une mise sous traitement hormonal et précède généralement une rechute clinique (Shipley et al., 1999 ; Sandler et al., 2000 ; Thames et al., 2003 ; Roach III et al., 2006).

3 Rechute clinique de cancer de la prostate

La rechute clinique du cancer est constatée lors d'un examen clinique tout comme le diagnostic présenté en 1.2.2. Elle peut être locale, régionale, distante ou être représentée par le décès du patient lié à son cancer. Les facteurs pronostiques pré-radiothérapie associés à une rechute biochimique, tels que le niveau de PSA initial, le stage du cancer et le score de Gleason sont également associés à une rechute clinique pour un cancer localisé (Kuban et al., 1995 ; Kestin et al., 1999 ; Roach III et al., 2007 ; Proust-Lima et al., 2008).

3.1 PSA et rechute clinique

D'autres facteurs tels que la cinétique de la dynamique du PSA sont associés à la rechute clinique (Zagars et al., 1995 ; Taylor et al., 2005 ; Shi et al., 2013), comme le temps de doublement de PSA (Pollack et al., 1994 ; Albertsen et al., 2004 ; D'Amico et al., 2004a) ou le niveau courant et la pente courante de PSA (Proust-Lima et al., 2008).

Hancock et al. (1995) et Sartor et al. (1997) ont de plus mis en évidence l'existence d'une corrélation entre le temps de doublement de PSA et le type de rechute. Selon Lee et al. (1997), un temps de doublement inférieur à 12 mois depuis la fin du traitement représenterait un prédicteur significatif de rechute clinique distante .

3.2 Prédiction individuelles de rechute

La prédiction d'un événement clinique à l'aide d'outils pronostiques est une question centrale en oncologie. En particulier, à la suite d'un premier traitement anti-cancéreux, la rechute biochimique ou clinique est un évènement d'intérêt majeur. Actuellement, les outils pronostiques classiques prennent en compte uniquement les caractéristiques individuelles des patients mesurés au moment du diagnostic ou considèrent uniquement de simples seuils

de mesures comme le temps de doublement du PSA ([Hancock et al., 1995](#) ; [Lee et al., 1997](#)) ou la règle de Phoenix : le "PSA nadir + 2ng/ml" ([Roach III et al., 2006](#)).

Ces outils pronostiques ont l'inconvénient de ne pas tenir compte de l'ensemble des informations disponibles. En particulier, ceux-ci font abstraction de l'histoire de la trajectoire post-traitement du biomarqueur et la plupart ne prennent pas en considération le caractère continu des mesures de PSA ([Thompson et al., 2005](#)).

Le caractère dynamique des données répétées du biomarqueur permet de proposer des *outils pronostiques dynamiques* considérant toute la trajectoire post-traitement du biomarqueur ([Henderson et al., 2002](#) ; [Taylor et al., 2005](#) ; [Yu et al., 2008](#) ; [Proust-Lima and Taylor, 2009a](#) ; [Rizopoulos, 2011b](#)). Ces données répétées permettent, entre autres, une meilleure visualisation de la progression de la maladie au cours du temps ainsi qu'une optimisation du suivi des patients. De tels outils pronostiques dynamiques sont basés sur des modèles statistiques. Pour prendre en compte le fait que l'évolution du biomarqueur et le risque de rechute sont fortement liés, une solution est de *modéliser conjointement* l'évolution longitudinale du biomarqueur et la survenue de la rechute.

Dans le cancer de la prostate spécifiquement, quelques outils pronostiques dynamiques issus de la modélisation conjointe ont été proposés ([Taylor et al., 2005](#) ; [Yu et al., 2008](#) ; [Proust-Lima and Taylor, 2009a](#)). Ces outils dynamiques ont permis d'avoir des prédictions de l'événement plus précises que les outils pronostiques classiques.

3.3 L'initiation d'hormonothérapie au cours du suivi

Dans le cancer de la prostate, après un traitement par radiothérapie, si l'évolution du PSA est suggestive d'une éventuelle rechute biochimique, le patient peut être amené à initier un second traitement (généralement hormonal). Ce nouveau traitement entraîne un changement de la dynamique du biomarqueur. Comme décrit sur la Figure [I.2](#), à l'ini-

tiation d'une hormonothérapie, les taux de PSA chutent sous l'effet du traitement puis, comme à la suite d'un traitement initial, ils se stabilisent ou peuvent légèrement augmenter. L'initiation d'hormonothérapie modifie aussi le risque de rechute clinique puisqu'il vise à retarder une éventuelle rechute du cancer. A la fois dans la modélisation des PSA et du risque d'événement et dans le développement d'outils pronostiques dynamiques, cette possible initiation à un nouveau traitement doit donc être prise en compte.

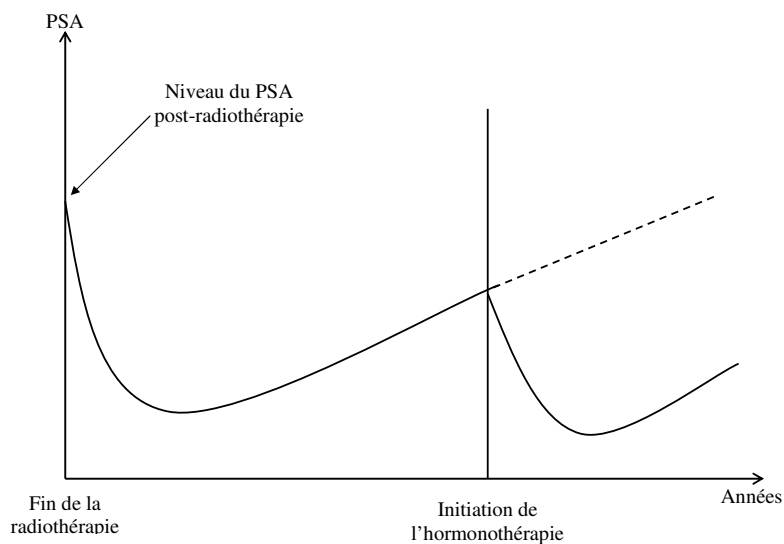


Figure I.2 – Changement de dynamique du PSA après initiation d'un nouveau traitement.

4 Objectifs de la thèse

L'objectif principal de cette thèse est de développer et valider des *outils pronostiques dynamiques*, prenant en compte toute la trajectoire du PSA, afin d'estimer de manière optimale le *risque de rechute clinique* chez des patients atteints de cancer de la prostate localisé et déjà traités par radiothérapie externe.

La présentation de ce travail s'articule en 6 chapitres. Nous commençons par proposer, dans un premier temps, un état de l'art des méthodes statistiques utilisées tout au long de cette thèse (chapitre II). Il y sera tout d'abord détaillé l'analyse des données longitudinales et l'analyse des données de survie. Nous présentons ensuite la méthodologie relative à la modélisation conjointe pour données longitudinales et de temps d'événement. Enfin la description des outils pronostiques dynamiques issus de ces modèles et leur validation clôturent le chapitre.

Le premier travail de la thèse, présenté dans le chapitre III, consiste à décrire et comparer différentes approches de modélisation conjointe des données longitudinales du PSA et du risque de rechute clinique à la fois en terme d'ajustement aux données et aux hypothèses sous-adjacentes, et en terme de pouvoir prédictif.

Le chapitre IV est consacré au développement et à la validation d'un outil pronostique dynamique dit "différentiel". En effet, cet outil permet de prédire le risque de rechute clinique suivant que le patient initie ou non immédiatement un second traitement de type hormonal. Dans ce travail, nous avons investigué différentes façons de modéliser la dépendance entre l'évolution longitudinale du PSA, l'initiation d'une hormonothérapie et le risque de rechute clinique. Les différents modèles ont été comparés en terme d'ajustement mais surtout en terme de pouvoir prédictif. La validation de l'outil pronostique a été faite dans le cas où une hormonothérapie est initiée immédiatement au temps de prédiction et dans le cas où aucune hormonothérapie n'est initiée dans la fenêtre de prédiction.

Chapitre I. Contexte épidémiologique

Le chapitre [V](#) revient enfin sur la progression de la maladie avec une description de la dynamique du PSA après le second traitement hormonal. Si la dynamique du PSA après la radiothérapie a été décrite et est bien connue, elle reste très peu connue après ce deuxième traitement. Dans la plupart des analyses statistiques, ces données post-hormonothérapie sont d'ailleurs systématiquement éliminées.

Pour finir, le chapitre [VI](#) conclut ce travail de thèse en présentant une discussion sur les analyses proposées et quelques perspectives pour la suite.

Chapitre II

Etat de l'art

Ce chapitre présente un état de lieux des méthodes statistiques utilisées dans cette thèse. On y présente l'analyse des données longitudinales et des données de survie puis les principes de la modélisation conjointe de biomarqueurs longitudinaux et de la survenue d'un événement d'intérêt pour finir par les outils pronostiques dynamiques issus de cette modélisation conjointe ainsi que leur évaluation.

1 Analyse de données longitudinales

L'analyse de données longitudinales est de plus en plus utilisée dans le domaine de la recherche médicale. Elle permet d'étudier un biomarqueur mesuré de manière répétée au cours du temps. Ces mesures répétées sur un même sujet induisent une certaine corrélation entre les mesures faites sur le même sujet. La théorie du modèle linéaire mixte est particulièrement adaptée pour analyser ce type de données corrélées.

1.1 Modèle linéaire mixte

Le modèle linéaire mixte est une extension du modèle linéaire classique avec, en plus des effets fixes, des effets aléatoires. Il a été introduit par [Harville \(1977\)](#) et décrit par [Laird and Ware \(1982\)](#); c'est le modèle de référence pour étudier l'évolution d'un biomarqueur

Gaussien mesuré au cours du temps. Ce modèle permet de prendre en compte l'évolution moyenne de la population via des effets fixes et les différentes évolutions individuelles à travers des effets aléatoires. Ces effets aléatoires, spécifiques à chaque sujet, représentent les déviations individuelles par rapport à l'évolution moyenne de la population.

Soit une population de N sujets. Notons $Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{in_i}))$, $i = 1, \dots, N$, le vecteur des n_i observations du sujet i mesurées aux temps t_{i1}, \dots, t_{in_i} . Le modèle linéaire mixte s'écrit de manière générale comme :

$$Y_i(t_{ij}) = X_i(t_{ij})^T \beta + Z_i(t_{ij})^T b_i + \epsilon_i(t_{ij}) \quad (\text{II.1})$$

où $X_i(t_{ij})$ et $Z_i(t_{ij})$ sont des vecteurs de covariables dépendant du temps de longueur p et q et respectivement associés au vecteur des effets fixes β et au vecteur des effets aléatoires b_i . Concernant les effets aléatoires, ils sont spécifiques à chaque sujet et sont supposés suivre une loi normale centrée et de matrice de variance-covariance D . Le vecteur des erreurs de mesures $\epsilon_i = (\epsilon_i(t_{i1}), \dots, \epsilon_i(t_{in_i}))$ est un vecteur Gaussien, centré, de matrice de variance covariance Σ_i (de dimension $n_i \times n_i$) et indépendant du vecteur des effets aléatoires b_i . Ces erreurs de mesures sont généralement supposées indépendantes entre elles, pour chaque sujet, et de même variance ($\Sigma_i = \sigma^2 I_{n_i}$). Mais on peut aussi les supposer corrélées entre elles ou bien ajouter aux erreurs indépendantes une structure de corrélation entre les mesures répétées par un processus autorégressif ou un mouvement brownien par exemple. Pour plus de simplicité, nous nous limiterons ici à des erreurs indépendantes.

La partie fixe $X_i(t_{ij}) \beta$ représente l'évolution moyenne du biomarqueur au cours du temps et la partie aléatoire $Z_i(t_{ij}) b_i$ les déviations individuelles par rapport à la moyenne. Le terme $X_i(t_{ij}) \beta + Z_i(t_{ij}) b_i$ représente donc la valeur du biomarqueur au temps t_{ij} dépourvue d'erreur de mesure que l'on nomme souvent la "vraie" valeur non observée du biomarqueur.

1.2 Estimation des paramètres

Pour un même sujet, les mesures répétées sont marginalement corrélées entre elles car elles partagent les mêmes effets aléatoires b_i . Donc, conditionnellement au vecteur des effets aléatoires b_i , les mesures répétées sont indépendantes entre elles. La densité marginale pour le sujet i s'écrit de la manière suivante :

$$f(Y_i; \theta) = \int f(Y_i|b_i; \beta) f(b_i; \alpha) db_i$$

avec $f(Y_i|b_i; \beta) = \prod_{j=1}^{n_i} f(Y_{ij}|b_i; \beta)$ la densité du vecteur de mesures conditionnellement aux effets aléatoires, $f(b_i; \alpha)$ la densité des effets aléatoires et $\theta = (\beta, \alpha)$ le vecteur de tous les paramètres des effets fixes β , des paramètres α dans D et dans Σ_i .

Si les différentes mesures sur un même sujet sont supposées corrélées, les sujets eux sont considérés comme indépendants dans ce type de modèle. En supposant l'indépendance entre les sujets, la vraisemblance du modèle s'écrit :

$$L(\theta) = \prod_{i=1}^N f(Y_i; \theta) \tag{II.2}$$

Dans le cas de modèles linéaires mixtes, l'écriture de $f(Y_i; \theta)$ se simplifie. En effet $Y_i \sim \mathcal{N}(X_i \beta, V_i = Z_i D Z_i^T + \Sigma_i)$ avec X_i et Z_i des matrices de dimension $n_i \times p$ et $n_i \times q$ et de vecteurs lignes $X_i(t_{ij})^T$ et $Z_i(t_{ij})^T$ respectivement.

La densité $f(Y_i; \theta)$ s'exprime donc comme :

$$f(Y_i; \theta) = \left(\frac{1}{2\pi} \right)^{n_i/2} |V_i|^{-1/2} \exp \left\{ -\frac{1}{2} (Y_i - X_i \beta)^T V_i^{-1} (Y_i - X_i \beta) \right\}$$

où $|V_i|$ représente le déterminant de la matrice V_i .

L'estimation du modèle se fait le plus souvent par la méthode de maximum de vraisemblance à travers des algorithmes itératifs tels que l'algorithme EM (Dempster et al., 1977) ou l'algorithme de Newton-Raphson (Lindstrom and Bates, 1988 ; Fletcher, 2000). Lorsque α , le vecteur des paramètres de la matrice de variance-covariance des effets aléatoires D et des erreurs de mesures Σ_i , est connu, l'estimateur sans biais des effets fixes est le BLUE ("Best Linear Unbiased Estimator") et est donné par :

$$\hat{\beta} = \left(\sum_{i=1}^N X_i^T V_i^{-1} X_i \right)^{-1} \sum_{i=1}^N X_i^T V_i^{-1} Y_i \quad (\text{II.3})$$

En pratique, α doit être estimé. Deux solutions peuvent être envisagées ; soit α est estimé par maximum de vraisemblance en remplaçant β par son estimateur $\hat{\beta}$ dans (II.2) ; soit la vraisemblance (II.2) est directement maximisée selon β et α .

1.3 Prédictions et résidus

Les effets aléatoires individuels peuvent être estimés à partir des estimations de θ comme $E(b_i|Y_i; \hat{\theta})$. Ces estimateurs sont appelés BLUP ("Best Linear Unbiased Predictor") ou estimateurs bayésiens empiriques et sont donnés par $\hat{b}_i = E(b_i|Y_i; \hat{\theta}) = \hat{D}Z_i^T \hat{V}_i^{-1} (Y_i - X_i \hat{\beta})$. Deux type de prédictions peuvent en être dérivés : des prédictions marginales aux effets aléatoires et des prédictions spécifiques aux effets aléatoires.

Les prédictions marginales sont données par $\hat{Y}_i = E(Y) = X_i \hat{\beta}$ tandis que les prédictions individuelles spécifiques aux sujets sont estimées par $\hat{Y}_i|\hat{b}_i = E(Y|\hat{b}_i) = X_i \hat{\beta} + Z_i \hat{b}_i$.

Les résidus étant calculés à partir des prédictions, on définit deux types de résidus longitudinaux : les résidus marginaux $Y_i - X_i \hat{\beta}$ et les résidus spécifiques aux sujets $Y_i - X_i \hat{\beta} - Z_i \hat{b}_i$.

1.4 Données manquantes dans les études longitudinales

Les données manquantes sont très courantes dans les études longitudinales et constituent un problème important dans les analyses et particulièrement dans l'estimation des paramètres. Au cours d'une étude, même si les dates de visites sont définies à l'avance, il peut arriver que des mesures prévues dans le protocole ne soient pas collectées. On distingue les *données manquantes monotones* et les *données manquantes intermittentes* (non monotones). On parle de données manquantes monotones lorsque plus aucune donnée n'est collectée au delà d'une certaine visite, c'est le cas lors d'une *sortie d'étude* (appelée "*dropout*"). En présence de ce type de données manquantes, il est essentiel de ne pas dissocier celles-ci du processus de mesure. Une classification des mécanismes de données manquantes a été proposée par [Rubin \(1976\)](#) et développée par [Little and Rubin \(1987, 2002\)](#).

Soit Y_i le vecteur de mesures défini dans la section 1.1. En supposant que Y_i soit composé d'une partie observée Y_i^o et d'une partie manquante Y_i^m , on obtient $Y_i = (Y_i^o, Y_i^m)$. Notons r_i le processus de données manquantes défini par :

$$r_i(t) = \begin{cases} 1 & \text{si } Y_i(t) \text{ est manquant} \\ 0 & \text{sinon} \end{cases}$$

La méthode proposée est basée sur le calcul de la densité du processus r_i conditionnellement au vecteur Y_i , notée $f(r_i|Y_i^o, Y_i^m)$. Les trois types de données manquantes définis sont :

- *MCAR* ("*Missing Completely at Random*") : On parle de données manquantes complètement aléatoires si la probabilité qu'une réponse soit manquante est indépendante de la variable étudiée. La densité du processus r_i s'exprime alors comme

$$f(r_i|Y_i^o, Y_i^m) = f(r_i)$$

- *MAR* ("Missing at Random") : On parle de données manquantes aléatoires si la probabilité qu'une réponse soit manquante dépend uniquement des données observées de la variable étudiée. Avec ce type de données manquantes, la densité du processus r_i se traduit comme

$$f(r_i|Y_i^o, Y_i^m) = f(r_i|Y_i^o)$$

- *MNAR* ("Missing Not at Random") : On parle de données manquantes non aléatoires ou informatives si la probabilité qu'une réponse soit manquante dépend des valeurs non observées de la variable étudiée. Dans ce contexte particulier, on exprime la densité du processus r_i par

$$f(r_i|Y_i^o, Y_i^m) = f(r_i|Y_i^m)$$

Les méthodes inférentielles classiques élaborées pour les données observées de manière complète peuvent ne pas être valides lorsque l'on a des données manquantes. En présence de données MCAR et MAR, le mécanisme de ces données manquantes *peut être ignoré* dans les méthodes d'inférence basées sur la vraisemblance, rendant ainsi sans biais les estimations obtenues (Rubin, 1976). En revanche, en présence de données de type MNAR, ce mécanisme *ne peut être ignoré* dans le calcul de la vraisemblance car cela induirait des problèmes de biais et de perte de puissance dans les estimations (Jacqmin-Gadda et al., 1999); d'où la nécessité de prendre en compte ce mécanisme de données manquantes.

Différentes familles de modèles ont été développées afin de prendre en compte le processus de données manquantes r_i . Elles consistent toutes à modéliser simultanément le vecteur de mesures $Y_i = (Y_i^o, Y_i^m)$ et le processus de données manquantes r_i . Il s'agit des *modèles de sélection* (Heckman, 1976; Wu and Carrol, 1988; Diggle and Kenward, 1994; Verbeke and Molenberghs, 2000; Fitzmaurice et al., 2008), des *pattern mixture models* et des *modèles à paramètre partagé* (Wu and Carrol, 1988; Wu and Bailey, 1988, 1989; Hogan and Laird, 1997; Verbeke and Molenberghs, 2000; Fitzmaurice et al., 2008).

Ces trois familles de modèles permettent de corriger le biais induit par le processus de données manquantes et sont basées sur une factorisation de la densité conjointe :

- Les pattern mixture models : $p(Y_i, r_i; \theta) = p(Y_i | r_i; \theta_y) p(r_i; \theta_r)$. L'évolution du biomarqueur Y est modélisée conditionnellement au processus de données manquantes.
- Les modèles de sélection : $p(Y_i, r_i; \theta) = p(Y_i; \theta_y) p(r_i | Y_i; \theta_r)$. On fait l'hypothèse dans ce modèle que la distribution du processus des données manquantes dépend directement des données complètes Y_i .
- Les modèles à paramètre partagé : $p(Y_i, r_i; \theta) = \int p(Y_i | b_i; \theta_y) p(r_i | b_i; \theta_r) p(b_i; \theta_b) db_i$. On suppose ici que la distribution conjointe dépend des effets aléatoires du modèle mixte qui sont partagés par les données complètes Y_i et par le processus r_i .

Ces deux derniers modèles ouvrent la voie de la modélisation conjointe telle que nous la détaillerons en section 3 et dans laquelle l'événement d'intérêt serait la sortie d'étude et les données manquantes seraient les données post-événement.

2 Analyse de données de survie

L'analyse de données de survie est souvent utilisée dans les études longitudinales de cohortes ou dans les essais cliniques. Elle permet d'étudier le délai passé depuis un temps d'origine jusqu'à la survenue d'un événement d'intérêt. Cet événement est souvent le décès d'où le terme "*survie*", mais on pourra aussi s'intéresser à la rechute d'un cancer ou la sortie d'étude comme évoqué précédemment.

2.1 Quelques définitions

Notons T^* la variable aléatoire continue et positive qui représente la durée de survie depuis l'entrée dans l'étude.

La loi de probabilité T^* est définie par différentes fonctions :

- La fonction de survie est la probabilité de survivre jusqu'à un temps t fixé :

$$S(t) = \mathbb{P}(T^* > t), \quad t \geq 0$$

- La fonction de répartition est la probabilité d'observer l'événement avant t :

$$F(t) = \mathbb{P}(T^* \leq t) = 1 - S(t)$$

- La fonction de densité de probabilité, notée $f(t)$, est la probabilité d'observer l'événement dans un petit intervalle de temps $[t, t + \Delta_t]$. Elle est définie par :

$$f(t) = \lim_{\Delta_t \rightarrow 0^+} \frac{\mathbb{P}(t \leq T^* < t + \Delta_t)}{\Delta_t} = F'(t) = \lambda(t) \exp\left(-\int_0^t \lambda(u) du\right)$$

- Le risque instantané, noté λ , est la probabilité d'observer l'événement dans un petit intervalle de temps $[t, t + \Delta_t]$, sachant que le sujet à survécu jusqu'au temps t :

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

- Le risque cumulé $\Lambda(t)$ est l'intégrale du risque instantané :

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

Dans les études longitudinales, il arrive souvent que l'on soit en présence de données incomplètes. En effet, l'événement d'intérêt n'est pas toujours observé durant l'étude à cause des processus de *censure* ou de *troncature*.

- Une durée de survie est *censurée à droite* si le sujet n'a pas subi l'événement à sa dernière visite. Au lieu d'observer T^* on va observer : $T = \min(T^*, C)$ et $\delta = \mathbb{1}_{\{T^* \leq C\}}$ où C est le temps de censure (indépendant). Il s'agit du problème de données incomplètes le plus fréquent dans les études longitudinales.
- Une durée de survie est dite *censurée à gauche* si le sujet a déjà subi l'événement avant une certaine date, sans qu'il soit possible d'en connaître l'exactitude. Au lieu d'observer T^* on va observer : $T = \max(T^*, C)$ et $\delta = \mathbb{1}_{\{T^* \geq C\}}$.
- Une durée de survie est *censurée par intervalle* lorsque l'événement n'est pas observé à un temps précis mais dans un intervalle de temps.

On parle aussi de *troncature* lorsqu'une durée de survie est conditionnelle à un événement Z . Une durée de survie est *tronquée à droite* par Z si T^* est observée seulement si $T^* < Z$ alors qu'elle est *tronquée à gauche* par Z si T^* est observée seulement si $T^* > Z$. Dans le cas où la durée de survie est tronquée à la fois à gauche et à droite, on parle de durée de survie *tronquée par intervalle*. Par la suite, nous nous limiterons au cas de données incomplètes le plus classique, à savoir les données de survie censurées à droite uniquement.

2.2 Modèle à risques proportionnels

Le modèle à risques proportionnels est le modèle de survie le plus standard pour étudier l'effet de covariables (âge, sexe, traitement, etc) sur la durée de survenue d'un événement d'intérêt (un décès, une rechute, une transplantation, l'initiation d'un traitement, etc) en présence de censure à droite. Pour le sujet i , $i = 1, \dots, N$, nous notons comme précédemment T_i^* le temps jusqu'à l'événement d'intérêt et C_i le temps de censure. C_i peut représenter le temps de suivi des sujets encore à risque à la fin de l'étude ou le temps

jusqu'à leur sortie d'étude. On pose $T_i = \min(T_i^*, C_i)$ et $\delta_i = \mathbb{1}_{\{T_i^* \leq C_i\}}$.

A un temps donné t , la fonction de risque instantané du sujet i est définie comme suit :

$$\lambda_i(t|W_i) = \lambda_0(t) \exp(W_i^T \gamma) \quad (\text{II.4})$$

où $\lambda_0(t)$ est la fonction de risque de base commune à tous les sujets et γ le vecteur de coefficients mesurant l'association entre le vecteur de covariables W_i et le risque instantané d'événement. L'exponentielle de chaque coefficient k , $\exp(\gamma_k)$, donne le risque relatif d'événement pour l'augmentation d'une unité de la covariable associée. Dans le cas où $\gamma > 0$, le risque instantané augmente et a contrario, si $\gamma < 0$ il diminue.

On distingue deux types de modèles à risques proportionnels : le modèle paramétrique dans lequel λ_0 est modélisé avec une fonction de risque de type Weibull ou Exponentiel par exemple, et le modèle semi-paramétrique dans lequel λ_0 reste non spécifié. Ce dernier constitue le modèle le plus connu, le modèle de Cox (Cox, 1972).

2.3 Estimation

L'estimation des paramètres se fait différemment suivant que le modèle est paramétrique ou semi-paramétrique. Dans le cas du modèle paramétrique, l'estimation se fait par maximisation directe de la vraisemblance qui s'écrit : $\prod_{i=1}^N \lambda_i(T_i)^{\delta_i} S(T_i)$.

Des exemples de modèles paramétriques incluent le modèle "piecewise" (une fonction constante par morceaux) où $\lambda_0(t) = \sum_{l=1}^L \xi_l \mathbb{1}_{\{v_{l-1} < t \leq v_l\}}$ avec ξ_l la valeur du risque sur chaque intervalle $]v_{l-1}, v_l]$; le modèle de Weibull où $\lambda_0(t) = \xi_1 \xi_2 (\xi_2 t)^{\xi_1 - 1}$; ou des modèles où le risque de base est approché par des splines qui sont des fonctions polynômiales par morceaux. Le risque peut être approché par une combinaison linéaire de splines positives, des M-splines : $\lambda_0(t) = \kappa_0 + \sum_{l=1}^L \kappa_l M_l(t, d)$ où $(\kappa_0, \kappa_1, \dots, \kappa_L)$ est le vecteur des coefficients de spline, L est le nombre total de nœuds, d est le degré et M la base de fonctions M-splines. Une alternative est d'approcher le logarithme du risque par des B-splines

comme proposé par Rizopoulos (2010).

L'estimation du modèle semi-paramétrique est basée sur la maximisation de la vraisemblance partielle. Cette vraisemblance partielle est obtenue par la factorisation de la vraisemblance ne retenant qu'une partie qui n'implique pas $\lambda_0(t)$. Ainsi le risque de base $\lambda_0(t)$ reste non paramétrique.

2.4 Résidus de martingale et de Cox-Snell

Les résidus de martingale sont définis comme la différence entre l'événement observé et le risque cumulé d'événement au temps d'événement, c'est-à-dire :

$$r_i^{\text{Mart}} = \delta_i - \int_0^{T_i} \lambda_i(t|W_i; \hat{\theta}) dt$$

Une fois sommés, ils peuvent être vus comme la différence entre le nombre d'événements observés et le nombre d'événements prédits. Ces résidus permettent ainsi d'évaluer la capacité du modèle à prédire l'événement. Cependant, ils permettent également en évaluant la forme fonctionnelle d'une covariable, c'est-à-dire la façon dont la covariable est associée à l'événement de vérifier l'hypothèse de log-linéarité du modèle .

Les résidus de Cox-Snell permettent de vérifier si le modèle ajuste bien les données. Ils sont basés sur le fait que si T_i est un temps d'événement ayant pour risque cumulé Λ alors $\Lambda(T_i)$ suit une loi exponentielle de paramètre 1. Les résidus de Cox-Snell sont donc définis comme :

$$r_i^{\text{CS}} = \int_0^{T_i} \lambda_i(t|W_i; \hat{\theta}) dt$$

Pour vérifier l'ajustement global du modèle, leur fonction de survie estimée par Kaplan-Meier qui doit donc être proche de la fonction de survie d'une loi exponentielle de paramètre 1 ($\exp(-t)$).

2.5 Inclusion de covariables dépendant du temps

Le modèle à risques proportionnels comme défini dans (II.4) peut prendre en compte des covariables constantes W_i , connues au "baseline" c'est à dire à l'entrée dans l'étude, mais aussi des covariables dépendant du temps $W_i(t)$ comme des conditions météorologiques, un traitement, etc.

L'écriture du modèle (II.4) devient alors :

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

Avec un risque de base modélisé de façon paramétrique, la vraisemblance s'écrit :

$$\begin{aligned} L(\theta) &= \prod_{i=1}^N \lambda_i(T_i)^{\delta_i} S(T_i) \\ &= \prod_{i=1}^N \left[\lambda_0(T_i) \exp(W_i(T_i)^T \gamma) \right]^{\delta_i} \exp\left(-\int_0^{T_i} \lambda_0(t) \exp(W_i(t)^T \gamma) dt\right) \end{aligned} \quad (\text{II.5})$$

L'estimation des paramètres dans (II.5) implique que W_i soit connu en tout temps.

Dans le modèle de Cox, de façon similaire, le calcul de la vraisemblance partielle nécessite que la covariable dépendant du temps soit connue à tous les temps d'événement ([Kalbfleisch and Prentice, 2002](#)).

Dans ce contexte, la vraisemblance partielle du modèle de Cox s'écrit comme :

$$L(\gamma) = \prod_{i=1}^N \left[\frac{\exp(W_i(T_i)^T \gamma)}{\sum_{j \in \mathcal{R}(T_i)} \exp(W_j(T_i)^T \gamma)} \right]^{\delta_i} \quad (\text{II.6})$$

où $\mathcal{R}(T_i)$ est l'ensemble des sujets encore à risque au temps d'événement du sujet i .

L'inclusion de covariables dépendant du temps en survie repose sur plusieurs hypothèses :

- seules les covariables dites *exogènes* ou *externes* peuvent être considérées dans cette approche. Une covariable *externe* a une trajectoire indépendante du processus de l'événement étudié. La valeur de la covariable au temps t n'est pas affectée par la survenue d'un événement à un temps antérieur s ($0 < s \leq t$). Cela se traduit par la relation :

$$\mathbb{P}(W_i(t)|W_i(s), T_i = s) = \mathbb{P}(W_i(t)|W_i(s), T_i > s)$$

où $W_i(s)$ est l'histoire de la covariable jusqu'au temps s .

La pollution de l'air est un bon exemple de covariable *externe*. Ces covariables *externes* sont à distinguer des covariables dites *internes* ou *endogènes* et décrites par [Kalbfleisch and Prentice \(2002\)](#). Les covariables dépendant du temps *internes* sont souvent des mesures prises sur les sujets et pour lesquelles l'observation dépend de l'événement : $\mathbb{P}(W_i(t)|W_i(s), T_i = s) \neq \mathbb{P}(W_i(t)|W_i(s), T_i > s)$. Par exemple, pour l'événement décès, les covariables internes ne peuvent être mesurées que si le sujet est encore en vie et à risque de l'événement. Des mesures répétées de PSA chez des patients atteints d'un cancer de la prostate ou le nombre de cellules de CD4 chez des patients infectés par le VIH sont des exemples de covariables internes.

- les covariables doivent être non bruitées, c'est-à-dire mesurées sans erreurs.
- les valeurs des covariables doivent être connues à tout temps, ou aux temps d'événement. En pratique, il n'est pas rare que W_i soit collectée de façon intermittente. Il est alors nécessaire d'imputer les valeurs manquantes, par exemple, par la dernière valeur connue. Cette méthode est connue sous le nom de *LOCF* ("*Last Observed Carried Forward*"). [Prentice \(1982\)](#) a montré qu'avec ce type d'imputation, les estimations d'un modèle de survie étaient biaisées.

Les modèles de survie ne sont donc pas adaptés en présence de covariables dépendant du temps telles que les biomarqueurs. En effet, les biomarqueurs ont comme caractéristique principale d'être des covariables internes ; la survenue de l'événement impacte la valeur du biomarqueur. De plus, ils sont mesurés avec des erreurs, que ces erreurs soient dues à l'appareil de mesure ou aux différentes variations biologiques induites par le sujet lui-même. Enfin, les biomarqueurs sont collectés uniquement au cours des visites du patient, ce qui veut dire qu'ils ne sont pas connus à tous temps mais seulement à des temps discrets.

3 Modélisation conjointe pour données longitudinales et de temps d'événement

Pour évaluer la relation entre un biomarqueur longitudinal et un temps d'événement, nous avons vu précédemment que les modèles à risques proportionnels classiques n'étaient pas adaptés puisque les biomarqueurs sont des covariables internes. Nous avons aussi vu en section 1.4 que la sortie d'étude informative engendrait des données manquantes du biomarqueur, ce qui est une source de biais dans les estimations du modèle linéaire mixte. Pour remédier à ces problèmes, il est nécessaire de modéliser *conjointement* les données longitudinales du biomarqueur et du temps d'événement. Les modèles qui en sont dérivés sont naturellement appelés des *modèles conjoints*.

3.1 Approche "Two-stage" : les prémisses de la modélisation conjointe

Suite aux limitations que présente le modèle de Cox, un modèle en deux étapes appelé modèle "Two-stage" a été développé. La première étape de ce modèle consiste à appliquer un modèle linéaire mixte sur les données longitudinales du biomarqueur (équation (1) -

II.7) et calculer les prédictions individuelles \hat{Y}_i de chaque sujet à tous les temps d'événement en utilisant les estimateurs bayésiens empiriques $\hat{b}_i = E(b_i | Y_i; \hat{\theta})$. La seconde étape permet d'ajuster un modèle de Cox en utilisant ces prédictions individuelles du biomarqueur, obtenues à la première étape, comme covariable dépendant du temps (équation (2) - II.7).

$$\begin{cases} Y_i(t) &= X_i(t) \beta + Z_i(t) b_i + \epsilon_i(t) & (1) \\ \lambda_i(t|Y_i) &= \lambda_0(t) \exp(\hat{Y}_i(t)\gamma) & (2) \end{cases} \quad (\text{II.7})$$

où (1) est estimée à travers la vraisemblance du modèle linéaire mixte (II.2) et (2) à travers la vraisemblance partielle du modèle de Cox (II.6).

Deux variantes de l'approche "Two-stage" existent. Dans la première version de cette approche proposée par Tsiatis et al. (1995), le modèle mixte était estimé à chaque temps d'événement T_k^* en utilisant uniquement les sujets à risque à T_k^* et les mesures du biomarqueur collectées jusqu'à ce temps. Les auteurs considéraient une approximation de la fonction de risque dans la vraisemblance partielle et utilisaient l'espérance conditionnelle du biomarqueur sachant l'histoire du sujet, encore à risque, jusqu'au temps t soit $E(Y_i(t) | Y_i^{(t)}, T_i^* \geq t)$ à la place de $\hat{Y}_i(t)$ dans II.7. Dans cette espérance, les paramètres utilisés changeaient donc à chaque temps d'événement. Le problème majeur de cette approche est qu'estimer un modèle mixte à chaque temps d'événement devient vite difficile à effectuer en pratique si bien que Ye et al. (2008) ont proposé une deuxième variante pour alléger l'estimation. Cette variante consiste à n'estimer le modèle mixte qu'une seule fois sur l'ensemble des données. Le problème majeur de cette seconde variante est qu'elle utilise les données postérieures à T_k^* dans l'estimation des prédictions individuelles aux différents temps d'événement, ce qui peut induire certains biais.

Comparées à la méthode LOCF, ces deux variantes de l'approche "Two-stage" permettent de réduire les biais mais ne les éliminent pas complètement (De Gruttola and Tu,

1994; Tsiatis et al. 1995; Hu et al., 1998; Dafni and Tsiatis, Dafni and Tsiatis; Ye et al., 2008; Albert and Shih 2010; Sweeting and Thompson, 2011). En effet, les données manquantes du biomarqueur dues à la sortie d'étude entraînent un biais dans les estimations des paramètres du modèle linéaire mixte. De plus, les incertitudes de ces estimations ne sont pas prises en compte dans les estimations du modèle de Cox; ce qui biaise au final les estimations de ce modèle.

Albert and Shih (2010) ont proposé, afin de compléter les données manquantes du biomarqueur, de générer des données suivant la distribution conditionnelle du biomarqueur sachant le temps d'événement. Une fois les données complètes obtenues, les deux étapes du "Two-stage" peuvent être appliquées. Les auteurs ont montré, à travers des simulations, que leur approche était moins biaisée que celle proposée par Ye et al. (2008).

3.2 Principe de la modélisation conjointe

Les *modèles conjoints* ont été développés afin de pallier aux différents problèmes induits par la modélisation séparée d'un biomarqueur longitudinal et d'un temps d'événement (Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Henderson et al., 2000; Wang and Taylor, 2001; Xu and Zeger, 2001).

Ces modèles conjoints offrent ainsi un cadre intéressant pour :

- (i) évaluer la trajectoire longitudinale du biomarqueur et son association avec des covariables, sans le biais de sélection induit par la sortie d'étude informative (Yu et al., 2004). En effet, comme décrit en section 1.4 les données manquantes du biomarqueur dues à une sortie d'étude induisent des biais dans les estimations des modèles mixtes classiques.
- (ii) évaluer le risque d'événement et son association avec des covariables incluant les données longitudinales du biomarqueur, sans le biais induit par les données manquantes

aux différents temps d'événement et les erreurs de mesure (Prentice, 1982). C'est le cas, par exemple, de l'étude de l'association entre le nombre de CD4 et la survie chez des patients atteints du SIDA, où le biomarqueur est mesuré périodiquement et avec un erreur (Tsiatis et al., 1995 ; Wulfsohn and Tsiatis, 1997).

- (iii) explorer directement l'association entre les processus longitudinal et de temps d'événement. Nous pouvons donner l'exemple de Proust-Lima et al. (2009b) qui étudient l'association entre le déclin cognitif et le risque de démence chez des personnes âgées. Un autre exemple, qui est abordé au cours de cette thèse, est l'étude de l'association entre l'évolution longitudinale du PSA et le risque de rechute clinique.
- (iv) pronostiquer le risque de survenue d'un événement d'intérêt en fonction de l'évolution d'un biomarqueur (Proust-Lima and Taylor, 2009a ; Rizopoulos, 2011b). Ce dernier point est l'objectif principal de cette thèse.

Les *modèles conjoints* consistent à combiner un *modèle linéaire mixte* et un *modèle de survie*, et à lier ces deux processus par une *structure latente commune* (Figure II.1). Cette structure latente capture l'association entre les deux processus, les rendant ainsi indépendants deux à deux conditionnellement à cette dernière. Les modèles conjoints permettent ainsi de modéliser en même temps les deux processus. A travers une vraisemblance conjointe, les paramètres des deux sous-modèles sont estimés simultanément.

En modélisation conjointe de données longitudinales et de temps d'événement, on peut distinguer principalement deux types d'approches : les modèles conjoints à effets aléatoires partagés avec une structure latente continue (des effets aléatoires) et les modèles conjoints à classes latentes avec une structure latente discrète (des classes latentes). Nous revenons en détail sur chacune d'elles dans les paragraphes suivants.

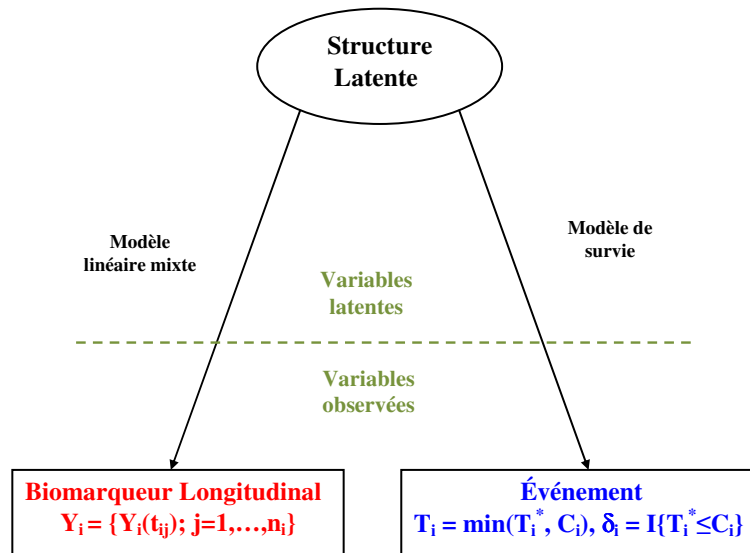


Figure II.1 – Graphe orienté structurel de la modélisation conjointe de données longitudinales et de temps d'événement

3.3 Modèles conjoints à effets aléatoires partagés

Les modèles à effets aléatoires partagés ont été développés à la fin des années 90 dans le cadre d'études longitudinales dans lesquelles un biomarqueur et un temps d'événement corrélés sont collectés en même temps (Wulfsohn and Tsiatis, 1997; Henderson et al., Henderson et al.). Ces modèles étendent directement le principe des modèles de survie avec covariables dépendant du temps et l'approche "Two-stage". Ils consistent à modéliser le profil individuel d'évolution du biomarqueur par un modèle mixte (voir section 1.1) et le risque d'événement comme, entre autre, une fonction des caractéristiques individuelles du biomarqueur. Ces caractéristiques sont des fonctions des effets aléatoires du modèle mixte, d'où le nom de modèle à effets aléatoires partagés. Plusieurs revues de ces modèles ont été publiées (Ibrahim et al., 2001; Tsiatis and Davidian, 2004; Ibrahim et al., 2010; Sousa, 2011; Wu et al., 2012).

3.3.1 Spécification du sous-modèle longitudinal

Pour le sujet i , $i = 1, \dots, N$, nous notons comme précédemment T_i^* le vrai temps de l'événement d'intérêt et C_i le temps de censure. On observe alors le temps $T_i = \min(T_i^*, C_i)$ et l'indicateur d'événement $\delta_i = \mathbb{1}_{\{T_i^* \leq C_i\}}$. De plus $Y_i = (Y(t_{i1}), \dots, Y(t_{in_i}))$ représente le vecteur des n_i observations du biomarqueur collectées de manière intermittente aux temps t_{i1}, \dots, t_{in_i} jusqu'au temps T_i .

Dans la suite, on définit également Y_i^* la version non bruitée de ces mesures qui représente la "vraie" valeur du biomarqueur. La trajectoire du biomarqueur est alors décrite par un modèle linéaire mixte comme présenté dans la section 1.1 :

$$\begin{aligned} Y_i(t_{ij}|X_{Li}, Z_i; \theta_y) &= Y_i^*(t_{ij}|X_{Li}, Z_i; \theta_y) + \epsilon_i(t_{ij}) \\ &= X_{Li}(t_{ij})^T \beta + Z_i(t_{ij})^T b_i + \epsilon_i(t_{ij}) \end{aligned} \quad (\text{II.8})$$

$X_{Li}(t_{ij})^T \beta$ modélise l'évolution moyenne du biomarqueur, $Z_i(t_{ij})^T b_i$ décrit les déviations individuelles par rapport à cette moyenne et $\epsilon_i(t_{ij})$ est une erreur de mesure. Plus précisément, β est le vecteur des effets fixes associés au vecteur des covariables dépendant du temps $X_{Li}(t_{ij})$; b_i est le vecteur des effets aléatoires supposé Gaussien $\mathcal{N}(0, D)$ et associé au vecteur des covariables dépendant du temps $Z_i(t_{ij})$. Le vecteur $\epsilon_i = (\epsilon_i(t_{i1}), \dots, \epsilon_i(t_{in_i}))$ des erreurs de mesures est supposé Gaussien, centré, indépendant et identiquement distribué $\mathcal{N}(0, \Sigma_i = \sigma I_{n_i})$; ϵ_i et b_i sont supposés indépendants. Le vecteur de tous les paramètres associés au processus longitudinal est noté θ_y .

Les premières modélisations, qui ont porté sur le nombre de CD4 chez des patients atteints du SIDA ([Faucett and Thomas \(1996\)](#) et [Wulfsohn and Tsiatis \(1997\)](#)) ont considéré une trajectoire linéaire du biomarqueur par une fonction $Z(t)$ linéaire en t . Depuis, des structures de modélisation plus complexes ont été proposées. Par exemple, [Pauler and Fin-](#)

kelstein (2002), Ghosh et al. (2011) et Tapsoba et al. (2011) ont considéré une trajectoire longitudinale en plusieurs phases d'évolution via des changements de pentes aléatoires. Brown et al. (2005) et Ding and Wang (2008) ont proposé d'approcher la trajectoire non-linéaire du biomarqueur par des splines.

Concernant les erreurs de mesure, les premières modélisations ont considéré des erreurs indépendantes (Tsiatis et al., 1995; Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Henderson et al., (2000) ce qui reste encore très utilisé (Rizopoulos, 2010; Sweeting and Thompson, 2011; Crowther et al., 2012). Par la suite, d'autres structures de corrélation ont été utilisées comme un processus Gaussian stationnaire (Xu and Zeger, 2001) ou un processus stochastique Ornstein-Uhlenbeck (Wang and Taylor, 2001).

3.3.2 Spécification du sous-modèle de survie

Le risque d'événement est modélisé à l'aide d'un modèle de survie, le plus souvent par un modèle à risques proportionnels. Nous avons donc choisi de nous limiter à ce cas. Le modèle le plus souvent utilisé est le suivant :

$$\lambda_i(t|b_i, X_{Ei}; \theta_e) = \lambda_0(t; \theta_e) \exp\left(X_{Ei}^T \gamma + Y_i^*(t) \varphi\right) \quad (\text{II.9})$$

où λ_0 est la fonction de risque de base; elle peut être paramétrique ou non. X_{Ei} est un vecteur de covariables associé au vecteur des paramètres γ qui mesure l'effet des covariables X_{Ei} sur le risque d'événement. La seule différence par rapport à un modèle de survie standard est l'ajout de $Y_i^*(t)$ qui est le "vrai" niveau courant du biomarqueur issu du modèle (II.8). Le paramètre φ mesure l'effet du niveau courant du biomarqueur sur le risque d'événement. Ce modèle (II.9) est un prolongement direct du modèle à risque proportionnel classique et du modèle "Two-stage" présentés précédemment. Le cas particulier $\varphi = 0$ correspond au modèle où le processus longitudinal et celui du temps d'événement sont

indépendants. Dans la suite, le vecteur des paramètres du modèle de survie est noté θ_e .

Différentes fonctions de dépendance peuvent être considérées. [Faucett and Thomas \(1996\)](#) et [Wulfsohn and Tsiatis \(1997\)](#) ont considéré le niveau courant du biomarqueur dans le modèle de survie comme présenté dans l'équation [II.9](#), mais d'autres fonctions de dépendance ont été envisagées par la suite comme une dépendance directe à travers les effets aléatoires ([Jacqmin-Gadda et al., 2004](#)) ou à travers le niveau courant et la pente courante du biomarqueur ([Yu et al., 2008](#)). Plutôt que d'introduire une fonction des effets aléatoires classiques, [Henderson et al. \(2000\)](#) ont introduit un processus latent Gaussien bivarié corrélé dont une composante intervenait dans le modèle longitudinal et l'autre dans le modèle de survie. D'autres exemples de dépendance, entre ces deux processus, peuvent être utilisées. Dans le chapitre suivant ([Chapitre III](#)), nous détaillerons et comparerons différentes spécifications de la dépendance entre le processus longitudinal et le temps d'événement.

3.3.3 Estimation du modèle

Les modèles conjoints à effets aléatoires partagés reposent sur l'hypothèse d'indépendance entre le processus longitudinal et celui du temps d'événement conditionnellement aux effets aléatoires (voir [Figure II.2](#)). Dans cette approche, le vecteur des effets aléatoires b_i a la particularité de prendre en compte à la fois la corrélation entre l'évolution de biomarqueur et la survenue de l'événement et la corrélation existante entre les mesures répétées d'un même sujet.

Soit $\theta = (\theta_y, \theta_e, \theta_b)$ le vecteur de tous les paramètres avec θ_y le vecteur des paramètres du sous-modèle longitudinal, θ_e le vecteur des paramètres du sous-modèle de survie et θ_b le vecteur des paramètres de la matrice de variance-covariance des effets aléatoires. Tous les paramètres des deux sous-modèles sont estimés simultanément à travers la log-vraisemblance conjointe qui s'écrit, grâce à l'hypothèse d'indépendance conditionnelle,

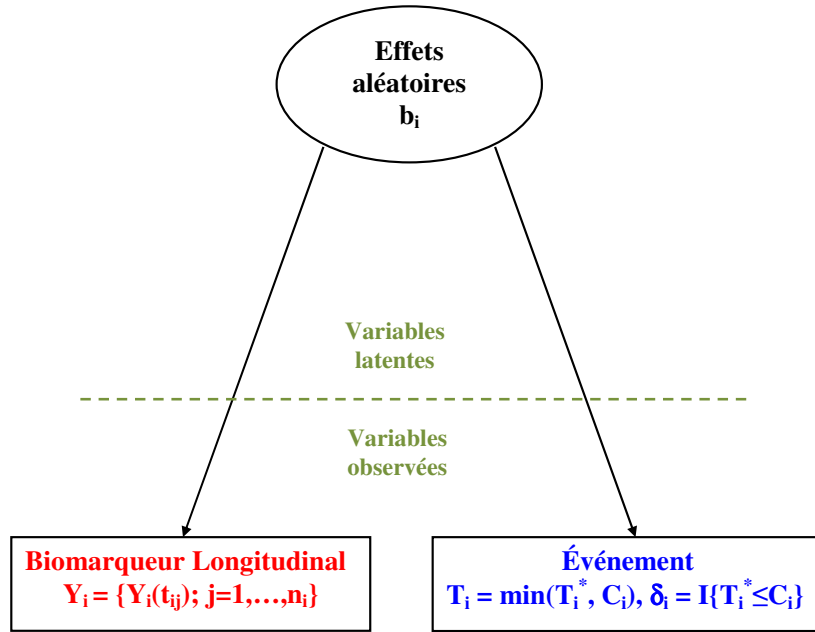


Figure II.2 – Graphe orienté du modèle conjoint à effets aléatoires partagés

comme :

$$l(\theta) = \sum_{i=1}^N \log \left(\int f(Y_i|b_i; \theta_y) \lambda_i(T_i|b_i; \theta_e)^{\delta_i} S_i(T_i|b_i; \theta_e) f(b_i; \theta_b) db_i \right) \quad (\text{II.10})$$

où $f(Y_i|b_i; \theta_y)$ et $f(b_i; \theta_b)$ sont les fonctions de densité de lois normales multivariées d'espérance $X_{Li} \beta + Z_i b_i$ et 0 respectivement et de matrices de variance-covariance Σ_i et D . Les matrices X_{Li} et Z_i sont des matrices de vecteurs lignes $X_{Li}(t_{ij})^T$ et $Z_i(t_{ij})^T$ respectivement et définies dans la section 3.3.1. Le terme $\lambda_i(T_i|b_i; \theta_e)^{\delta_i} S_i(T_i|b_i; \theta_e)$ représente la densité d'un temps d'événement pour données censurées à droite, $\lambda_i(T_i|b_i; \theta_e)$ est la fonction de risque instantané et $S_i(T_i|b_i; \theta_e) = \exp\left(-\int_0^{T_i} \lambda_i(t|b_i; \theta_e) dt\right)$ la fonction de survie associée introduites dans la section 2.1.

Les intégrales présentes dans la fonction de survie et sur les effets aléatoires n'ont généralement pas de solutions analytiques et nécessitent d'être approximées à chaque calcul de

la log-vraisemblance. Les techniques d'intégration numériques telles que les quadratures de Gauss ou la méthode de Monte-Carlo (Wulfsohn and Tsiatis, 1997 ; Henderson et al., 2000 ; Lin et al., 2002) sont habituellement utilisées. Lorsque la dimension des effets aléatoires est très grande, Rizopoulos et al. (2009) suggèrent d'utiliser la méthode d'approximation de Laplace.

Les méthodes de quadrature de Gauss sont des méthodes d'intégration numérique qui permettent d'approcher la valeur de l'intégrale d'une fonction. Elles consistent à remplacer le calcul de l'intégral par une somme pondérée de la fonction prise à un certain nombre de points du domaine d'intégration. La qualité de l'approximation augmente avec le nombre de points de quadrature. Cependant, si les points de quadrature ne sont pas bien positionnés suivant l'endroit où la fonction a sa plus grande masse, cela peut donner des mauvaises approximations de l'intégrale. Pour avoir plus de précision dans les approximations, il peut être utile de repositionner les points de quadrature là où l'intégrande a d'importantes variations. Cette technique s'appelle la méthode de quadrature de Gauss *adaptive*. Elle permet aussi d'obtenir de bonnes approximations avec un petit nombre de points. Cependant, elle nécessite en contrepartie à chaque itération et pour chaque sujet de replacer les points de quadrature autour du mode de la fonction ce qui se fait le plus souvent pour un algorithme itératif d'estimation .

Dans des modèles aussi complexes que les modèles conjoints à effets aléatoires partagés par exemple, approcher des intégrales telles que celle sur les effets aléatoires peut donc être, très demandeur en temps de calcul même par la méthode *adaptive*. Récemment, Rizopoulos (2012) a proposé une technique d'intégration de Gauss-Hermite dite *pseudo-adaptive* qui permet d'obtenir une meilleure approximation de l'intégrale sur les effets aléatoires que la méthode classique de Gauss-Hermite. Tout comme la quadrature *adaptive*, elle consiste à centrer l'intégrale. Mais au lieu de center sur le mode de la densité conjointe, la méthode *pseudo-adaptive* le fait sur le mode de la distribution a posteriori

des effets aléatoires du modèle linéaire mixte. Ainsi, cette technique pseudo-adaptative ne nécessite pas d'utiliser beaucoup de points de quadrature par rapport à la méthode classique et est numériquement plus simple et rapide que la méthode adaptative (Rizopoulos, 2010).

L'estimation du modèle conjoint à effets aléatoires partagés se fait par la méthode du maximum de vraisemblance. Zeng and Cai (2005) ont montré que cet estimateur du maximum de vraisemblance a de bonnes propriétés asymptotiques. L'algorithme de type Newton-Raphson ou l'algorithme EM, où les effets aléatoires sont considérées comme des données manquantes, peuvent être utilisés pour l'estimation (Wulfsohn and Tsiatis, 1997; Henderson et al., 2000). Une combinaison des deux algorithmes peut être éventuellement considérée du fait de la lenteur de convergence de l'algorithme EM lorsque l'on se rapproche du maximum (Lin et al., 2002; Rizopoulos et al., 2009). Notons également que l'estimation du modèle peut être réalisée en utilisant une approche bayésienne (Faucett and Thomas, 1996; Wang and Taylor, 2001; Xu and Zeger, 2001; Ghosh et al., 2011).

Les packages R JM (Rizopoulos, 2010; Rizopoulos, 2012) et joiner (<http://cran.r-project.org/web/packages/joiner/>) permettent d'estimer les modèles conjoints à effets aléatoires partagés. Le package JM a souvent été utilisé dans cette thèse; soit directement, soit après avoir modifié le code source lorsque c'était nécessaire. Ce package permet d'estimer des modèles conjoints à effets aléatoires partagés pour des données longitudinales Gaussiennes et des données de temps d'événement dans le cadre du maximum de vraisemblance.

Il permet de considérer trois sortes de dépendance entre le processus longitudinal et celui du temps d'événement : dépendance à travers le niveau courant du biomarqueur, à travers la pente courante du biomarqueur, ou à travers les deux. Il permet également de spécifier le type de fonction de risque de base (weibull, constante par morceaux, splines, etc), le type de méthode d'intégration numérique pour les effets aléatoires (quadrature

Gauss-Hermite pseudo-adaptative ou non adaptative, approximation de Laplace), l'algorithme d'estimation (EM, Newton-Raphson, ou la combinaison des deux). Il permet enfin de faire de l'évaluation de modèles avec les prédictions et les résidus. Le package JM est relativement simple d'utilisation grâce à un tutoriel détaillé et représente également un outil flexible de par ses différentes déclinaisons pour l'estimation de modèles conjoints à effets aléatoires partagés.

3.3.4 Évaluation du modèle

Comme tout modèle statistique, le modèle conjoint à effets aléatoires partagés doit être évalué. Son évaluation passe par l'étude des prédictions et des résidus des deux sous-modèles. Des prédictions marginales et spécifiques aux sujets peuvent être définies pour le modèle longitudinal. Les résidus qui sont dérivés de ces prédictions, comme présenté dans la section 1.3, permettent de vérifier certaines hypothèses du modèle linéaire mixte notamment les hypothèses de normalité ou d'homoscédasticité des erreurs de mesures. Les résidus du modèles de survie, martingale et Cox-Snell, présentés dans la section 2.4 pour un modèle de survie classique, restent valables également pour le modèle conjoint à effets aléatoires partagés et permettent d'évaluer le sous-modèle de survie.

3.3.5 Extensions du modèle

Diverses extensions du modèle conjoint à effets aléatoires partagés ont été proposées. On peut citer notamment le développement de modèles conjoints pour des données longitudinales multiples et des temps d'événement (Lin et al., 2002 ; Brown et al., 2005 ; Thiébaud et al., 2005 ; Chi and Ibrahim, 2006 ; Rizopoulos and Ghosh, 2011a ; He and Luo, 2013). Cette modélisation permet de prendre en compte les interactions entre plusieurs biomarqueurs longitudinaux et leurs associations avec la survenue d'un événement.

[Williamson et al. \(2008\)](#), [Elashoff et al. \(2008\)](#) et [Huang et al. \(2011\)](#) ont proposé une modélisation conjointe de données longitudinales et de plusieurs temps d'événement. Ce type de modélisation instaure la notion de risques compétitifs et permet de quantifier l'association entre les données répétées et les différents types d'événement considérés.

Les modèles de guérison ("cure models") sont des modèles de survie qui ont été développés afin de considérer la possible guérison de la maladie. En effet, certains patients guérissent et sont immunisés contre une possible rechute. Certains auteurs ont également étendu ces modèles de guérison pour incorporer des données longitudinales ([Law et al., 2002](#); [Brown and Ibrahim, 2003](#); [Yu et al., 2004](#); [Yu et al., 2008](#)).

En guise de dernier exemple, citons le développement de modèles conjoints de données longitudinales et de temps d'événement utilisant un modèle linéaire généralisé permettant une flexibilité dans la modélisation des données longitudinales discrètes ou non Gaussiennes ([Rizopoulos and Ghosh, 2011a](#); [Viviani et al., 2013](#); [He and Luo, 2013](#)).

3.3.6 Domaines d'application du modèle

Les domaines d'application du modèle conjoint à effets aléatoires partagés sont diverses. Par exemple, [Wulfsohn and Tsiatis \(1997\)](#) ont étudié la trajectoire du nombre de CD4 et son association avec la survie chez des sujets atteints de VIH. [Piccorelli and Schluchter \(2012\)](#) ont, quant à eux, modélisé conjointement la fonction pulmonaire et la survie chez des patients atteints de mucoviscidose. Plus récemment, [Abdi et al. \(2013\)](#) ont investigué la relation entre l'exposition longitudinale à l'acide mycophénolique, qui est un inhibiteur utilisé pour prévenir les rejets après la transplantation d'un organe, et le risque de rejet après une transplantation rénale. Appliqué à la maladie de Parkinson, [He and Luo \(2013\)](#) ont modélisé conjointement les réponses d'items multi-niveaux, comme covariables longitudinales, et un événement terminal dépendant pour évaluer l'effet d'un traitement chez des patients atteints précocement de la maladie de Parkinson. [Jacqmin-Gadda et al. \(2006\)](#)

et [Graham et al. \(2011\)](#) ont modélisé conjointement l'évolution du déclin cognitif et le risque de démence. Pour finir, plusieurs modèles conjoints ont été définis pour analyser la dynamique des PSA et la survenue d'une rechute clinique ([Pauler and Finkelstein, 2002](#); [Yu et al., 2008](#)).

3.3.7 Limites du modèle

Les modèles conjoints à effets aléatoires partagés présentent quelques inconvénients. Tout d'abord, la vraisemblance du modèle n'a pas de forme exacte et doit être approchée par les techniques d'intégration numérique. Au cours de cette approximation, le temps de calcul peut augmenter de manière exponentielle en fonction de la complexité du modèle et du nombre de points de quadratures utilisés ([Rizopoulos, 2010](#)). Le problème est que si un nombre de point trop faible est utilisé, même avec les méthodes pseudo-adaptative ou adaptative, l'approximation sera mauvaise, ce qui peut avoir des conséquences sur la qualité des estimations.

De plus, les effets aléatoires jouent le rôle de structure de corrélation à la fois entre le processus longitudinal et le processus de survie, et entre les mesures répétées du biomarqueur chez un même sujet. Ainsi, il n'y a pas de stricte différenciation entre ces deux types de corrélation. Une autre caractéristique essentielle de ces modèles est qu'ils supposent a priori un certain type d'association entre le biomarqueur et le temps d'événement. Par exemple, on peut supposer que le niveau courant du biomarqueur est l'unique source de corrélation entre le processus longitudinal et de survie. Ce type d'hypothèse a priori sur la nature de la dépendance est primordiale pour tester des hypothèses spécifiques, notamment dans la recherche de biomarqueur de substitution. Cela étant, ces modèles peuvent aussi reposer sur une hypothèse d'association fautive et aussi potentiellement biaiser l'inférence statistique. Enfin, ces modèles conjoints reposent sur un modèle linéaire mixte standard et donc supposent une homogénéité des profils d'évolution du biomarqueur dans toute la

population là où on pourrait attendre une population hétérogène vis à vis de l'événement au moins.

Face à ces problèmes numériques, ces problèmes d'hypothèse sur la nature de la dépendance et afin de pouvoir prendre en compte l'existence d'une population hétérogène avec différents profils d'évolution du biomarqueur, de nouveaux modèles ont été développés : ce sont les modèles conjoints à classes latentes.

3.4 Modèles conjoints à classes latentes

Les modèles conjoints à classes latentes sont plus récents et moins populaires que les modèles à effets aléatoires. Ils sont inspirés des modèles de mélange fini et cherchent à prendre en compte la corrélation entre les processus longitudinal et de survie par l'identification de sous-populations de sujets. Ils consistent à considérer que la population d'étude n'est plus homogène, c'est-à-dire caractérisée par une trajectoire moyenne du biomarqueur et des déviations individuelles Gaussiennes par rapport à celle-ci, mais qu'elle est intrinsèquement hétérogène, c'est-à-dire constituée de plusieurs sous-groupes de personnes.

Cette hypothèse d'hétérogénéité est vraisemblable dans beaucoup de contextes de santé surtout en présence d'une population à risque d'événements cliniques comme dans la plupart des études de cohorte. Chez des patients ayant été traités par radiothérapie par exemple, les trajectoires de PSA post-traitement peuvent être très hétérogènes entre les sujets ayant eu ou non une rechute clinique.

Les modèles conjoints à classes latentes peuvent être vus comme des extensions directes des modèles mixtes à classes latentes qui permettent, en l'absence d'événement associé, d'identifier des clusters de sujets ayant des profils homogènes et à les lier à travers des classes latentes ([Verbeke and Lesaffre, 1996](#) ; [Muthén and Shedden, 1999](#)).

3.4.1 Spécification du modèle

On suppose que la population d'étude de N sujets est composée de G sous-populations homogènes, non identifiées, et on note c_i la classe ou sous-population à laquelle appartient le sujet i ($i = 1, \dots, N$). Chaque sujet appartient à une et une seule classe.

Cette variable c_i étant latente, c'est la probabilité π_{ig} d'appartenance du sujet i à la classe g que l'on modélise. On définit pour cela le plus souvent un modèle de régression logistique multinomiale :

$$\pi_{ig} = \mathbb{P}(c_i = g | X_{pi}; \xi) = \frac{\exp(\xi_{0g} + X_{pi}^T \xi_{1g})}{\sum_{l=1}^G \exp(\xi_{0l} + X_{pi}^T \xi_{1l})} \quad (\text{II.11})$$

où ξ_{0g} est l'intercept pour la classe g et ξ_{1g} le vecteur de paramètres spécifiques à la classe g associé au vecteur de covariables X_{pi} . Pour que le modèle soit identifiable, on suppose $\xi_{01} = 0$ et $\xi_{11} = 0$. Le paramètre ξ_{1g} représente le log de l'odds ratio de la probabilité d'appartenir à la classe g par rapport à la classe 1 (classe de référence) pour une augmentation d'une unité du vecteur de covariables X_{pi} . Le vecteur des paramètres est noté $\xi = ((\xi_{0g}, \xi_{1g})_{g=2, \dots, G})$.

Chaque classe latente est caractérisée par une trajectoire du biomarqueur et un risque d'événement spécifique à la classe, et conditionnellement à la classe latente, les deux processus sont supposés indépendants. Ce type de modèle est donc un cas particulier du modèle conjoint décrit dans la Figure II.1 et dans laquelle la structure latente est une variable discrète. Comme schématisé en Figure II.3, le modèle conjoint à classes latentes suppose que le biomarqueur est relié à la classe latente à travers des effets aléatoires.

La trajectoire du biomarqueur dans chaque classe g ($g = 1, \dots, G$) est décrite par un modèle linéaire mixte :

$$Y_i(t_{ij}|c_i = g, X_{Li}, Z_i; \theta_y) = X_{Li}(t_{ij})^T \beta + Z_i(t_{ij})^T b_{ig} + \epsilon_i(t_{ij}) \quad (\text{II.12})$$

où β est le vecteur des effets fixes communs à toutes les classes et est associé au vecteur des covariables $X_{Li}(t_{ij})^T$; $b_{ig} = b_i|_{c_i=g} \sim \mathcal{N}(\mu_g, D_g)$ est le vecteur des effets aléatoires spécifiques à chaque classe et est associé au vecteur des covariables $Z_i(t_{ij})^T$; μ_g est le vecteur des paramètres des effets fixes dans la classe g . D_g représente la matrice de variance-covariance des effets aléatoires, elle peut être commune à toutes les classes ($D_g = D$) ou spécifique à chaque classe.

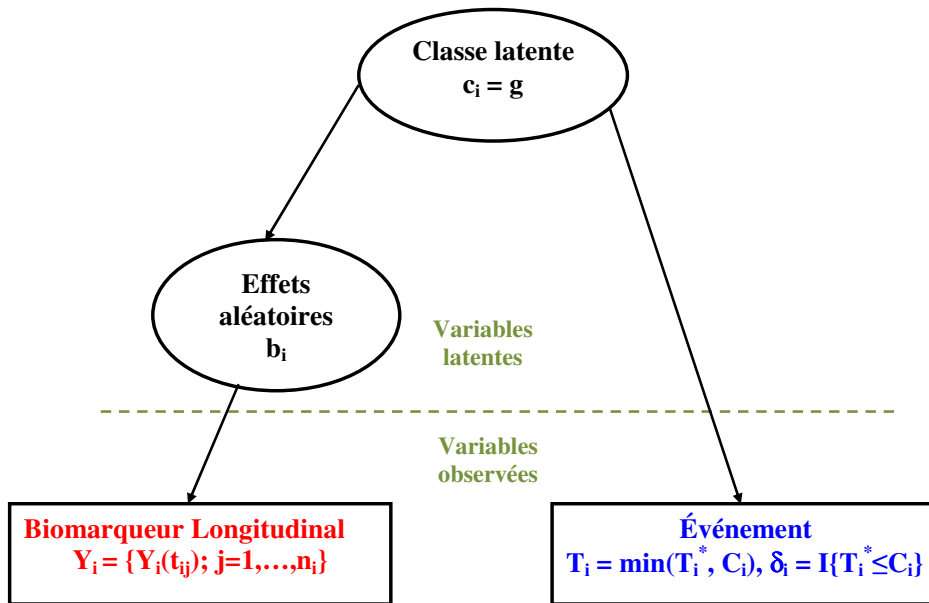


Figure II.3 – Graphe orienté du modèle conjoint à classes latentes

Une façon équivalente de décrire ce type de modèle est de définir que le vecteur des effets aléatoires b_i suit un mélange de G lois Gaussiennes multivariées, c'est-à-dire $b_i \sim \sum_{g=1}^G \pi_{ig} \mathcal{N}(\mu_g, D_g)$ où π_{ig} est la probabilité définie dans (II.11) et ϵ_i est le vecteur des erreurs de mesure comme dans un modèle mixte standard (section 1.1), i.e $\epsilon_i \sim \mathcal{N}(0, \Sigma_i)$.

Les erreurs peuvent être indépendantes, c'est-à-dire $\Sigma_i = \sigma^2 I_{n_i}$, ou corrélées avec différentes structures de corrélation. On note θ_y le vecteur des paramètres dans (II.12). Remarquons que ce modèle peut également prendre en compte des effets fixes spécifiques aux classes en remplaçant β par β_g .

Le modèle (II.12) modélise la corrélation entre les données répétées du biomarqueur à la fois à l'intérieur de chaque sous-population à travers la classe latente et au sein de chaque trajectoire individuelle à travers les effets aléatoires.

Le risque d'événement dans chaque classe g est décrit par un modèle à risques proportionnels classique (section 2) :

$$\lambda_i(t|c_i = g, X_{Ei}; \theta_e) = \lambda_{0g}(t; \zeta_g) \exp\left(X_{Ei}^T \gamma_g\right) \quad (\text{II.13})$$

où λ_{0g} est la fonction de risque de base spécifique à la classe g et décrite par le vecteur de paramètres ζ_g ; elle peut être de type Weibull, constante par morceaux ("piecewise") ou approchée par des splines comme décrit dans la section 2.3. X_{Ei} est un vecteur de covariables associé au vecteur des paramètres γ_g . Le vecteur total des paramètres du modèle de survie est $\theta_e = ((\zeta_g, \gamma_g)_{g=1, \dots, G})$.

D'autres modélisations de la survenue de l'événement ont été étudiées. [Muthén and Shedden \(1999\)](#), [Lin et al. \(2000\)](#), [Beunckens et al. \(2008\)](#) et [Proust-Lima et al. \(2007\)](#) ont utilisé un modèle de régression logistique pour définir la probabilité d'événement tandis que [Lin et al. \(2002\)](#) ont ajouté un terme de fragilité au modèle à risques proportionnels.

3.4.2 Estimation du modèle

L'estimation des modèles conjoints à classes latentes se fait généralement pour un nombre de classes fixé G . Nous notons $\theta_G = (\xi, \theta_y, \theta_e)$ le vecteur de tous les paramètres pour

un modèle à G classes. En utilisant l'hypothèse d'indépendance conditionnelle entre le processus longitudinal et le processus de survie sachant la classe latente, la log-vraisemblance du modèle s'écrit :

$$l(\theta_G) = \sum_{i=1}^N \log \left(\sum_{g=1}^G \pi_{ig} f(Y_i | c_i = g; \theta_y) \lambda_i(T_i | c_i = g; \theta_e)^{\delta_i} S_i(T_i | c_i = g; \theta_e) \right) \quad (\text{II.14})$$

où π_{ig} est la probabilité pour le sujet i d'appartenir à la classe g et est définie en (II.11); $f(Y_i | c_i = g; \theta_y)$ est la densité d'une loi normale multivariée de moyenne $X_{Li} \beta + Z_i \mu_g$ et de matrice de variance-covariance $Z_i D_g Z_i^T + \Sigma_i$ où X_{Li} et Z_i sont des matrices de vecteurs lignes $X_{Li}(t_{ij})^T$ et $Z_i(t_{ij})^T$ respectivement.

Le terme de droite $\lambda_i(T_i | c_i = g; \theta_e)^{\delta_i} S_i(T_i | c_i = g; \theta_e)$ représente la densité d'un temps d'événement pour des données censurées à droite, $\lambda_i(T_i | c_i = g; \theta_e)$ la fonction de risque instantané et $S_i(T_i | c_i = g; \theta_e)$ sa fonction de survie associée issues du modèle défini en (II.13).

Cette vraisemblance a une expression analytique et ne nécessite donc pas d'intégration numérique comme celle d'un modèle à effets aléatoires partagés. L'intégrale sur les effets aléatoires est remplacée par une somme sur les classes latentes.

L'estimation peut se faire, pour un nombre de classes fixé G , par la méthode du maximum de vraisemblance à travers l'algorithme itératif comme celui de [Marquardt \(1963\)](#) et les variances des paramètres peuvent être obtenues par l'inverse de la matrice Hessienne ([Proust-Lima and Taylor, 2009a](#); [Proust-Lima et al., 2009b](#); [Proust-Lima et al., 2012](#)). Un algorithme EM peut aussi bien être utilisé pour estimer les modèles conjoints, c'est le cas de [Lin et al. \(2002\)](#), [Lin et al. \(2004\)](#), [Beunckens et al. \(2008\)](#) et [Zhang and Wang \(2010\)](#). [Garre et al. \(2008\)](#) a proposé une approche bayésienne.

Pour déterminer le nombre optimal de classes latentes, le modèle doit également être estimé pour différentes valeurs du nombre de classes G et le modèle avec le nombre de

classes optimal est retenu. Ce choix est fait le plus souvent à partir du critère d'information bayésien (BIC) (Schwarz, 1978) qui est la vraisemblance pénalisée par le nombre de paramètres et le nombre de sujets du modèle (Hawkins et al., 2001 ; Zhang and Cheng, 2004 ; Lin et al., 2002 ; Proust-Lima and Taylor, 2009a). D'autres critères statistiques existent et dans certains cas, des critères cliniques peuvent aussi intervenir telles que l'identification d'un nombre maximum fixé de classes.

L'estimation des modèles conjoints à classes latentes peut se faire à travers la fonction Jointlcm du package R lcm (http://cran.r-project.org/web/packages/lcm/index.html).

3.4.3 Classification a posteriori

A la différence du modèle conjoint à effets aléatoires partagés, le modèle conjoint à classes latentes permet de faire de la classification de sujets. Cela se fait à partir des probabilités a posteriori d'appartenance aux classes latentes. Pour les modèles conjoints à classes latentes, deux probabilités a posteriori sont obtenues à partir de la formule de Bayes : la probabilité a posteriori sachant le biomarqueur longitudinal et le temps d'événement :

$$\hat{\pi}_{ig}^{Y,T} = \mathbb{P}(c_i = g | Y_i, (T_i, \delta_i); \hat{\theta}) = \frac{\hat{\pi}_{ig} f(Y_i | c_i = g; \theta_y) \lambda_i(T_i | c_i = g; \theta_e)^{\delta_i} S_i(T_i | c_i = g; \theta_e)}{\sum_{l=1}^G \hat{\pi}_{il} f(Y_i | c_i = l; \theta_y) \lambda_i(T_i | c_i = l; \theta_e)^{\delta_i} S_i(T_i | c_i = l; \theta_e)}$$

et la probabilité a posteriori sachant le biomarqueur uniquement :

$$\hat{\pi}_{ig}^Y = \mathbb{P}(c_i = g | Y_i; \hat{\theta}) = \frac{\hat{\pi}_{ig} f(Y_i | c_i = g; \theta_y)}{\sum_{l=1}^G \hat{\pi}_{il} f(Y_i | c_i = l; \theta_y)}$$

La probabilité $\hat{\pi}_{ig}^{Y,T}$ est utilisée pour évaluer la qualité d'ajustement du modèle conjoint tandis que $\hat{\pi}_{ig}^Y$ est plutôt utilisée pour faire de la prédiction d'événement. Dans les deux cas, les probabilités permettent de classer les sujets dans les différentes classes. Pour un sujet i , son vecteur des probabilités est donné par $\pi_i = (\pi_{i1}, \dots, \pi_{iG})$ avec $\sum_{g=1}^G \pi_{ig} = 1$.

Ce sujet est classé dans la composante g si $\hat{\pi}_{ig} = \max_{l=1, \dots, G} \{\hat{\pi}_{il}\}$ c'est-à-dire la composante à laquelle il a la plus grande probabilité d'appartenir.

3.4.4 Évaluation du modèle

Comme pour le modèle à effets aléatoires partagés, le modèle à classes latentes a besoin d'être évalué. A partir des classifications *a posteriori* des sujets présentée ci-dessus, on peut évaluer la qualité discriminante du modèle ([Muthén and Shedden, 1999](#) ; [Beunckens et al., 2008](#)). Une bonne classification correspondrait à ce que chaque sujet ait une probabilité proche de 1 d'appartenir à une classe et des probabilités d'appartenir aux autres classes proches de 0.

De même que pour les modèles conjoints à effets aléatoires partagés, les modèles conjoints à classes latentes permettent de calculer des prédictions relatives au modèle longitudinal. La différence est que ces prédictions sont spécifiques aux classes latentes : pour un sujet i , à un temps t_{ij} et dans une classe g , on peut obtenir une prédiction marginale ou spécifique à chaque sujet. A partir de ces prédictions, des prédictions individuelles peuvent être obtenues en calculant la moyenne sur toutes les classes ou des prédictions spécifiques aux classes par la moyenne des prédictions sur tous les sujets ([Proust-Lima et al., 2012](#)). Des résidus marginaux ou spécifiques aux sujets peuvent aussi être déduits des prédictions comme cela est présenté dans la section [1.3](#).

Les modèles conjoints à classes latents sont basés sur l'hypothèse d'indépendance entre les processus longitudinal et de temps d'événement conditionnellement à la classe latente. [Jacqmin-Gadda et al. \(2010\)](#) ont proposé un test pour vérifier si la dépendance n'existe pas également à travers les effets aléatoires. Il s'agit d'un test du score qui pose comme hypothèse nulle que la dépendance entre le biomarqueur longitudinal et le temps d'événement existe uniquement à travers les classes latentes et comme hypothèse alternative que la dépendance existe à travers les classes latentes et au moins un des effets aléatoires

(Figure II.4). Cela revient à avoir le modèle de survie (II.15) sous l'hypothèse alternative :

$$\lambda_i(t|c_i = g, X_{Ei}; \theta_e, \eta) = \lambda_{0g}(t; \zeta_g) \exp(X_{Ei} \gamma_g + b_{ig}^T \eta) \quad (\text{II.15})$$

et tester l'hypothèse nulle $H_0 : \eta = 0$. La statistique du score est le gradient par rapport à η de la log-vraisemblance conjointe dérivée du modèle de survie (II.15) et évaluée en $\eta = 0$. Si l'hypothèse nulle est rejetée, les classes latentes ne capturent pas toute la corrélation entre le processus longitudinal et celui du temps d'événement. On se retrouve alors en présence d'un modèle conjoint à classes latentes avec effets aléatoires. Ce type de modèles conjoints alliant des classes latentes et des effets aléatoires partagés a été proposé par Beunckens et al. (2008) et Garre et al. (2008).

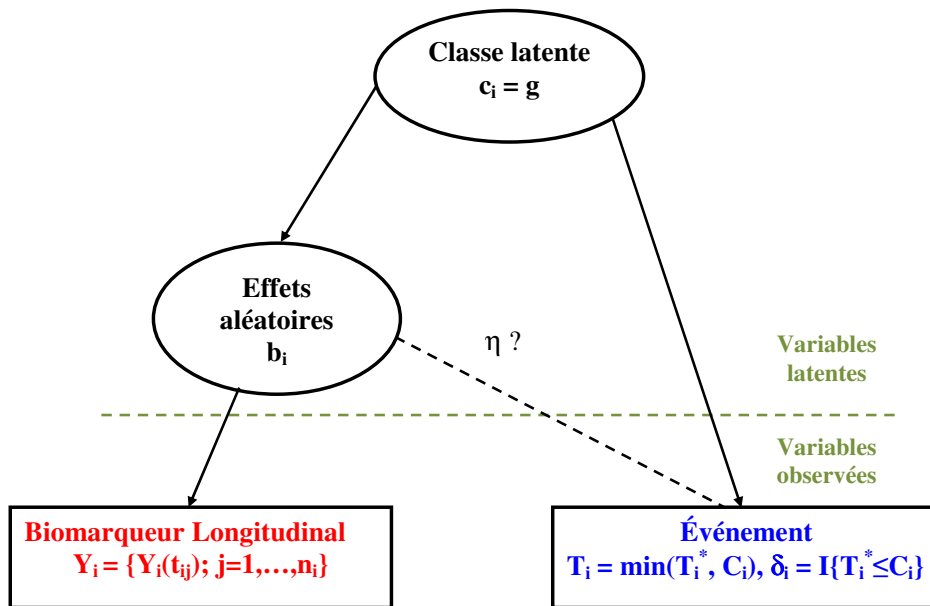


Figure II.4 – Graphe orienté du modèle conjoint avec une dépendance entre le biomarqueur longitudinal et le temps d'événement à la fois à travers les effets aléatoires et la classe latente

3.4.5 Extensions et limites du modèle

D'autres extensions du modèle conjoint à classes latentes ont été développées. Parmi celles-ci, nous pouvons citer des modèles conjoints avec des risques compétitifs (Huang

et al., 2011) ou des modèles conjoints à classes latentes avec plusieurs biomarqueurs longitudinaux (Proust-Lima et al., 2007 ; Proust-Lima et al., 2009b).

L'estimation des modèles conjoints à classes latentes pose un certain nombre de problèmes. En plus d'être calculée pour une valeur fixe de G , la vraisemblance peut avoir plusieurs modes et les algorithmes cités précédemment peuvent alors converger vers un maximum local. Afin de pallier ce problème, il est nécessaire d'estimer le modèle conjoint à classes latentes plusieurs fois à partir de différentes valeurs initiales. Les différentes vraisemblances obtenues sont comparées et le vecteur des paramètres estimés qui donne la vraisemblance maximale et donc représentant le maximum global est sélectionné (Biernacki et al., 2003 ; Berchtold, 2004).

Au contraire des modèles conjoints à effets aléatoires partagés, les modèles conjoints à classes latentes ne permettent pas de tester des formes pré-définies de relation entre un biomarqueur et un temps d'événement. De plus, dans certains cas, les classes latentes peuvent également être difficilement interprétables. Néanmoins, elles constituent un outil simple pour décrire des trajectoires dans une population sans faire aucune hypothèse sur la nature de la dépendance, ce qui peut se révéler utile lorsque l'on veut faire de la prédiction notamment.

4 Outils pronostiques dynamiques

Les outils pronostiques dynamiques consistent à utiliser toutes les informations disponibles jusqu'à un certain temps (noté s), en particulier des données répétées de biomarqueur, afin de prédire le risque que l'événement survienne entre les temps s et $s+t$ (Figure II.5). Ils fournissent aux cliniciens une quantification du risque de l'événement d'intérêt qui varie en fonction de la dynamique du biomarqueur. Ces outils sont dits "*dynamiques*" car ils peuvent être mis à jour à chaque nouvelle mesure disponible.

Ces prédictions dynamiques individuelles de l'évènement d'intérêt peuvent permettre au clinicien de suivre, à tout temps, la progression de la maladie. Cette information permet ainsi d'adapter pour chaque patient un suivi individuel adéquat et d'initier d'éventuels (nouveaux) traitements si nécessaires. C'est le principe de la médecine personnalisée (Welsh and Powis, 2009).

A cause de leur caractère dynamique, ces outils peuvent être construits à partir de modèles conjoints pour données longitudinales et de temps d'évènement. Comme pour tout outil pronostique, un outil pronostique dynamique a besoin d'être validé. Cette validation peut être complexe en présence de données censurées comme c'est souvent le cas dans les études longitudinales. Nous présentons dans cette section le calcul de ces prédictions dynamiques et leur validation.

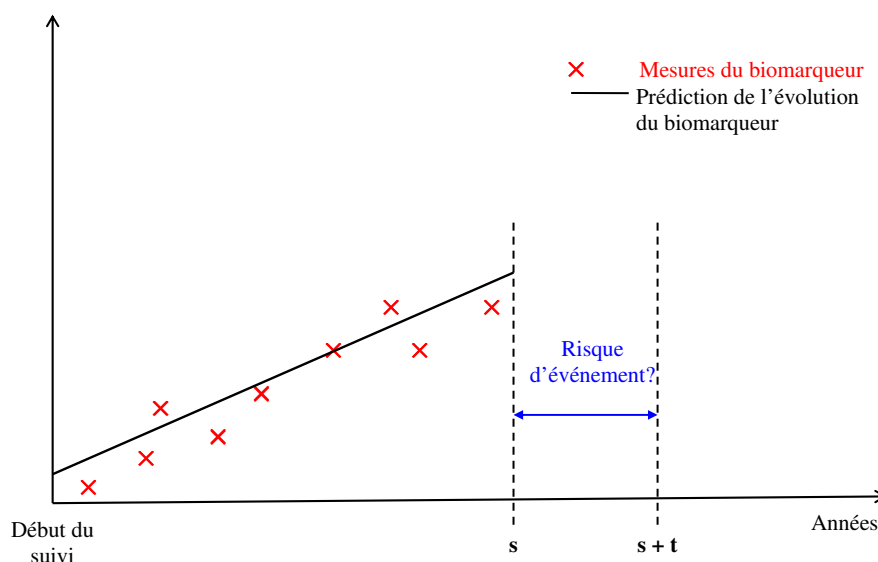


Figure II.5 – Outil pronostique dynamique

4.1 Prédictions individuelles dynamiques

Soit $s \geq 0$ un temps de prédiction et $t \geq 0$ un horizon de temps. Les prédictions dynamiques consistent en la probabilité prédite que l'évènement survienne entre les temps

s et $s + t$ sachant les covariables au diagnostic et les données du biomarqueur collectées jusqu'au temps s (Proust-Lima and Taylor, 2009a; Rizopoulos, 2011b). Pour un sujet i , T_i^* le temps d'événement, $Y_i^{(s)} = \{Y_i(t_{ij}), j = 1, \dots, n_i \text{ tel que } t_{ij} \leq s\}$ l'histoire du biomarqueur jusqu'en s , $X_{Li}^{(s)} = \{X_{Li}(t_{ij}), j = 1, \dots, n_i \text{ tel que } t_{ij} \leq s\}$ l'histoire jusqu'en s d'un vecteur de covariables dépendant du temps, X_{Si} représente le vecteur de covariables au diagnostic et θ le vecteur des paramètres du modèle.

Les prédictions individuelles dynamiques sont dérivées des modèles conjoints et, quel que soit le modèle utilisé, modèle conjoint à effets aléatoires partagés ou modèle conjoint à classes latentes, la probabilité prédite que l'événement survienne entre s et $s + t$ est donnée par :

$$p_i(s, t; \theta) = \mathbb{P}(T_i^* \leq s + t | T_i^* \geq s, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta)$$

De tels outils pronostiques dynamiques ont été introduits dans le contexte des modèles conjoints à effets aléatoires et des modèles conjoints à classes latentes (Proust-Lima and Taylor, 2009a; Rizopoulos, 2011b)

- Pour les modèles conjoints à effets aléatoires partagés, les probabilités prédites d'événement sont données par (Rizopoulos, 2011b) :

$$p_i(s, t; \theta) = \int_{b_i} \mathbb{P}(T_i^* \leq s + t | T_i^* \geq s, b_i, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta) \times f(b_i | T_i^* \geq s, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta) db_i \quad (\text{II.16})$$

- Pour les modèles conjoints à classes latentes, la formule est similaire à (II.16) sauf que l'intégrale sur les effets aléatoires est remplacée par une somme sur les classes latentes (Proust-Lima and Taylor, 2009a) :

$$p_i(s, t; \theta) = \sum_{g=1}^G \mathbb{P}(T_i^* \leq s + t | T_i^* \geq s, c_i = g, X_{Li}^{(s)}, X_{Si}; \theta) \times \mathbb{P}(c_i = g | T_i^* \geq s, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta) \quad (\text{II.17})$$

Une estimation ponctuelle de la probabilité prédite d'événement peut être donnée par $p_i(s, t; \hat{\theta})$ où $\hat{\theta}$ est l'estimateur de maximum de vraisemblance. Toutefois la méthode de Monte-Carlo peut tout aussi bien être utilisée afin d'approximer directement la distribution a posteriori de $p_i(s, t; \theta)$. Cette méthode présente l'avantage de fournir des intervalles de confiance. Elle consiste à générer un échantillon de L vecteurs de paramètres $\theta^{(l)}$ ($l = 1, \dots, L$) suivant la distribution asymptotique des paramètres estimés du modèle conjoint $\mathcal{N}(\hat{\theta}, \widehat{V}(\hat{\theta}))$ et calculer $p_i(s, t; \theta^{(l)})$. Une estimation de la probabilité d'événement est obtenue via le calcul de la médiane (ou de la moyenne) des $p_i(s, t; \theta^{(l)})$ et les quantiles d'ordre 2,5 et 97,5% donnent eux l'intervalle de confiance à 95%.

Dans ce travail de thèse, avons estimé les probabilités individuelles d'événement comme décrit ci-dessus. Cependant, il existe une approche légèrement différente proposée par [Rizopoulos \(2011b\)](#) qui utilise aussi la technique de Monte-Carlo. A partir d'un grand échantillon de vecteurs de paramètres $\theta^{(l)} \sim \mathcal{N}(\hat{\theta}, \widehat{V}(\hat{\theta}))$ ($l = 1, \dots, L$) nous calculons $p_i(s, t; \theta^{(l)})$ alors que [Rizopoulos \(2011b\)](#) génère aussi un vecteur des effets aléatoires $b_i^{(l)}$ ($l = 1, \dots, L$) suivant la distribution a posteriori des effets aléatoires $f(b_i | T_i^* \geq s, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta^{(l)})$ et les probabilités d'événement sont obtenues conditionnellement à ces effets aléatoires ($\mathbb{P}(T_i^* \leq s + t | T_i^* \geq s, b_i^{(l)}, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta^{(l)})$) et comme nous, il conserve un estimation ponctuelle avec la médiane ou la moyenne et l'intervalle de confiance à 95% est donné par les quantiles d'ordre 2,5 et 97,5%.

Pour clore cette section, on peut citer une dernière approche de prédiction dynamique, appelée prédiction dynamique par "landmark". Celle-ci consiste à estimer un modèle de survie uniquement chez les sujets encore à risque à un temps s , appelé point de repère ("landmark"), et sur toutes les mesures collectées jusqu'à ce temps. Cela implique que l'on a autant de vecteurs de paramètres estimés que de points de repère. La probabilité prédite d'événement est ainsi calculée en utilisant uniquement la valeur du biomarqueur au point de repère ([Van Houwelingen, 2007](#); [Parast et al., 2012](#)) ou des résumés de l'histoire

du biomarqueur au point de repère. Nous n'avons néanmoins pas considéré ce type de prédiction dynamique dans cette thèse.

4.2 Évaluation de l'outil

Afin de s'assurer des performances de l'outil pronostique, il est nécessaire d'évaluer sa capacité prédictive ou discriminative. Dans le modèle de survie classique, deux principales mesures sont couramment utilisées : une mesure de précision prédictive, basée sur l'erreur de prédiction qui peut être absolue ou quadratique, comparant le statut de survie observé et la survie prédite ([Graf et al., 1999](#) ; [Schemper and Henderson, 2000](#)) et une mesure de précision discriminative permettant d'évaluer comment le modèle discrimine les patients ayant un fort risque de subir l'événement de ceux ayant un faible risque ([Harrell et al., 1982](#) ; [Pencina et al., 2008](#)).

Ces deux mesures ont été adaptées pour des covariables dépendant du temps comme les biomarqueurs ([Henderson et al., 2002](#) ; [Heagerty and Zheng, 2005](#) ; [Gerds and Schumacher, 2006](#) ; [Zheng and Heagerty, 2007](#) ; [Schoop et al., 2008](#) ; [Proust-Lima and Taylor, 2009a](#) ; [Schoop et al., 2011](#) ; [Rizopoulos, 2011b](#)).

Une autre mesure de pouvoir prédictif, plus récente et issue de la théorie de l'information, peut également être envisagée : c'est l'entropie croisée pronostique ([Commenges et al., 2012](#) ; [Proust-Lima et al., 2012](#)).

Dans cette thèse, nous avons fait le choix d'évaluer les outils pronostiques par leur pouvoir prédictif et non par leur pouvoir discriminant. En effet, nous nous intéressons d'avantage à la capacité de l'outil à bien prédire l'événement via le calcul de probabilités d'événement et de leurs erreurs associées. De ce fait, nous avons uniquement utilisé l'erreur de prédiction quadratique et l'entropie croisée pour évaluer la qualité prédictive des outils pronostiques.

4.2.1 Erreur de prédiction quadratique ou score de Brier

L'erreur de prédiction est une mesure qui compare à un temps donné le statut de survie observé et la survie prédite. C'est une mesure de pouvoir prédictif qui est définie comme l'espérance d'une fonction de perte $L : E \left[L \left(\eta(u) - \hat{S}(u) \right) \right]$ où $\eta(u)$ est le statut de survie observé à un temps u valant 0 ou 1 et $\hat{S}(u)$ est la survie prédite.

L'erreur de prédiction quadratique appelée aussi score de Brier (BS) est définie par une fonction de perte L quadratique. Un estimateur du score de Brier est la moyenne de la différence au carré du statut de survie observé et de la survie prédite (Brier, 1950). Dans le contexte de la prédiction dynamique, le score de Brier à un temps $s + t$ est estimé par :

$$\text{BS}(s, t) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left(\eta_i(s + t) - \hat{S}_i(s + t|s) \right)^2 \quad (\text{II.18})$$

où N_s est le nombre de sujets encore à risque au temps de prédiction s , $\eta_i(s + t)$ est le statut de survie observé au temps $s + t$ valant 0 ou 1, $\hat{S}_i(s + t|s) = 1 - p_i(s, t; \hat{\theta})$ est la survie prédite au temps $s + t$ sachant que le sujet est encore à risque au temps de prédiction s et considérant toute l'histoire du biomarqueur jusqu'au temps s . $p_i(s, t; \hat{\theta})$ représente la probabilité prédite d'événement donnée par (II.16) ou (II.17).

Cependant, dans ce contexte de prédiction dynamique, le statut de survie au temps $s + t$ peut être censuré. En effet, au temps $s + t$, on peut rencontrer une de ces trois possibilités :

- (i) le sujet est encore à risque à $s + t$ ce qui veut dire que $T_i > s + t$ et son statut de survie est $\eta(s + t) = 1$.
- (ii) le sujet a déjà observé l'événement avant ou en $s + t$ donc $T_i \leq s + t$ et $\delta_i = 1$, ce qui veut dire que $\eta(s + t) = 0$.

(iii) le sujet est censuré avant ou en $s + t$ donc $T_i \leq s + t$ et $\delta_i = 0$. Pour ce sujet, le statut de survie est inconnu et la contribution au score de Brier ne peut pas être calculée.

S'il n'y a pas de données censurées, le score de Brier est donné par l'équation (II.18). Afin de compenser la perte d'information due à la censure, deux estimateurs du score de Brier ont été proposés.

Un premier estimateur consiste à pondérer les contributions individuelles au score de Brier (II.18) en utilisant la technique IPCW ("Inverse Probability of Censoring Weighting") (Graf et al., 1999 ; Gerds and Schumacher, 2006 ; Schoop et al., 2011 ; Proust-Lima et al., 2012). Cela revient à avoir :

$$\hat{BS}_1(s, t) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left[\frac{\mathbb{1}_{\{T_i > s+t\}}}{\hat{G}(s+t)/\hat{G}(s)} \left(1 - \hat{S}(s+t|s)\right)^2 + \frac{\delta_i \mathbb{1}_{\{T_i \leq s+t\}}}{\hat{G}(T_i)/\hat{G}(s)} \left(0 - \hat{S}(s+t|s)\right)^2 \right] \quad (\text{II.19})$$

où $\hat{G}(\cdot)$ est l'estimateur de Kaplan-Meier de la fonction de survie de la distribution de la censure (Graf et al., 1999 ; Schoop et al., 2011) ou bien estimée par un modèle de régression (Gerds and Schumacher, 2006). Dans cette estimation, les observations sont pondérées suivant leur probabilité d'être observées. Le premier terme correspond aux sujets encore à risque à $s + t$ (i) et le second terme aux sujets qui ont eu l'événement (ii).

Le cas (iii) des sujets censurés dans la fenêtre de prédiction intervient uniquement dans la pondération des observations.

Un second estimateur a été proposé et est donné par :

$$\begin{aligned} \hat{BS}_2(s, t) = & \frac{1}{N_s} \sum_{i=1}^{N_s} \left[\mathbb{1}_{\{T_i > s+t\}} \left(1 - \hat{S}(s+t|s)\right)^2 + \delta_i \mathbb{1}_{\{T_i \leq s+t\}} \left(0 - \hat{S}(s+t|s)\right)^2 \right. \\ & \left. + (1 - \delta_i) \mathbb{1}_{\{T_i \leq s+t\}} \left\{ \left(1 - \hat{S}(s+t|s)\right)^2 \frac{\hat{S}(s+t|s)}{\hat{S}(T_i|s)} + \left(0 - \hat{S}(s+t|s)\right)^2 \left(1 - \frac{\hat{S}(s+t|s)}{\hat{S}(T_i|s)}\right) \right\} \right] \end{aligned} \quad (\text{II.20})$$

Le premier terme correspond à la contribution des sujets encore à risque à $s + t$ (i), le second terme aux sujets ayant eu l'événement avant ou à $s + t$ (ii) et le troisième terme aux sujets qui ont été censurés avant ou à $s + t$ (iii) et pour qui on suppose qu'ils soient encore à risque ou qu'ils aient subi l'événement en leur affectant les poids $\hat{S}_i(s + t|s)/\hat{S}_i(T_i|s)$ et $1 - \hat{S}_i(s + t|s)/\hat{S}_i(T_i|s)$ respectivement (Schemper and Henderson, 2000; Henderson et al., 2002; Proust-Lima et al., 2012). Pour les sujets censurés, les poids qui leur sont affectés sont prédits en utilisant directement le modèle conjoint. La principale limite de cet estimateur est qu'il suppose indirectement que le modèle est bien spécifié en l'utilisant pour imputer la contribution des sujets censurés.

4.2.2 L'entropie croisée pronostique

L'entropie croisée est issue de la théorie de l'information. Pour une variable aléatoire Z de densité f , l'entropie peut être considérée comme un risque, qui est l'espérance d'une fonction de perte $-\log f(Z)$.

En prédiction dynamique, on s'intéresse à pronostiquer l'événement d'intérêt sachant que le sujet est encore à risque et sachant également toute son histoire. Dans ce contexte, l'entropie croisée pronostique (CE) est alors basée sur l'espérance de la densité conditionnelle de l'événement. A partir d'un temps de prédiction s , l'entropie croisée pronostique est définie comme : $E \left[-\log \left(f_{T|Y^{(s)}, T^* \geq s} \right) | T^* \geq s \right]$ où $f_{T|Y^{(s)}, T^* \geq s}$ est la densité conditionnelle de l'événement, dérivée du modèle conjoint, sachant l'histoire du biomarqueur jusqu'au temps s . Contrairement à d'autres mesures de qualité prédictive telles que le score de Brier, l'entropie croisée ne se base pas directement sur la probabilité conditionnelle d'événement. Elle est définie par une fonction de densité, la densité de la partie survie du modèle conjoint conditionnellement à l'histoire du biomarqueur. L'entropie croisée pronostique est estimée par :

$$\hat{\text{CE}} = \frac{1}{N_s} \sum_{i=1}^{N_s} -F_i(\hat{\theta}, s) \quad (\text{II.21})$$

où N_s est le nombre de sujets encore à risque au temps de prédiction s , $\hat{\theta}$ est le vecteur des paramètres estimés du modèle conjoint et F_i est la contribution individuelle à la log-vraisemblance conditionnelle (Commenges et al., 2012; Proust-Lima et al., 2012). Comme cela sera détaillé dans le chapitre III, la log-vraisemblance conditionnelle est définie :

- pour les modèles à effets aléatoires partagés par

$$F_i(\hat{\theta}, s) = \mathbb{1}_{\{T_i \geq s\}} \log \left(\frac{\int_{b_i} f_Y(Y_i^{(s)} | b_i, X_{L_i}^{(s)}; \hat{\theta}) \lambda_i(T_i | b_i, X_{S_i}; \hat{\theta})^{\delta_i} S(T_i | b_i, X_{S_i}; \hat{\theta}) f_b(b_i) db_i}{\int_{b_i} f_Y(Y_i^{(s)} | b_i, X_{L_i}^{(s)}; \hat{\theta}) S(s | b_i, X_{S_i}; \hat{\theta}) f_b(b_i) db_i} \right)$$

- pour les modèles à classes latentes par

$$F_i(\hat{\theta}, s) = \mathbb{1}_{\{T_i \geq s\}} \log \left(\frac{\sum_{g=1}^G \pi_{ig} f_Y(Y_i^{(s)} | c_i = g, X_{L_i}^{(s)}; \hat{\theta}) \lambda_i(T_i | c_i = g, X_{S_i}; \hat{\theta})^{\delta_i} S(T_i | c_i = g, X_{S_i}; \hat{\theta})}{\sum_{g=1}^G \pi_{ig} f_Y(Y_i^{(s)} | c_i = g, X_{L_i}^{(s)}; \hat{\theta}) S(s | c_i = g, X_{S_i}; \hat{\theta})} \right)$$

Comme elle est calculée à partir de la vraisemblance, l'entropie dépend de l'indicateur d'événement δ_i qui est connu en tout temps alors que le score de Brier dépend du statut de survie qui n'est pas toujours connu. En effet, $\delta_i = 1$ si le sujet a observé l'événement et 0 sinon. Dans le cas d'une censure, l'entropie a l'avantage par rapport au score de Brier de pouvoir être calculée sans faire d'hypothèse supplémentaire sur la distribution de la censure.

Les deux mesures de qualité prédictive présentées ci-dessus sont évaluées à chaque temps de prédiction s uniquement chez les sujets encore à risque. Pour des grandes valeurs de s , on dispose de peu d'individus mais ces sujets encore à risque disposent de plus d'informations acquises au cours de leurs suivis. Les populations sur lesquelles ces mesures sont évaluées sont donc différentes à chaque temps s , d'où la nécessité de comparer la qualité prédictive entre modèles pour chaque temps s d'intérêt.

4.2.3 Validation interne/externe

Afin d'évaluer les performances prédictives d'un modèle, il est essentiel d'utiliser d'autres données que celles utilisées lors de l'estimation du modèle. En effet, une validation de l'outil pronostique sur les données d'estimation peut engendrer un problème de *sur-optimisme*. L'outil va avoir tendance à trop bien prédire l'événement.

[Altman and Royston \(2000\)](#) évoquent l'importance d'évaluer les performances d'un outil pronostique sur des données externes. Ils distinguent deux types de validation : une validation "clinique" qui consiste à valider l'outil pronostique par des observations cliniques et une validation "statistique" qui consiste à valider l'outil pronostique en terme d'adéquation et de qualité prédictive. Dans notre cas, on se limitera uniquement à une validation statistique de l'outil pronostique.

Cependant, disposer à la fois de données d'estimation et de données de validation externe n'est pas toujours possible en pratique. Une méthode alternative est d'utiliser la technique de *validation croisée* ([Altman and Royston, 2000](#) ; [Gerds and Schumacher, 2007](#)). Elle consiste à diviser l'échantillon de données en un échantillon d'apprentissage et un échantillon de test. Le modèle est alors estimé sur l'échantillon d'apprentissage et validé sur l'échantillon de test ([Royston et al., 2004](#) ; [May et al., 2004](#)).

Pour des modèles aussi complexes que les modèles conjoints, faire de la validation croisée peut être très lourd en temps de calcul. [Commenges et al. \(2013\)](#) ont ainsi proposé un critère universel d'approximation de la validation croisée par la méthode "leave-one-out". Cela permet d'évaluer la qualité prédictive d'un modèle directement sur les données d'estimation en corrigeant le sur-optimisme, mais également de l'évaluer sur des données externes.

[Commenges et al. \(2012\)](#) ont proposé une adaptation de cette technique universelle de validation croisée approximée à l'entropie croisée pronostique.

Dans cette thèse, nous avons utilisé cette méthode de validation croisée approximée pour évaluer l'entropie croisée pronostique. Nous l'avons également adaptée pour le score de Brier.

Chapitre III

Prédiction dynamique du risque de rechute clinique de cancer de la prostate : comparaison d’approches

Dans ce chapitre nous comparons différentes approches de modélisation conjointes dans le cadre de la prédiction dynamique de risque de rechute clinique de cancer de la prostate à partir des données répétées de PSA chez des patients traités par radiothérapie externe pour un cancer de la prostate localisé. On compare, en terme d’ajustement mais surtout en terme de pouvoir prédictif, tout d’abord des modèles conjoints à effets aléatoires partagés avec plusieurs type de dépendance entre le processus longitudinal et le risque de rechute clinique, puis deux approches différentes de modélisation conjointe à savoir les modèles conjoints à effets aléatoires partagés et les modèles conjoints à classes latentes.

1 Introduction

En modélisation conjointe à effets aléatoires partagés, différents types de dépendance entre les données longitudinales du biomarqueur et le temps d’événement peuvent être considérés. Ils correspondent généralement à une dépendance directe à travers les effets aléatoires ([Jacqmin-Gadda et al., 2004](#)), ou une dépendance à travers le niveau courant du biomarqueur ([Faucett and Thomas, 1996](#) ; [Wulfsohn and Tsiatis, 1997](#)) ou à travers le niveau courant et la pente courante ([Yu et al., 2008](#)). Cependant, ce choix de fonction de dépendance est a priori rarement discuté ou justifié.

Chapitre III. Prédiction dynamique du risque de rechute clinique de cancer de la prostate : comparaison d'approches

Dans le cas où l'on veut évaluer l'hypothèse qu'un biomarqueur puisse être utilisé comme un biomarqueur de substitution pour l'événement d'intérêt, l'utilisation d'une dépendance à travers le niveau courant du biomarqueur peut être justifié. Cependant, il est important, lorsque l'on fait de la prédiction, d'effectuer le meilleur choix de modélisation possible afin de traduire au mieux la dépendance entre les deux processus.

L'objectif de ce chapitre est d'investiguer, à travers les modèles conjoints à effets aléatoires partagés, la relation entre le PSA et le risque de rechute clinique chez des patients atteints de cancer de la prostate localisé et traités uniquement par radiothérapie. Nous avons étudié plusieurs types de dépendance entre l'évolution du PSA et le risque de rechute clinique. Les différents modèles conjoints considérés ont été comparés en terme d'adéquation aux données mais également en terme de qualité prédictive dans le contexte de la prédiction dynamique. Pour évaluer la qualité prédictive des modèles, nous avons adapté l'entropie croisée pronostique, présentée pour les modèles conjoints à classes latentes par [Commenges et al. \(2012\)](#), aux modèles conjoints à effets aléatoires partagés. Ce travail a fait l'objet d'un article publié dans le Journal de la Société Française de Statistique et présenté dans la section 2.

Ce type de comparaison peut également être fait suivant différentes approches de modélisation conjointe. Dans un second article, nous avons comparé des modèles conjoints à effets aléatoires partagés et des modèles conjoints à classes latentes, toujours en termes d'ajustement et de pouvoir prédictif. Dans cette analyse, le pouvoir prédictif des modèles a été évalué à travers l'entropie croisée pronostique mais également à travers l'erreur quadratique de prédiction (score de Brier) sur des données externes. Ma part de travail dans cet article a concerné l'implémentation et l'analyse des modèles conjoints à effets aléatoires partagés uniquement. Ce travail qui a fait l'objet d'un article publié dans le journal "Statistical Methods in Medical Research" est résumé dans la section 3 et présenté en annexe.

2 Article

Résumé de l'article

Au cours de ces dix dernières années, la recherche en modélisation conjointe s'est développée de manière accrue dans le domaine des biostatistiques et de la recherche médicale. Ce type de modélisation permet d'étudier simultanément la relation entre un biomarqueur longitudinal et un temps d'événement. Parmi les différents modèles appartenant à cette famille de modèles conjoints, les modèles à effets aléatoires partagés ont notamment reçu le plus d'attention. Ils sont construits à l'aide d'un modèle mixte pour le biomarqueur longitudinal et d'un modèle de survie pour le temps d'événement. Ils ont la particularité d'inclure les caractéristiques du modèle mixte comme variables explicatives. Ces modèles étendent naturellement le cadre des modèles de survie incluant des variables explicatives dépendant du temps et offrent ainsi une flexibilité dans l'exploration du lien entre un biomarqueur longitudinal et un risque d'événement.

Après une brève revue de la méthodologie relative aux modèles à effets aléatoires partagés, cet article a pour objectif de détailler son implémentation et son évaluation à travers un exemple réel d'étude de progression du cancer de la prostate chez des patients ayant été traités par radiothérapie. Pour ce faire, nous investiguons différentes spécifications de dépendance entre le biomarqueur longitudinal, l'antigène spécifique de la prostate (PSA), et le risque de rechute clinique afin de mieux comprendre le lien existant entre ces deux processus. Ces différents modèles conjoints sont comparés en terme de qualité d'ajustement et d'adéquation aux hypothèses du modèle conjoint mais également en termes de pouvoir prédictif évalué à l'aide de l'entropie croisée pronostique. En effet, outre une meilleure compréhension de la relation entre l'évolution d'un biomarqueur et la survenue d'une rechute clinique, une des perspectives dans les études sur le cancer, et en particulier le cancer de la prostate, est de fournir des outils pronostiques dynamiques de rechute clinique basés sur l'histoire du biomarqueur longitudinal.

Shared random-effect models for the joint
analysis of longitudinal and time-to-event
data : application to the prediction of
prostate cancer recurrence

Mbéry Sène^{1,2,3*}, Carine A. Bellera^{3,4} and Cécile Proust-Lima^{1,2,3}

¹ INSERM, ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, F-33000
Bordeaux, France

² Univ. Bordeaux, ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, F-33000
Bordeaux, France

³ INSERM, ISPED, CIC-EC7, F-33000 Bordeaux, France

⁴ Department of Clinical Epidemiology and Clinical Research, Institut Bergonié,
Regional Comprehensive Cancer Center, Bordeaux, France.

Correspondence to : Mbéry Sène, INSERM U897, ISPED, Université de Bordeaux Se-
galen, 146 rue Léo Saignat, 33076 Bordeaux Cedex, FRANCE

Mbery.Sene@isped.u-bordeaux2.fr

Abstract : In the last decade, joint modeling research has expanded very rapidly in biostatistics and medical research. This type of models enables the simultaneous study of a longitudinal marker and a correlated time-to-event. Among them, the shared random-effect models that define a mixed model for the longitudinal marker and a survival model for the time-to-event including characteristics of the mixed model as covariates received the main interest. Indeed, they extend naturally the survival model with time-dependent covariates and offer a flexible framework to explore the link between a longitudinal biomarker and a risk of event.

The objective of this paper is to briefly review the shared random-effect model methodology and detail its implementation and evaluation through a real example from the study of prostate cancer progression after a radiation therapy. In particular, different specifications of the dependency between the longitudinal biomarker, the prostate-specific antigen (PSA), and the risk of clinical recurrence are investigated to better understand the link between the PSA dynamics and the risk of clinical recurrence. These different joint models are compared in terms of goodness-of-fit and adequation to the joint model assumptions but also in terms of predictive accuracy using the expected prognostic cross-entropy. Indeed, in addition to better understand the link between the PSA dynamics and the risk of clinical recurrence, the perspective in prostate cancer studies is to provide dynamic prognostic tools of clinical recurrence based on the biomarker history.

Keywords : Joint models, Shared random-effect models, Dynamic predictions, Prognostic cross-entropy, Predictive accuracy, Prostate cancer.

Résumé : Dans la dernière décennie, la recherche en modélisation conjointe s'est développée très rapidement dans le domaine des biostatistiques et de la recherche médicale. Ce type de modèles permet d'étudier simultanément un marqueur longitudinal et un temps d'événement corrélés. Parmi eux, les modèles à effets aléatoires partagés, qui définissent un modèle mixte pour le marqueur longitudinal et un modèle de survie pour le temps d'événement incluant les caractéristiques du modèle mixte comme variables explicatives, ont reçu le plus d'attention. En effet, ces modèles étendent naturellement le modèle de survie avec variables explicatives dépendantes du temps et offrent un cadre flexible pour explorer le lien entre le biomarqueur longitudinal et le risque d'événement.

L'objectif de cet article est de passer brièvement en revue la méthodologie du modèle à effets aléatoires partagés et de détailler son implémentation et son évaluation à travers un exemple réel d'étude de progression de cancer de la prostate après une radiothérapie. En particulier, différentes spécifications de la dépendance entre le biomarqueur longitudinal, l'antigène spécifique de la prostate (PSA), et le risque de rechute clinique sont investiguées pour bien comprendre le lien entre la dynamique du PSA et le risque de rechute clinique. Ces différents modèles conjoints sont comparés en termes de qualité d'ajustement et d'adéquation aux hypothèses du modèle conjoint mais aussi en termes de pouvoir prédictif en utilisant la cross-entropie pronostique. En effet, en plus de mieux comprendre le lien entre la dynamique de PSA et le risque de rechute clinique, la perspective dans les études sur le cancer de la prostate est de fournir des outils pronostiques dynamiques de rechute clinique basés sur toute l'histoire du biomarqueur.

Mots clés : Modèles conjoints, Modèles à effets aléatoires partagés, Prédictions dynamiques, Cross-entropie pronostique, Pouvoir prédictif, Cancer de la prostate.

2.1 Introduction

In cohort studies, time-to-event data and repeated measures of biomarkers data are often collected for each subject. There exist standard methods for separately analyzing these data; mixed models (Laird and Ware, 1982) are used to describe the biomarker repeated measures and assess the association with covariates while survival models, mainly proportional hazard models (Cox, 1972), are used to evaluate the risk of the event and the effect of its predictors. These approaches are however not optimal. First, these processes are usually dependent so that separated analyses may lead to biased inference, and second, interest is frequently in the understanding of their association so that the two processes need to be modeled simultaneously.

To assess the relationship between a longitudinal biomarker and a time-to-event, an intuitive model would be the proportional hazard Cox model (Cox, 1972) with the biomarker as a standard time-dependent covariate. But this technique is not adapted with biomarker repeated data. Indeed, first the Cox model assumes that the time-dependent covariate is external (or exogenous) which means that the value of the covariate at time t is not affected by the occurrence of an event at time u , with $t > u$ (Kalbfleisch and Prentice, 2002). This is not the case with biomarkers that constitute a typical example of internal (or endogenous) time-dependent covariate. Second, the Cox model assumes that the time-dependent covariate is observed at all the event times, which is most always not the case in practice. To solve this problem, missing values at the event times can be imputed using the LOCF (Last Observation Carried Forward) method that replaces an unobserved value at an event time by the most recent observed value. However, this method can strongly bias the estimates (Tsiatis and Davidian, 2004), especially with irregularly and infrequent measures of the biomarker. Finally, the standard Cox model assumes that the time-dependent covariate is measured without any error where as the observed biomarker is noisy measure.

To remedy this, joint models have been developed. They aim at modeling simultaneously the longitudinal biomarker and time-to-event processes, and at characterizing their relationship. As a consequence, they provide an interesting framework to (i) evaluate the longitudinal biomarker trajectory and its association with covariates without the bias introduced by the informative time-to-event, (ii) evaluate the risk of event and its association with covariates, including the longitudinal biomarker data and (iii) directly explore the association between the longitudinal and survival processes. The principle of joint models is to model the biomarker repeated measures using a mixed model, to model the risk of event using a survival model and to link the two models using a common latent structure (Faucett and Thomas, 1996 ; Wulfsohn and Tsiatis, 1997 ; Henderson et al., 2000). This common latent structure captures the association between the processes so that the two processes are conditionally independent given the latent structure.

There exist two kinds of joint models for longitudinal and time-to-event data : the shared random-effect models (SREM) and the joint latent class models (JLCM). The SREM extends directly the idea of the survival model with time-dependent covariates by considering as covariates some characteristics of the mixed model defined for the longitudinal biomarker (Faucett and Thomas, 1996 ; Wulfsohn and Tsiatis, 1997 ; Henderson et al., 2000 ; Tsiatis and Davidian, 2004 ; Jacqmin-Gadda et al., 2004 ; Ibrahim et al., 2010 ; Sweeting and Thompson, 2011 ; Wu et al., 2012). These characteristics are functions of the individual random-effects of the mixed model that capture the individual deviations to the mean trajectory of the biomarker. The most classical example of shared characteristic is the individual current error-free level of the biomarker predicted from the mixed model that is included as a time-dependent covariate in the survival model (Faucett and Thomas, 1996 ; Wulfsohn and Tsiatis, 1997). Another example is to include in the survival model directly the vector of random-effects. The alternative, the JLCM, relies on a different idea. It assumes that the population is heterogeneous and can be divided in a finite number of homogeneous subgroups (or classes), each class being characterized by a specific trajectory

of the biomarker and a specific risk of event (Lin et al., 2002; Proust-Lima and Taylor, 2009a; Proust-Lima et al., 2012). The latent class structure can be seen as a latent stratification that entirely captures the dependency between the two processes. See Proust-Lima et al. (2012) for a review of JLCM.

In the last decade, joint modeling research has expanded very rapidly in biostatistics and medical research with a preference for the SREM approach. Indeed, it extends naturally the survival model with time-dependent covariates and offers a flexible framework to explore the link between a longitudinal biomarker and a risk of event. Besides, it allows to improve the efficiency in assessment of treatment effects and other prognostic factors (Ibrahim et al., 2010). Different extensions of these models have been developed; with multiple events (Huang et al., 2011), or multiple longitudinal data (Rizopoulos and Ghosh, 2011a), or multivariate longitudinal and multivariate survival data (Chi and Ibrahim, 2006), or with cured fraction (Law et al., 2002; Yu et al., 2004; Yu et al., 2008), or in presence of data clustering with common frailty (Ratcliffe et al., 2004), or with a proportional cumulative hazard time-to-event submodel instead of standard proportional hazard (Crowther et al., 2012).

In addition to evaluating the association between a longitudinal and event process, joint models (JLCM and SREM) have also allowed very recently the development of a new type of prediction tools that incorporate the repeated biomarker measures to predict the risk of event (Proust-Lima and Taylor, 2009a; Rizopoulos, 2011b; Proust-Lima et al., 2012). Because of the longitudinal nature of the biomarker, these predictive tools are dynamic : whatever the time s , they provide the risk of event from s given all the information collected until s so that, in practice, the dynamic prediction can be systematically updated when new information is collected. As a consequence, these dynamic predictive tools potentially provide more accurate predictions than standard prognostic tools based only on information at baseline. Dynamic prognostic tools are of particular interest in chronic

diseases such as cancers for predicting a clinical event (recurrence, death,...) where in addition to standard prognostic factors, repeated measures of a biomarker are collected for monitoring the patients (Yu et al., 2008 ; Proust-Lima and Taylor, 2009a).

In this context, the aim of this paper is to review briefly the SREM methodology and detail its implementation and evaluation through a real example from the study of prostate cancer progression after a radiation therapy. In particular, different specifications of the dependency between the longitudinal biomarker, the prostate-specific antigen (PSA), and the risk of clinical recurrence are investigated to better understand the link between the PSA dynamics and the risk of clinical recurrence. These different joint models are compared in terms of goodness-of-fit and adequation to the joint model assumptions but also in terms of predictive accuracy. Indeed, in addition to better understand the link between the PSA dynamics and the risk of event, the perspective in prostate cancer studies is to provide powerful dynamic prognostic tools of clinical recurrence based on the biomarker history.

The paper is organized as follows. Section 2 presents the SREM model and its estimation. Section 3 focuses on dynamic prognostic tools derived from these models and their validation using a measure of predictive accuracy. The method is applied to prostate cancer progression in section 4. Finally, section 5 concludes.

2.2 Shared random-effect models

For subject i , $i = 1, \dots, N$, we note T_i^* the time to the event of interest and C_i the censoring time. We observe the time $T_i = \min(T_i^*, C_i)$ and $E_i = \mathbf{1}_{\{T_i^* \leq C_i\}}$, the indicator of event. We also observe $Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{in_i}))$, the n_i -vector of repeated measures of the biomarker collected intermittently at times $(t_{i1}, \dots, t_{in_i})$ until the observed time T_i . In general, covariates are denoted by X . For clarity, we distinguish in the following X_{Li} the matrix of (possibly time-dependent) covariates involved in the longitudinal submodel and X_{Si} the vector of time-independent covariates involved in the survival submodel.

2.2.1 Longitudinal submodel

Repeated measures of the biomarker are analyzed by a linear mixed model ([Laird and Ware, 1982](#)). Specifically, we assume that the repeated measures $Y_i(t_{ij})$ are noisy measures of $Y_i^*(t_{ij})$ the true unobserved biomarker value for $j = 1, \dots, n_i$. We model the mean change over time of $Y_i^*(t_{ij})$ by taking into account the correlation within the biomarker repeated measures of a same subject :

$$\begin{aligned} Y_i(t_{ij}) &= Y_i^*(t_{ij}) + \epsilon_i(t_{ij}) \\ &= X_{Li}(t_{ij})^T \beta + Z_i(t_{ij})^T b_i + \epsilon_i(t_{ij}) \end{aligned} \quad (\text{III.1})$$

where $X_{Li}(t_{ij})$ and $Z_i(t_{ij})$ are p-vector and q-vector of time-dependent covariates associated respectively with the p-vector of fixed effects β and the q-vector of Gaussian random effects b_i with mean $\mathbf{0}$ and variance-covariance matrix B . In the following, we will also use X_{Li} and Z_i the design matrices with respectively $X_{Li}(t_{ij})^T$ and $Z_i(t_{ij})^T$ as row vectors j ($j = 1, \dots, n_i$). In (III.1), the fixed part $X_{Li} \beta$ represents the mean trajectory of the biomarker over time while the random part $Z_i b_i$ defines the individual deviation relative to the mean trajectory. The vector of errors of measurements is $\epsilon_i = (\epsilon_i(t_{i1}), \dots, \epsilon_i(t_{in_i}))^T$. We assume ϵ_i are independent and follow a multivariate Gaussian distribution of mean $\mathbf{0}$ and diagonal variance-covariance matrix $\Sigma_i = \sigma^2 I_{n_i}$; ϵ_i and b_i are independent. In addition to the independent errors, autocorrelated errors could be included to further take into account the correlation between the repeated measures for a same subject. For example, [Henderson et al. \(2000\)](#) considered a stationary Gaussian process with mean zero while [Wang and Taylor \(2001\)](#) considered an integrated Ornstein-Uhlenbeck stochastic process. Originally, [Tsiatis et al. \(1995\)](#), [Faucett and Thomas \(1996\)](#) and [Wulfsohn and Tsiatis \(1997\)](#) used a linear mixed model with only a random intercept and a random slope but any functions of time can be considered in $X_{Li}(t_{ij})$ and $Z_i(t_{ij})$ to capture at best the trajectory of the biomarker.

2.2.2 Survival submodel

The risk of event could be modelled using any survival model but in practice, proportional hazard models are mostly considered and defined as follows :

$$\lambda_i(t|X_{Si}, b_i) = \lambda_0(t) e^{X_{Si}^T \gamma + h(b_i, t)^T \eta} \quad (\text{III.2})$$

where $\lambda_0(t)$ is the baseline hazard function, γ is the r-vector of coefficients defining the association between the r-vector of covariates X_{Si} (that can be time dependent) and the survival time, and $h(b_i, t)$ is a multivariate function of the random-effects b_i defined in (III.1) and associated with the vector of parameters η . The coefficients η measure the association between the longitudinal and survival processes while $h(b_i, t)$ defines the nature of the dependence between the two processes. Different functions $h(b_i, t)$ can be defined. Usual examples include the current true level of the biomarker ($h(b_i, t) = Y_i^*(t)$) (Wulfsohn and Tsiatis, 1997; Ibrahim et al., 2010; Crowther et al., 2012), the current slope of the biomarker ($h(b_i, t) = \partial Y_i^*(t)/\partial t$), both ($h(b_i, t) = (Y_i^*(t), \partial Y_i^*(t)/\partial t)^T$) (Yu et al., 2008) or directly the individual deviations ($h(b_i, t) = b_i$) (Jacqmin-Gadda et al., 2004). In the application, we will go back on these functions and detail more specific ones in the context of prostate cancer progression.

Other types of dependency can be assumed. For example, instead of considering strictly random-effects shared between the two submodels, Henderson et al. (2000) considered two correlated zero-mean Gaussian processes, one for each submodel with the assumption that the correlation between the longitudinal and the survival processes were entirely captured by the correlation between the two zero-mean Gaussian processes.

In the present paper, we chose to not detail further this option and concentrate on models where the dependency between the submodels is captured by standard random-effects and functions of them.

2.2.3 Maximum likelihood estimation

Shared random-effect models can be estimated within the maximum likelihood framework by considering the joint likelihood from the two submodels.

Let θ be the whole vector of parameters defined in (III.1) and (III.2). Using the assumption of independence between the longitudinal and the survival processes conditionally to the random effects, the joint log-likelihood of the observed data is :

$$\begin{aligned}
 l(\theta) &= \log \left[\prod_{i=1}^N \left(\int_{b_i} f_Y(Y_i|X_{Li}, b_i; \theta) f_T(T_i|X_{Si}, b_i; \theta) f_b(b_i; \theta) db_i \right) \right] \\
 &= \sum_{i=1}^N \log \left(\int_{b_i} f_Y(Y_i|X_{Li}, b_i; \theta) \lambda_i(T_i|X_{Si}, b_i; \theta)^{E_i} S_i(T_i|X_{Si}, b_i; \theta) f_b(b_i; \theta) db_i \right)
 \end{aligned}
 \tag{III.3}$$

where f_b and f_Y are multivariate Gaussian density functions of b and Y with respectively mean 0 and $X_{Li} \beta + Z_i b_i$, and variance-covariance matrix B and Σ_i ; $\lambda_i(T_i|X_{Si}, b_i; \theta)$ is the hazard function defined in (III.2) and taken at the observed time T_i and $S_i(T_i|X_{Si}, b_i; \theta) = e^{-\int_0^{T_i} \lambda_i(t|X_{Si}, b_i; \theta) dt}$ is the derived survival function.

From this, the maximum likelihood estimates can be obtained by iterative algorithms such as EM or Newton-Raphson algorithm (Wulfsohn and Tsiatis, 1997; Henderson et al., 2000; Rizopoulos et al., 2009; Rizopoulos, 2010). This estimator has good asymptotic properties as shown by Zeng and Cai (2005). We note that other authors used a bayesian approach to estimate these joint models (Faucett and Thomas, 1996; Wang and Taylor, 2001; Guo and Carlin, 2004; Ibrahim et al., 2001).

Equation (III.3) involves two integrals that do not have analytic solutions. They are usually approximated by numerical integration with Gauss-Hermite and Gauss-Kronrod quadratures for the integrals over respectively the random effects and the survival function (Wulfsohn and Tsiatis, 1997; Henderson et al., 2000; Rizopoulos, 2010). The numerical

approximations of the integrals, mostly the Gauss-Hermite one for the random effects, slow the calculations. Indeed, the integral over the random-effects is usually multidimensional (size q). When q remains small (less than 3), the Gauss-Hermite quadrature remains the standard method but in higher dimension settings, alternative methods may be preferred to reduce the computational time like the Laplace method as proposed by Rizopoulos et al. (2009) or a Monte Carlo method. We note that the structure of B does not intervene in the computational complexity, only the number q of random-effects is limiting.

To improve the accuracy and reduce the number of nodes in Gaussian quadratures, adaptive versions have been proposed. They consist in centering and rescaling the integral (Lesaffre and Spiessens, 2001) around its modal value, so that the nodes are systematically placed at the optimum position. This technique remains however time consuming as it requires a subject-and-iteration-specific optimization to define the optimum position. To retain the same precision but simplify the numerical aspect, Rizopoulos (2012) proposed a pseudo-adaptive version where the integral is centered and rescaled according to the posterior distribution of the random-effects defined in the linear mixed model in (III.1) but estimated separately and once for all in a first step.

There exist two R packages to estimate SREM : JM (Rizopoulos, 2010) and `joineR` (<http://cran.r-project.org/web/packages/joineR>). In this work, we used directly the JM program online, or modified the JM source code when necessary to include specific functions $h(b_i, t)$. We used a Newton-Raphson algorithm for the maximization of the likelihood and a pseudo-adaptive Gaussian quadrature rule with nine points to approximate the integral over the random effects in (III.3) (Rizopoulos, 2012). The variance-covariance matrix of the parameters $\widehat{V}(\hat{\theta})$ was estimated by the inverse of the Hessian matrix.

2.3 Dynamic predictions

2.3.1 Individual prediction

Individual dynamic predictions of event can be derived from any joint model. They consist in the individual predicted probability of event between times s and $s + t$ given the biomarker data $Y_i^{(s)} = \{Y_i(t_{ij}), j = 1, \dots, n_i, \text{ such as } t_{ij} \leq s\}$ collected until time s and the matrix of explanatory covariates $X_{Li}^{(s)} = \{X_{Li}(t_{ij}), j = 1, \dots, n_i, \text{ such as } t_{ij} \leq s\}$ and X_{Si} (Proust-Lima and Taylor, 2009a; Rizopoulos, 2011b; Proust-Lima et al., 2012). In a SREM, this reduces to :

$$\begin{aligned}
 p_i(s, t; \theta) &= \mathbb{P} \left(T_i \leq s + t | T_i \geq s, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta \right) \\
 &= \frac{\int_{b_i} f_{Y^{(s)}} \left(Y_i^{(s)} | X_{Li}^{(s)}, b_i; \theta \right) [S_i(s | X_{Si}, b_i; \theta) - S_i(s + t | X_{Si}, b_i; \theta)] f_b(b_i; \theta) db_i}{\int_{b_i} f_{Y^{(s)}} \left(Y_i^{(s)} | X_{Li}^{(s)}, b_i; \theta \right) S_i(s | X_{Si}, b_i; \theta) f_b(b_i; \theta) db_i}
 \end{aligned}
 \tag{III.4}$$

where f_b and S_i are defined in section 2.2.3, and $f_{Y^{(s)}}$ is the multivariate Gaussian density function with mean $X_{Li}^{(s)} \beta + Z_i^{(s)} b_i$ and variance-covariance matrix $\Sigma_i^{(s)}$, superscript (s) meaning that only information before time s is considered.

2.3.2 Dynamic prognostic tool

Once a joint model is estimated, dynamic prognostic tools can be constructed from its estimates $\hat{\theta}$ and the variance of its estimates $\widehat{V}(\hat{\theta})$. At a time s of prediction, for a new subject i with biomarker history $Y_i^{(s)}$ and covariates $X_{Li}^{(s)}$ and X_{Si} , a dynamic prognostic tool can be computed as the predicted probability of event defined in (III.4) and computed at the point estimates $\hat{\theta}$. However, as this dynamic prognostic tool is an estimate, its uncertainty should be taken into account. This can be done by approximating

the posterior distribution of $p_i(s, t; \theta)$ defined in (III.4) using a Monte-Carlo method. It consists in generating a large set of $\theta^{(d)}$ ($d = 1, \dots, D$) from the asymptotic distribution of the estimates $\mathcal{N}\left(\hat{\theta}, \widehat{\mathbb{V}}(\hat{\theta})\right)$, and compute $p_i(s, t; \theta^{(d)})$. The median of $p_i(s, t; \theta^{(d)})$ gives a point estimate over the D draws while the 2.5% and 97.5% percentiles give the 95% confidence bands (Proust-Lima and Taylor, 2009a; Proust-Lima et al., 2012). Using also a Monte-Carlo method to generate a large set of $\theta^{(d)}$ ($d = 1, \dots, D$), Rizopoulos (2011b) proposed to generate random-effects $b_i^{(d)}$ from their posterior distribution and compute the predicted probability of event given $b_i^{(d)}$ instead of computing directly $p_i(s, t; \theta^{(d)})$.

2.3.3 Evaluation of the predictive accuracy

To evaluate the predictive accuracy of dynamic prognostic tools, different measures can be used. The quadratic error of prediction called the Brier score and developed in survival models (Gerds and Schumacher, 2006; Gerds and Schumacher, 2007) was extended to the dynamic setting of joint models (Henderson et al., 2002; Proust-Lima and Taylor, 2009a; Proust-Lima et al., 2012; Schoop et al., 2011). It consists in the expectation of the squared difference between the observed status of event and the predicted survival. The methodology of the area under the ROC curve (AUC) was also extended to the longitudinal setting (Zheng and Heagerty, 2007; Rizopoulos, 2011b) to assess how a longitudinal biomarker discriminated between patients with low and high risks of event. Recently, another predictive accuracy measure was developed from the information theory. This is the expected prognostic observed cross-entropy (EPOCE) (Commenges et al., 2012; Proust-Lima et al., 2012) which is defined as $E\left[-\log\left(f_{T|Y^{(s)}, T^* \geq s}(T)\right) | T^* \geq s\right]$, that is the expectation of the log of the conditional density $f_{T|Y^{(s)}, T^* \geq s}$ of the time of event given the history of the marker until the time of prediction s .

One major problem when evaluating the predictive performances of a model is the expected over-optimism of the performances when estimated on the data on which the

model was estimated. One way to counter that is to validate the model on external data but this is not always possible in practice. An alternative is to compute the measure on the estimating data but using a cross-validation technique (Gerds and Schumacher, 2007). However this is numerically too intensive with joint models. In contrast to the AUC or Brier score, a formula approximating the leave-one-out cross-validation was proposed by Commenges et al. (2012) for the EPOCE estimator so that this estimator provides a correct measure of predictive accuracy that can be directly computed on the estimating data in a very short computation time.

In this context, we chose to focus in this work only on the EPOCE to evaluate the predictive accuracy of the joint models on the estimating data by correcting the over-optimism. The approximated cross-validated estimator of the EPOCE called CVPOL_a is defined by :

$$\text{CVPOL}_a(s) = -\frac{1}{N_s} \sum_{i=1}^{N_s} F_i(\hat{\theta}, s) + N \text{Trace}(H^{-1}K_s) \quad (\text{III.5})$$

where N_s is the number of subjects still at risk at time s , $F_i(\hat{\theta}, s)$ is the individual contribution to the conditional log-likelihood and is defined below in (III.6), H is the Hessian matrix of the joint log-likelihood defined in (III.3), and $K_s = \frac{1}{N_s(N-1)} \sum_{i=1}^N \mathbf{1}_{\{T_i \geq s\}} \hat{v}_i(s) \hat{d}_i^T$ with $\hat{v}_i(s)$ and \hat{d}_i^T the gradients of the individual contributions respectively to the conditional log-likelihood computed in s using only $Y_i^{(s)}$ and defined in (III.6), and the joint log-likelihood defined in (III.3) and computed using all measurements of Y_i of subject i . These gradients are computed numerically using finite differences. In (III.5), the individual contribution to the conditional log-likelihood of the time-to-event given that the subject is still at risk at time s and given his biomarker history until s is :

$$F_i(\hat{\theta}, s) = \mathbf{1}_{\{T_i \geq s\}} \log \left(\frac{\int_{b_i} f_Y(Y_i^{(s)} | X_{L_i}^{(s)}, b_i; \theta) \lambda(T_i | X_{S_i}, b_i; \theta)^{E_i} S_i(T_i | X_{S_i}, b_i; \theta) f_b(b_i; \theta) db_i}{\int_{b_i} f_Y(Y_i^{(s)} | X_{L_i}^{(s)}, b_i; \theta) S_i(s | X_{S_i}, b_i; \theta) f_b(b_i; \theta) db_i} \right) \quad (\text{III.6})$$

The $CVPOL_a$ has two components. The first part $\left(-\frac{1}{N_s} \sum_{i=1}^{N_s} F_i(\hat{\theta}, s)\right)$ estimates the apparent cross-entropy in the sample while the second term $(N \text{Trace}(H^{-1}K_s))$ provides the correction for over-optimism computed by the approximated leave-one-out cross-validation. As a measure of risk, the lower the $CVPOL_a$, the better the predictive accuracy of the model. Difference in EPOCE can also be computed with a 95% tracking interval ([Commenges et al., 2012](#)) making possible the comparison of predictive accuracy between two models.

We note that for this work, the predictive accuracy was assessed in the entire remaining window of time from the time of prediction s but the EPOCE estimate could also be computed to evaluate the predictive accuracy of a model in a shorter window of horizons in order to focus on the ability to predict the event in the next τ years.

2.4 Application to prostate cancer data

The objective of this application was to better understand the dependency between the dynamics of the prostate specific antigen (PSA) and the risk of clinical recurrence after a first treatment by radiation therapy for a localized prostate cancer, and to explore the predictiveness of joint models based on the PSA dynamics.

Through this application, we also aimed at showing how shared random-effect models could be used in practice, how their goodness-of-fit could be evaluated, and how dynamic predictive tools could be derived and validated.

2.4.1 Description of the dataset

The data come from an hospital cohort of the University of Michigan that included patients treated with external beam radiation therapy (EBRT) for a localized prostate cancer and followed-up from 1988 to 2004. The present sample is constituted of 459 men among whom 74 (16.1%) had a clinical recurrence defined as any distant metastases, nodal recurrence or any palpable or biopsy-detected local recurrence three years or later after radiation, any local recurrence within three years of EBRT if the last PSA value was over 2ng/ml and death from prostate cancer. PSA measurements were collected from the end of EBRT to clinical recurrence or lost to follow-up. Patients included in the sample were required to have at least 1-year follow-up without clinical recurrence, did not initiate any second treatment during the follow-up and had at least two PSA measurements before the end of follow-up.

Three main prognostic factors known at diagnosis were considered in the analysis and included in the vector of time-independent covariates X_i :

1. The level of PSA at diagnosis was defined as a continuous covariate in the log scale ; The median was 2.1 (in log of ng/ml) with an interquartile range of [1.7, 2.6] (in log of ng/ml)
2. T-stage indicates how far the cancer has developed ; it ranges from 1 when the cancer is small and contained within the prostate, to 4 when the cancer has spread to other organs, or bones. In the application, T-stage was split in 2 categories, 3-4 *versus* 1-2, with respectively 8.9% and 91.1% of the patients in each category.
3. The Gleason score quantifies the aggressiveness of prostate cancer ; it ranges from 2 to 10, 10 indicating the higher aggressiveness. In the application, Gleason score was considered in 3 categories, 7 and 8-10 *versus* 2-6, with respectively, 37.7%, 7.4% and 54.9% of the patients.

Chapitre III. Prédiction dynamique du risque de rechute clinique de cancer de la prostate : comparaison d'approches

Patients had a median follow-up time of 5.2 years with an interquartile range of [2.7, 7.7] years and a median time to clinical recurrence of 2.8 years with an interquartile range of [1.9, 4.3] years after EBRT.

PSA repeated measures were analyzed in the logarithmic scale $\log(PSA + 0.1)$ to satisfy the normality assumption of the linear mixed model. In the following, the term biomarker and log PSA will refer to $\log(PSA + 0.1)$. Figure III.1 presents the individual observed trajectories of log PSA according to the occurrence of a clinical recurrence during the follow-up. In the two subgroups, the log PSA trajectory is characterized by a very rapid PSA drop in the first year after EBRT and a long-term increase. However, the intensity of the long-term increase differs greatly according to the subgroup : patients who recurred tend to have a faster PSA rise than patients who did not.

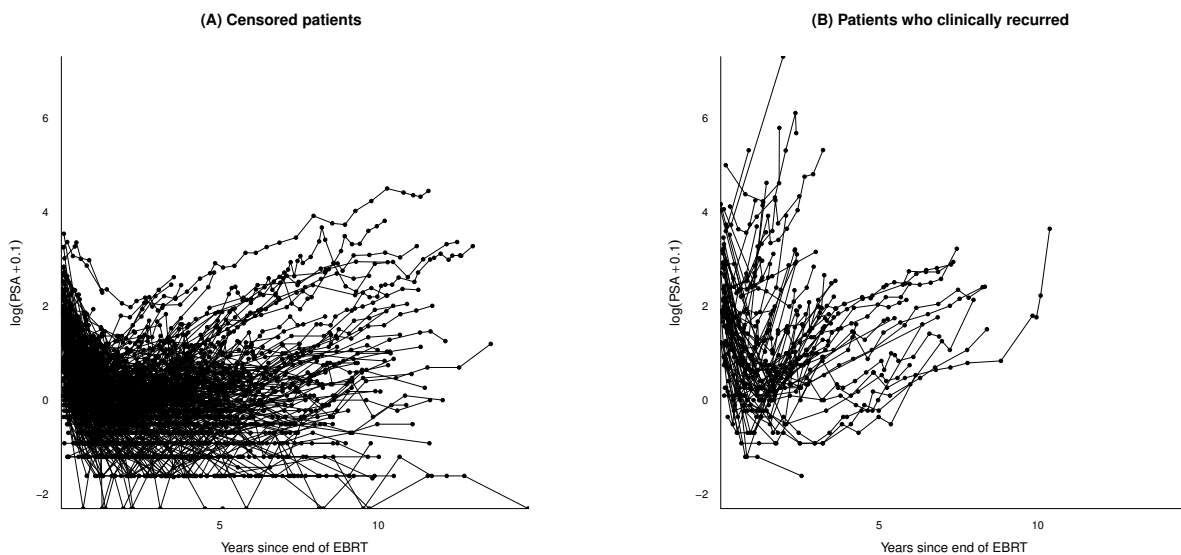


Figure III.1 – Individual observed trajectories of $\log(PSA + 0.1)$ over time for (A) censored patients and (B) patients who clinically recurred.

2.4.2 Specification of the joint models

To describe the dynamics of the PSA after EBRT and its link with time to clinical recurrence, we explored different SREM models that only differed by the function $h(b_i, t)$ assumed to summarize the dependency between the PSA dynamics and the risk of clinical recurrence.

The log PSA trajectory was systematically described by the following linear mixed model :

$$\begin{aligned}
 Y_i(t) &= \log(PSA + 0.1) \\
 &= Y_i^*(t) + \epsilon_i(t) \\
 &= \left(\mu_0 + X_i^T \beta_0 + b_{0i}\right) + \left(\mu_1 + X_i^T \beta_1 + b_{1i}\right) f_1(t) + \left(\mu_2 + X_i^T \beta_2 + b_{2i}\right) t + \epsilon_i(t), \forall t \in \mathbb{R}^+
 \end{aligned}
 \tag{III.7}$$

The biphasic trajectory of log PSA was captured by $f_1(t) = ((1+t)^{-1.5} - 1)$ that modeled the initial decline of log PSA after the end of EBRT and t that modeled the linear long-term rise in log PSA. [Proust-Lima et al. \(2008\)](#) showed through a pooled analysis of five large cohorts, by profile maximum likelihood technique, that these functions correctly captured the nonlinear biphasic PSA trajectory after EBRT. The vector $\mu = (\mu_0, \mu_1, \mu_2)$ represents the intercepts in the different phases; the vector of random effects $bi = (b_{i0}, b_{i1}, b_{i2})$ is multivariate Gaussian with vector of mean $\mathbf{0}$ and unstructured variance-covariance matrix. We remind that the vector X_i includes the three baseline prognostic factors.

The time-to-recurrence was systematically modeled using the following proportional hazard model :

$$\lambda_i(t|X_i, b_i) = \lambda_0(t) e^{X_i^T \gamma + h(b_i, t)^T \eta}
 \tag{III.8}$$

in which the log-baseline hazard, $\log(\lambda_0(t))$ was approximated using B-splines with three internal knots placed at the quartiles of the observed event times.

We explored the nature of the dependency between the log PSA dynamics described in (III.7) and the risk of clinical recurrence described by considering the following functions for $h(b_i, t)$ in (III.8) :

(a) Current level of log PSA : $h_a(b_i, t) = Y_i^*(t)$.

This is the most classical function assumed in a SREM. It assumes that the current true level of the biomarker is predictive of the risk of event.

(b) Current slope of log PSA : $h_b(b_i, t) = \partial Y_i^*(t)/\partial t$.

This model assumes that the current change of the biomarker is predictive of the risk of event.

(c) Current level and current slope of log PSA : $h_c(b_i, t) = (Y_i^*(t), \partial Y_i^*(t)/\partial t)^T$.

With $h_c(b_i, t)$, the risk of event is linked both to the current true level of the biomarker and the current change in the biomarker.

(d) Lagged level and lagged slope of log PSA : $h_d(b_i, t) = (Y_i^*(t-1), \partial Y_i^*(t-1)/\partial t)^T$.

Instead of considering that the current biomarker information explains the risk of event, a lag can be introduced as in $h_d(b_i, t)$ where we assumed a latency of 1 year between the log PSA characteristics and their impact on the risk of event.

(e) Transformed current level of log PSA : $h_e(b_i, t) = \Gamma(Y_i^*(t))$.

A SREM with $h_a(b_i, t) = Y_i^*(t)$ makes a log-linear assumption between the current true level and the risk of event. To relax this assumption, a transformation of the current level can be considered with function Γ . In the PSA progression after EBRT, [Proust-Lima et al. \(2008\)](#) found that the transformation $\Gamma(Y_i^*(t)) = \text{logit}^{-1}((Y_i^*(t) - 0.71)/0.44)$ better modeled the effect of the current level of PSA and the risk of recurrence.

(f) Transformed current level and current slope of PSA : $h_f(b_i, t) = (\Gamma(Y_i^*(t)), \partial Y_i^*(t)/\partial t)^T$.

To relax both models (c) and (e), the most adapted transformed current true level of log PSA and the current slope of log PSA can be simultaneously considered to describe the impact of PSA dynamics on the risk of recurrence.

(g) Random effects : $h_g(b_i, t) = (b_{0i}, b_{1i}, b_{2i})^T$.

Instead of considering some functions of the level of the biomarker, the individual deviations can be directly included in the survival model to explain the impact of the PSA dynamics. In that case, the function $h_g(b_i)$ linking both processes is time-independent.

We only considered the seven functions above but other functions $h(b_i, t)$ could be investigated according to the question of research.

2.4.3 Joint models estimations

Estimation of these seven models ((a) - (g)) is summarized in Table III.1 with goodness-of-fit statistics and estimates of parameters η that summarize the link between the PSA dynamics and the risk of recurrence adjusted for other prognostic factors. Whatever the assumed dependency between the PSA dynamics and the risk of recurrence, the seven models concluded to a significant association between the two processes. However, according to the Akaike criterion, the goodness-of-fit of the models varied substantially.

Models (a) and (b) underlined that the current true level of log PSA and the current slope of log PSA were associated with the risk of recurrence, higher current level or slope predicting higher risk of recurrence. Model (c) showed that these two features of PSA dynamics were independent predictors of the risk of recurrence. Their effects were partly reduced when considered simultaneously as predictors of the risk of recurrence but they both remained highly significant ($p < 0.0001$ and $p < 0.0001$).

Model (d) illustrates how some lag in the effect of a time-dependent covariate can be introduced in joint models, with the exploration here of a latency of 1 year between the characteristics of the PSA trajectory and their impact on the risk of recurrence. In the present example, the AIC was better when assuming an association between the risk of recurrence and the level and slope of log PSA at the current time (model (c)) rather than

Chapitre III. Prédiction dynamique du risque de rechute clinique de cancer de la prostate : comparaison d'approches

Table III.1 – Goodness-of-fit statistics and parameter estimates (standard-deviation) of the association between PSA dynamics and risk of recurrence adjusted for other prognostic factors from joint models (a) to (g).

Model	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Dependency	$Y_i^*(t)$	$\frac{\partial Y_i^*(t)}{\partial t}$	$\left(\frac{Y_i^*(t)}{\frac{\partial Y_i^*(t)}{\partial t}} \right)$	$\left(\frac{Y_i^*(t-1)}{\frac{\partial Y_i^*(t-1)}{\partial t}} \right)$	$\Gamma(Y_i^*(t))$	$\left(\frac{\Gamma(Y_i^*(t))}{\frac{\partial Y_i^*(t)}{\partial t}} \right)$	b_i
L	-2644.15	-2635.97	-2628.55	-2637.62	-2611.06	-2605.57	-2630.70
AIC	5356.29	5339.93	5327.09	5345.25	5290.11	5281.14	5333.41
# parameters	34	34	35	35	34	35	36
Level [†]	0.95 (0.08)	-	0.49 (0.12)	0.44 (0.15)	7.34 (0.78)	5.45 (0.84)	-
Slope [‡]	-	3.14 (0.35)	2.04 (0.39)	2.19 (0.37)	-	1.08 (0.32)	-
Random effects :							
b_{0i} (Intercept)	-	-	-	-	-	-	1.03 (0.30)
b_{1i} ($f_1(t)$)	-	-	-	-	-	-	-0.32 (0.12)
b_{2i} (t)	-	-	-	-	-	-	4.05 (0.55)

[†] Corresponds to the effect of the current true log PSA level in models (a) and (c); corresponds to the effect of the true log PSA level one year before in model (d), and corresponds to the effect of the transformed true log PSA level in models (e) and (f).

[‡] Corresponds to the effect of the current slope of log PSA in models (b), (c) and (f); corresponds to the effect of the slope one year before in model (d).

one year before (model (d)). However, the parameter estimates were roughly the same in the two models.

Models (a), (c) and (d) assume a log-linear effect of the level of log PSA on the risk of recurrence, which means that the effect of 1 additional unit of log PSA is constant whatever the level of log PSA. As suggested by previous works (Proust-Lima et al., 2008), a non-linear effect of the current level of log PSA was explored via a logistic transformation in model (e) and combined with a joint effect of the current slope in model (f). From the seven models investigated, the latter model assuming a dependency via the transformed current log PSA level and the current slope provided the best fit with an improvement of 46 points of AIC compared to the model assuming a dependency via the crude current level of log PSA in addition to the slope. The logistic transformation makes the effect of the level of PSA increase in the range 0 to 4 ng/mL, and become maximal around 4ng/mL. This result is in accordance with clinical assumptions regarding the range of meaningful

PSA levels. Let note that when considering the logistic transformation of the level of log PSA (model (f)) in the survival model instead of the crude true log-PSA level (model (c)), the association between the current slope and the risk of recurrence was still significant but its intensity was divided by two.

Finally, instead of assuming a dependency of the risk of recurrence on the PSA dynamics via the transformation and/or derivative of the current level of log PSA, the last model (g) included the individual deviations of the PSA dynamics, that are the random-effects, as predictors of the risk of recurrence in the survival model. Although the model used 1 additional parameter to summarize the dependency, it gave a worse fit than the models including the current slope of log PSA and the transformed (or crude) current level of log PSA with an AIC higher by 52.3 points compared to model (f) (and respectively 6.3 points compared to model (c)). Interestingly, as a linear long-term trend was assumed for the log PSA trajectory in (III.7), $\partial Y_i^*(t)/\partial t$ reduces approximately to $(\mu_2 + X_i^T \beta_2 + b_{2i})$ after a few years, and effects of $\partial Y_i^*(t)/\partial t$ and b_{2i} adjusted for X_i in the survival model can be quantitatively compared. So with an effect estimated around 4, model (g) gives a larger impact of the long-term slope on the risk of recurrence than models (c) and (f) with estimates of the effect of the long-term slopes around 2 and 1 respectively.

2.4.4 Goodness-of-fit assessment

In joint models, like in all statistical models, model assumptions can be checked with a comparison of the predicted *versus* the observed values and an inspection of the residuals.

Figure III.2 compares respectively the marginal $(X_{Li}(t_{ij})^T \hat{\beta})$ and the subject-specific $(X_{Li}(t_{ij})^T \hat{\beta} + Z_i(t_{ij})^T \hat{b}_i)$ with \hat{b}_i the empirical Bayes estimate) predictions from the linear mixed model with the observed log PSA values. Observations and predictions are averaged by intervals of times having the same amount of PSA measures.

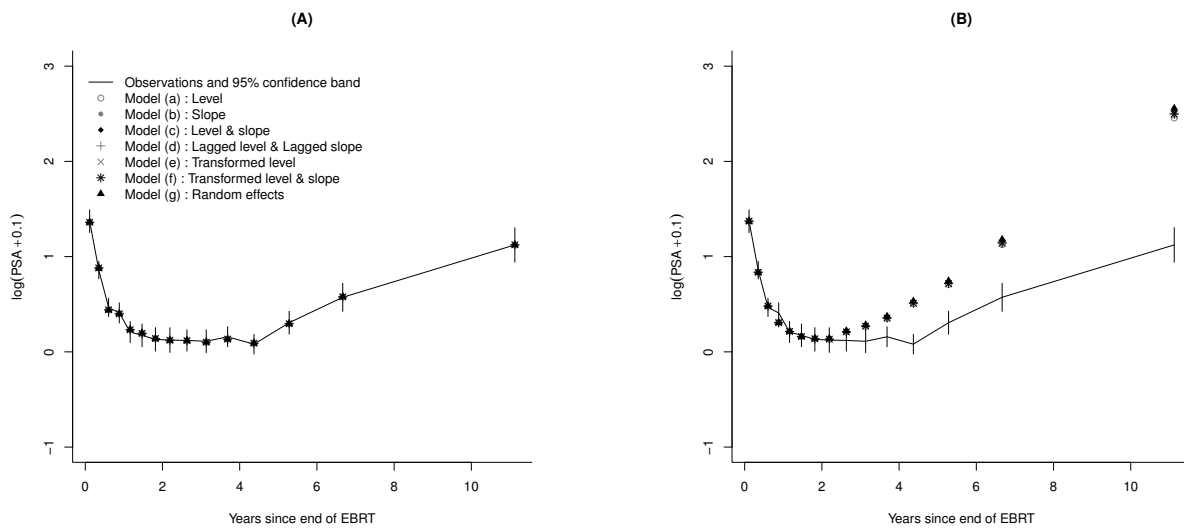


Figure III.2 – Comparison of the subject-specific predictions (A) and the marginal predictions (B) with the observed PSA values averaged by intervals of times having the same amount of PSA measures.

First, whatever the assumed dependency between processes, the longitudinal predictions are very similar between models. This was expected since the structure of the longitudinal model was the same in all the models and all the models took into account (even if differently) the association with the risk of recurrence. Second, the marginal predictions deviate from the observed mean trajectory from 2-3 years after the end of EBRT but the subject-specific predictions show a very good fit to the observed data. This deviation of the marginal predictions is due to the informative dropout caused by the clinical recurrence and was again expected. In the seven models, the residuals of the longitudinal submodel were very similar and did not exhibit any heteroscedasticity problem or model misspecification (results not shown).

For the goodness-of-fit of the survival submodel, we only focused on the four joint models (b), (c), (f) and (g). Figure III.3 displays the predicted survival curves and the Kaplan-Meier estimate of the time-to-recurrence. The predicted survival curves were computed as $\frac{1}{N} \sum_{i=1}^N S_i(t|\hat{b}_i)$ where \hat{b}_i is the empirical Bayes estimate of b_i .

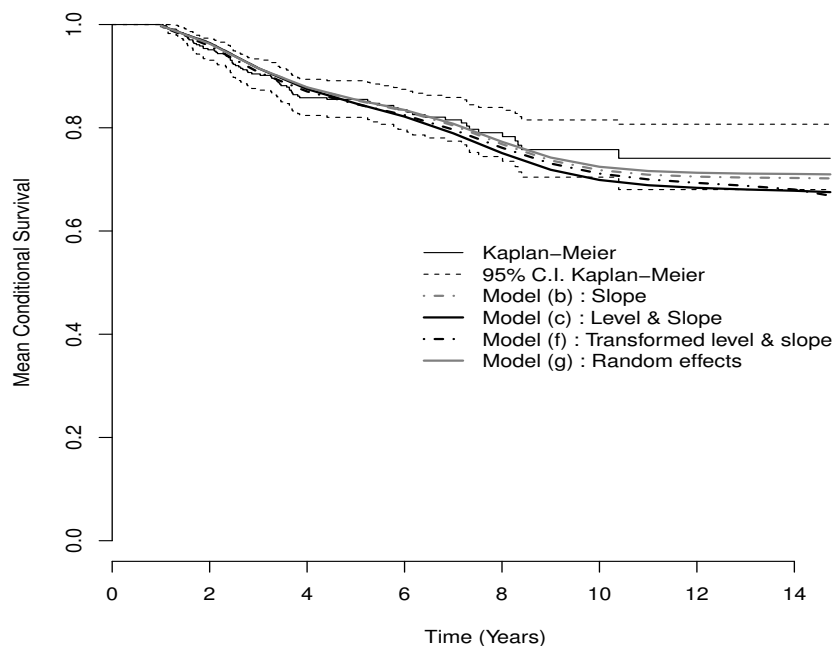


Figure III.3 – Mean predicted subject-specific survival curves from four joint models and observed Kaplan-Meier survival curve with 95% confidence bands.

All the predicted survival curves remained in the 95% confidence interval of the Kaplan-Meier estimate showing an overall correct goodness-of-fit even if the curves slightly differed one from the other.

We used also the Cox-Snell residuals computed as $\left(\Lambda_i(T_i|\hat{b}_i) = \int_0^{T_i} \lambda_i(t|\hat{b}_i) dt\right)$ to evaluate the overall goodness-of-fit of the survival models. Figure III.4 displays the Kaplan-Meier estimate of the Cox-Snell residuals from the 4 models. They are relatively close to the theoretical survival curve of a unit exponential random variable ($\exp(-t)$) showing the overall satisfying goodness-of-fit of the four survival models.

Using the Martingale residuals computed as $\left(E_i - \Lambda_i(T_i|\hat{b}_i)\right)$, we specifically assessed in Figure III.5 the log-linearity assumption of the covariates linking the PSA dynamics and the risk of recurrence. In model (c) with the current level and slope, the residuals are satisfying according to the slope of log PSA but exhibit a deviation in the larger values of the current level of log PSA. When assuming instead a transformation for the current

log PSA level in model (f), this deviation disappears, showing a very good adequation to the log-linearity assumption with the logistic transformation of the current level of log PSA and the current slope. In model (b) assuming a dependency via the slope only, the residuals also highlight a deviation in the large values of the predicted slope. The same phenomenon is observed in model (g) where the dependency is directly on the individual random-effects. While the residuals are correct for the random intercept and the random short-term drop, a deviation from the zero-mean is observed for larger random long-term slopes.

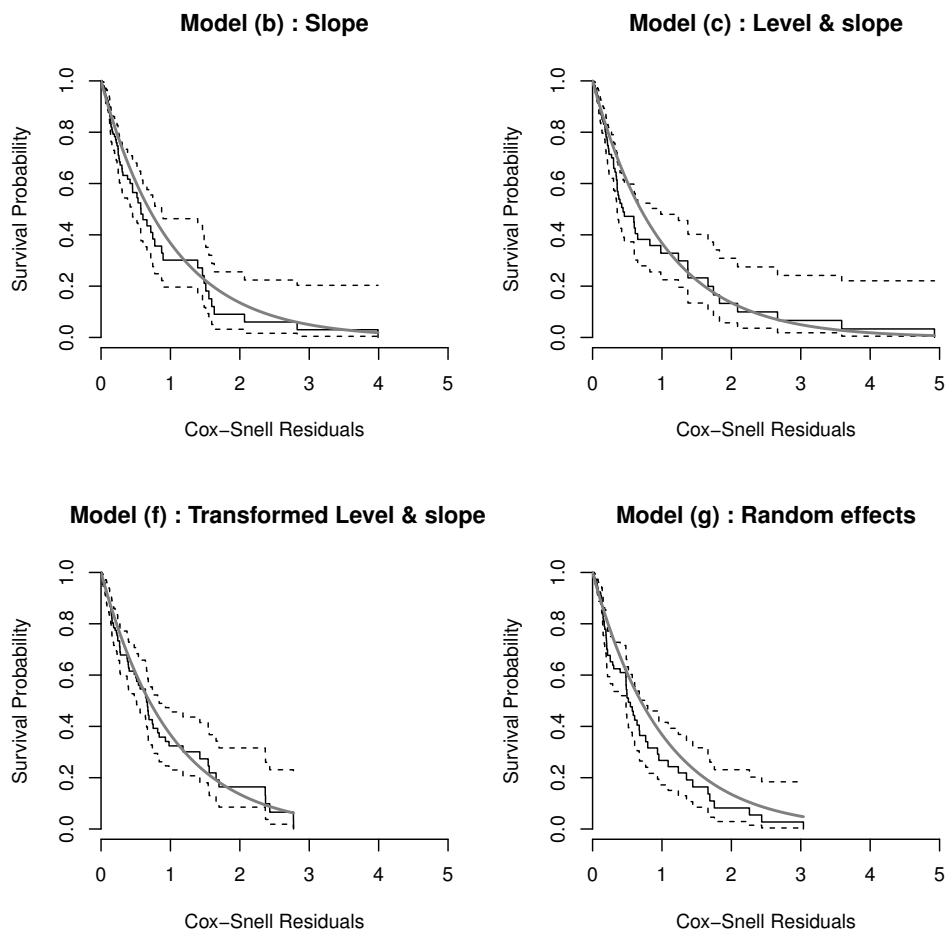


Figure III.4 – Kaplan-Meier estimates of the Cox-Snell residuals from four joint models ((b), (c), (f) and (g)) in plain black line, their 95% confidence bands in dashed black line, and the theoretical survival curve of a unit exponential random variable in plain grey line.

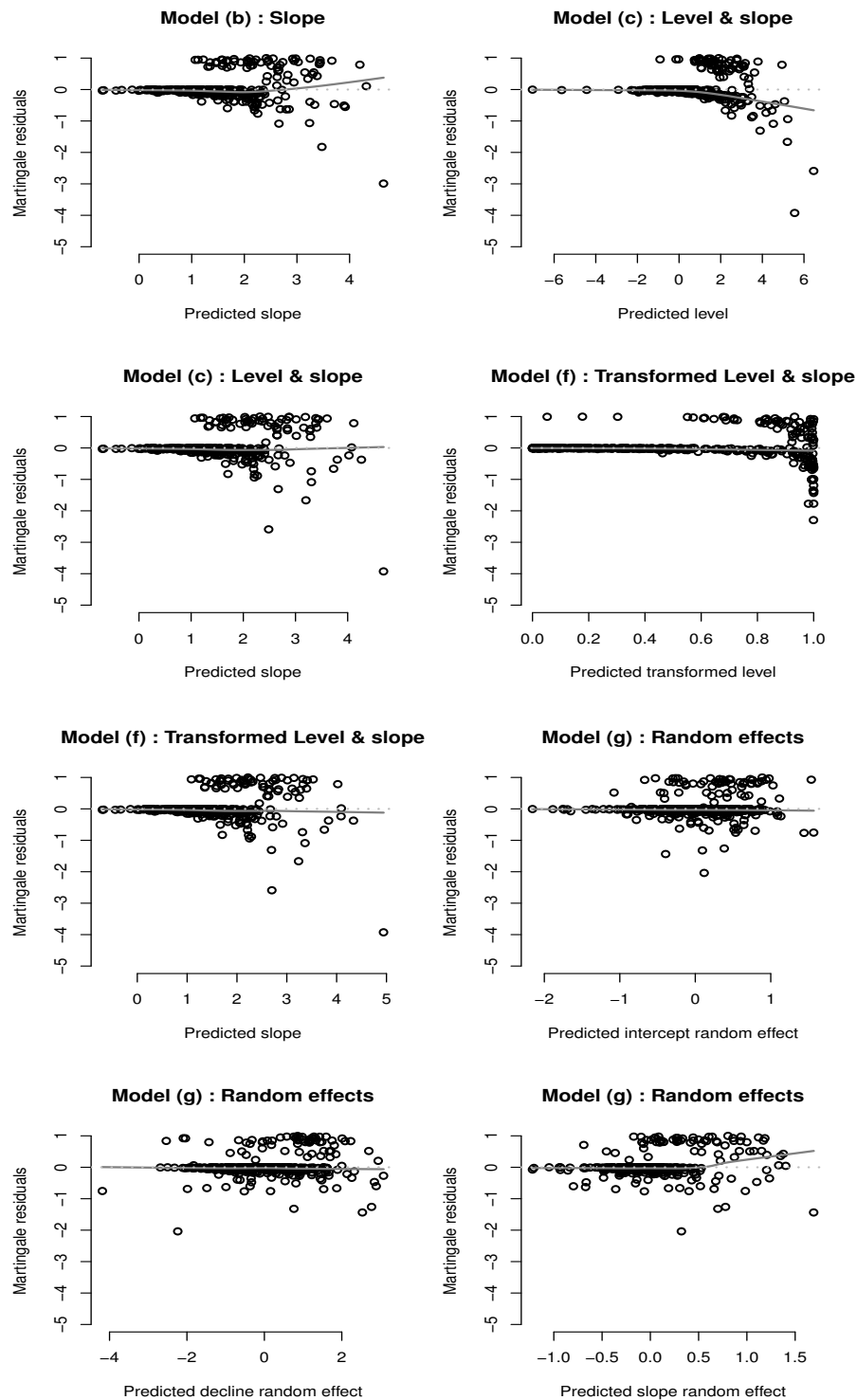


Figure III.5 – Martingale residuals *versus* predicted covariates summarizing the dependency between the PSA dynamics and the risk of recurrence in four joint models ((c), (f), (b) and (g)).

2.4.5 Predictive accuracy assessment

To evaluate the predictiveness of these joint models, we computed the approximated cross-validated prognostic likelihoods that estimate the EPOCE in Figure III.6 (A), and the differences between models with 95% tracking intervals in Figure III.6 (B). The abscissa of the graphs represents the times s of prediction so that each point gives the model ability to predict the risk of event from time s (in Figure III.6 (A)) and the difference in model abilities (in Figure III.6 (B)) for subjects still at risk at time s according to the PSA repeated measures collected before s . We remind that the EPOCE is a risk so that the lower the EPOCE estimate, the better the predictive accuracy of the model.

Interestingly, the model with the best goodness-of-fit, that is model (f), does not have the best predictive accuracy for all the times of prediction between 1 year and 6 years after the end of EBRT. Its predictive accuracy is the best (lowest EPOCE estimate) when using information in the first 2.25 years after the end of EBRT on patients still at risk at these times. However, after 2.25 years (and especially after 3 years), its predictive accuracy becomes a lot worse compared to the ones of models (b), (c) and (g).

Compared to the goodness-of-fit measures, the predictive accuracy measures focus only on subjects still at risk at the time of prediction. Yet, in our dataset, 50% of the events occurred before 2.8 years and 75% occurred before 4 years. This means that most of the information available is concentrated in the first years after the end of EBRT, giving more importance to that period of time in the estimation process compared to longer times after EBRT. As a consequence, models that gave a very good fit to the data in the first years as model (f) had better overall goodness-of-fit. This is confirmed by the predicted survival curves in Figure III.3 which show a slightly better fit of model (f) in the first years after the end of EBRT (and a slightly better fit of model (g) in longer times after end of EBRT).

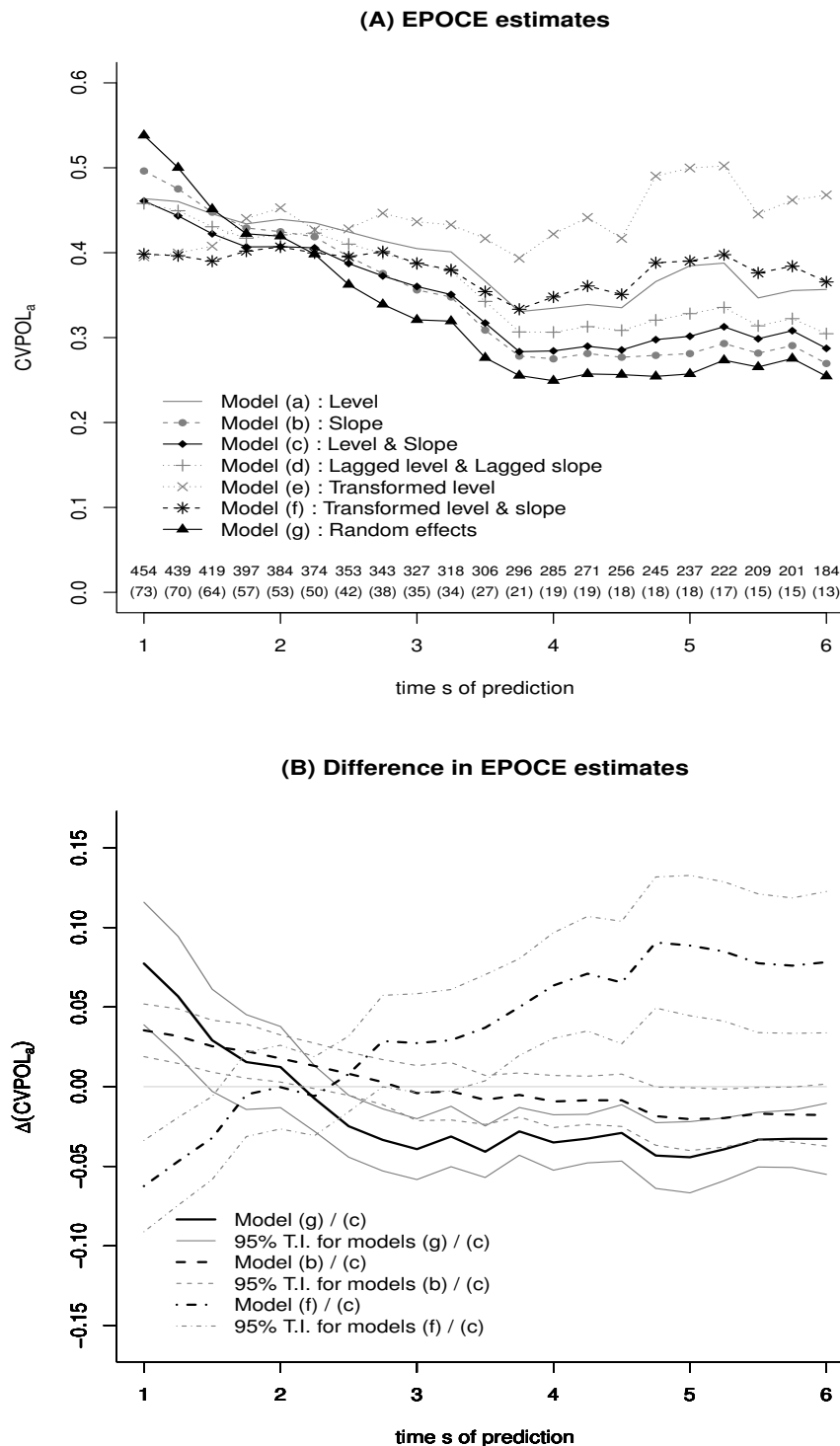


Figure III.6 – (A) $CVPOL_a$ computed from the 7 shared random-effect models from times of prediction from 1 to 6 years after the end of radiation therapy. For each time of prediction, the number of subjects still at risk (and the number of clinical recurrence) are indicated. (B) Differences in EPOCE estimate with 95% tracking interval (T.I.) between 4 joint models from times of prediction from 1 to 6 years after the end of radiation therapy.

Finally, we note that in the longer times of prediction, the models having the smallest EPOCE were those assuming a dependency through the random-effects or through the current log PSA slope. This suggests that after a few years, the slope of log PSA could be the major predictor of the risk of recurrence on subjects still at risk of recurrence. Indeed, although most of the models included the slope as a predictor, the intensity of the effect of the current slope was a lot more important in model (g) with the random-effects (parameter estimate around 4) and in model (b) with only the current slope (parameter estimate around 3) compared to models (c) and (f) including the current log PSA level also (and where the parameter estimates of the effect of the current slope were around 2 and 1).

2.4.6 Individual predictions

To illustrate how joint models can be used to provide dynamic individual predictions using the methodology described in section 2.3.2, Figure III.7 displays dynamic individual predictions of the risk of clinical recurrence computed at an horizon of 3 years and updated every 6 months from the joint model (f). The patient had a T-stage of 2, a Gleason score of 7, an initial PSA of 9.7ng/mL and had a clinical recurrence after 3.8 years following the end of EBRT. The figure illustrates how the dynamic predictive tool can be updated at each new measurement of the biomarker, with here a progressive increase in the predicted risk of recurrence over years.

2.5 Conclusion

This paper aimed at explaining through an application on a real dataset how shared random-effect models could be used in practice, how their assumptions could be evaluated, and how individual predictions could be derived from them.

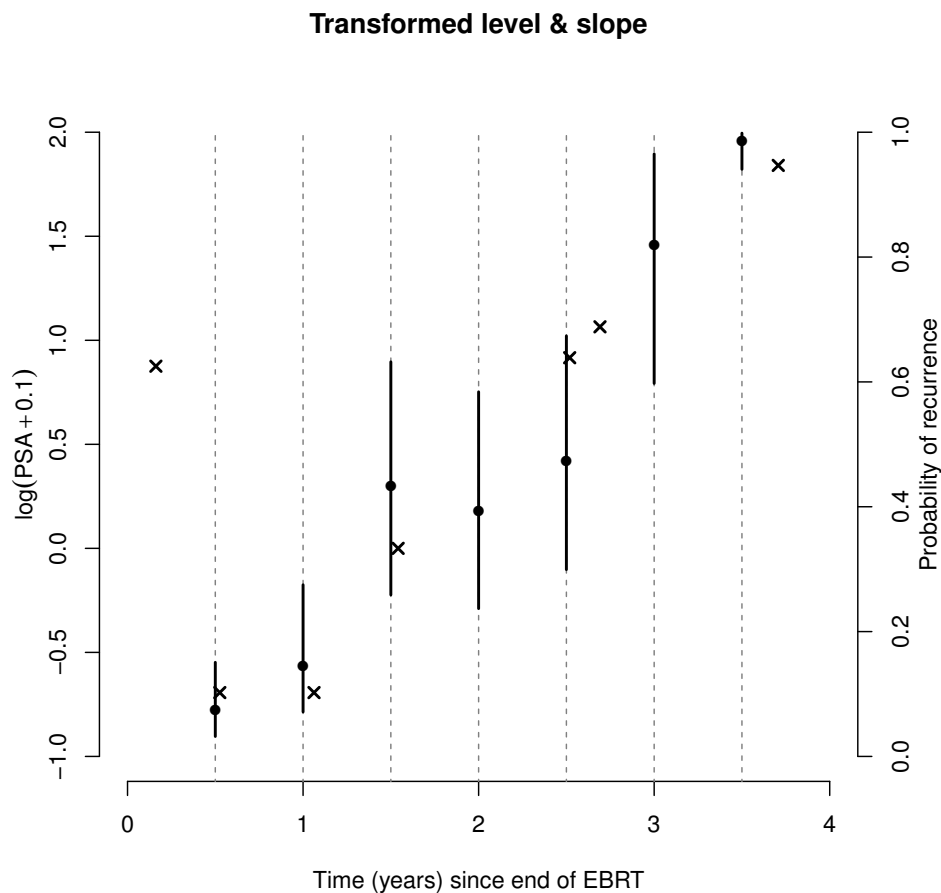


Figure III.7 – Individual dynamic predicted probability of clinical recurrence (●) within 3 years and updated every 6 months computed from model (f) and its 95% confidence bands (bold plain line) obtained by a Monte Carlo method with 2000 draws; repeated PSA measures are denoted by ×. The patient had a T-stage of 2, a Gleason score of 7, an initial PSA of 9.7ng/mL and had a clinical recurrence after 3.8 years following the end of EBRT.

We specifically focused here on SREM that differed by the nature of the dependency assumed between the biomarker dynamics and the risk of the event. Indeed, most applications of the SREM assume that only the current level of the biomarker impacts the risk of event. Yet, first the log-linear assumption induced by the inclusion of the current level of the biomarker in the proportional hazard model may not hold, and second, other characteristics of the biomarker trajectory can better summarize the dependency between the two processes than the current level. For example, in the application, we first found using Martingale residuals that the log-linear assumption of the current true level of the

biomarker did not hold and that a logistic transformation of this current true level should be considered instead. Second, we found that the current slope of the biomarker was the most important predictor of the risk of recurrence.

As we were interested both in the fit of the models to the data, and the models ability to predict the risk of event, we evaluated both the goodness-of-fit and the predictive accuracy of the models. Interestingly, we found that the model with the best Akaike criterion and adequate Martingale residuals (i.e. the best goodness-of-fit) was not the best model for predicting the risk of recurrence whatever the time elapsed since the end of the treatment. This illustrates the difference between these two types of evaluation. The goodness-of-fit measures use all the information available while the predictive accuracy measures use only a part of the longitudinal biomarker (the history of the biomarker up to time s) and only a part of the sample (the subjects still at risk at time s). This raises the question of the methodology which should be favored when developing dynamic predictive tools. When interested only in the predictive ability of a model and the development of a dynamic predictive tools, predictive accuracy measures may be preferable over goodness-of-fit measures.

Among predictive accuracy measures, we only used the prognostic cross-entropy ([Proust-Lima et al., 2012](#); [Commenges et al., 2012](#)) which offers a natural way to focus on the ability to predict the event from the history of the marker. Yet, other measures could be investigated as the Brier score ([Gerds and Schumacher, 2006](#); [Gerds and Schumacher, 2007](#); [Proust-Lima et al., 2012](#)) and AUC ([Zheng and Heagerty, 2007](#); [Rizopoulos, 2011b](#)). Nevertheless, it should be kept in mind that the over-optimism with predictive accuracy measures computed on the internal data always needs to be corrected. This was done using an approximated cross-validation formula for the EPOCE but no such approximation is available for the Brier score or AUC yet, and classical cross-validation technique is computationally too intensive with joint models to be considered in practice. We also note that

the EPOCE has the advantage to provide a tracking interval of the difference between EPOCE of two models. This is very useful to appreciate the gain in predictive accuracy of a model over another, and has not been developed yet for the other measures.

3 Comparaison entre les modèles à effets aléatoires partagés et les modèles à classes latentes

Dans cette seconde analyse, nous avons comparé les modèles conjoints à classes latentes et les modèles conjoints à effets aléatoires partagés. A travers une application sur les mêmes données que celles de la section précédente (données de la cohorte hospitalière de l'Université de Michigan), les modèles ont été comparés en terme de qualité d'ajustement aux données mais également en terme de pouvoir prédictif du risque de rechute clinique.

Le pouvoir prédictif des modèles a été évalué par l'erreur quadratique de prédiction sur des données externes et à travers l'entropie croisée pronostique à la fois sur les données d'estimation par validation croisée approximée et sur des données externes auxquelles nous avons accès uniquement pour cette validation. Les données de validation concernent également des patients atteints de cancer de la prostate et dont les critères d'inclusion étaient similaires à ceux de la cohorte d'estimation des modèles. L'article est disponible en annexe à la page [203](#) et résumé ci-dessous.

Données utilisées

Les données utilisées sont celles de la cohorte hospitalière provenant de l'Université de Michigan et composée de 459 sujets dont 74 (16.1%) ont observé une rechute clinique ([Taylor et al., 2005](#)). La seconde cohorte utilisée dans cette analyse sert à effectuer la validation externe de l'outil pronostique. Elle regroupe 719 patients issus du centre de l'Agence du cancer de la Colombie-Britannique de Vancouver. Parmi eux, 126 (17.5%) ont observé une rechute clinique ([Pickles et al., 2003](#)). Nous avons accès à ces données uniquement dans le cadre de la validation.

Spécification des modèles

Dans ce travail, pour les modèles conjoints à effets aléatoires nous n'avons pas considéré exactement la même spécification que dans la section précédente. Afin de comparer les deux types d'approche plus facilement, nous avons supposé dans chacune d'elles une évolution longitudinale du PSA modélisée à l'aide d'un modèle linéaire mixte prenant en compte les différentes phases de la trajectoire comme montré dans la Figure I.1 du chapitre I, et le risque de rechute clinique via un modèle à risques proportionnels en considérant une fonction de risque de base de type Weibull. Le modèle conjoint à classes latentes est estimé avec un nombre de classes allant de 1 à 5. Le cas où ce modèle contient une seule classe correspond alors au modèle où l'évolution longitudinale du PSA et le risque de rechute clinique sont complètement indépendants. Le modèle conjoint à effets aléatoires partagés est estimé, quant à lui, en considérant une dépendance soit à travers le niveau courant du PSA, soit à travers la pente courante ou soit par la combinaison de ces deux valeurs. Le modèle à classes latentes avec une classe constitue donc aussi un cas particulier de ce modèle conjoint à effets aléatoires partagés.

Les facteurs pronostiques standards, c'est à dire le taux de PSA pré-radiothérapie, le stade du cancer ainsi que le score de Gleason, ont été inclus dans le modèle longitudinal, dans le modèle de survie et dans le modèle d'appartenance aux classes pour les modèles à classes latentes.

Estimation des modèles

Les résultats sur la qualité d'ajustement des différents modèles sont présentés dans la Table I de la page 12 de l'article en annexe de ce manuscrit.

Quel que soit le nombre de classes ($G > 2$), l'ajustement du modèle à classes latentes est largement meilleur (BIC=5132,42 ; 5059,54 ; 5068,35 et 5101,35 pour $G = 2, 3, 4$ et

5 classes) comparé à celui du modèle sans association entre les deux processus ($G = 1$ avec $BIC=5594,71$). On remarque cependant que le nombre de paramètres augmente rapidement avec le nombre de classes, ce qui n'est pas sans conséquence sur le temps d'estimation de ces modèles.

Tout type de dépendance confondue, l'ajustement du modèle conjoint à effets aléatoires partagés est également amélioré ($BIC=5479,55$; $5450,26$ et $5445,07$ pour une dépendance à travers le niveau courant, la pente courante et les deux) comparé au modèle sans association. De plus, le nombre de paramètres reste raisonnable.

Puisque ces derniers modèles n'incluent qu'un ou deux paramètres supplémentaires, il n'était pas surprenant que l'ajustement soit donc bien meilleur pour les modèles conjoints à classes latentes avec un gain en BIC d'au moins 300 points sur les modèles conjoints à effets aléatoires partagés.

Qualité prédictive

Le pouvoir prédictif des différents modèles est évalué à l'aide du score de Brier et de l'entropie croisée pronostique. Chacune des deux mesures est calculée à partir de différents temps de prédiction allant de 1 à 6 années après la fin de la radiothérapie. Pour chaque temps de prédiction, les mesures sont calculées uniquement chez les sujets encore à risque. Les deux mesures sont calculées sur le modèle à classes latentes avec un nombre de classes allant de 1 à 5 ainsi que sur le modèle à effets aléatoires partagés doté d'une dépendance à travers le niveau courant et la pente courante du PSA. Plus les mesures sont petites, meilleurs sont les modèles.

Concernant l'entropie croisée pronostique, elle est évaluée à la fois sur les données d'estimation par validation croisée approximée (voir Figure 3 (A) page 14) et sur les données de validation (voir Figure 4 (A) page 15). Des différences d'entropie sont également calculées entre des paires de modèles permettant ainsi de les comparer entre eux (voir

Figure 3 (B) page 14 et Figure 4 (B) page 15).

Sur les données d'estimation, quel que soit le nombre de classes ($G > 2$), la qualité prédictive du modèle à classes latentes est meilleure que celle du modèle sans association. Il en est de même pour le modèle conjoint à effets aléatoires partagés. Les modèles avec $G = 3, 4$ ou 5 classes ont des qualités prédictives proches. Les différences d'entropie montrent qu'il n'y a pas de différence significative entre les modèles avec $G = 3$ et $G = 4$ classes. Elles montrent également que le modèle avec $G = 2$ classes et le modèle à effets aléatoires partagés sont moins bons que le modèle avec $G = 4$ classes sur le court-terme, mais qu'il n'y a de différence significative sur le long-terme.

Sur les données de validation, les qualités prédictives des différents modèles sont très proches et on voit qu'il n'y a pas de différence significative entre le modèle à effets aléatoires partagés et les modèles avec $G = 3$ classes ou $G = 4$ classes outre autour de 2.5 ans après la fin de la radiothérapie. Même s'il n'y a pas de différence significative entre les modèles avec $G = 2$ et $G = 4$ classes sur le long-terme, le modèle avec $G = 4$ classes apparaît aussi le meilleur modèle sur le court-terme.

Le score de Brier est évalué sur les données de validation et les résultats sont présentés dans sur la Figure 4 (C) et la Figure 4 (D) de la page 15 de l'article. Comme évoqué dans la section 4.2.1 du chapitre II, il est nécessaire de prendre en compte la censure dans l'estimation du score de Brier. Les deux estimateurs présentés en section 4.2.1 ont été calculés ici.

Les résultats des deux estimateurs du score de Brier vont dans le même sens. Ils montrent que l'erreur de prédiction est réduite avec le modèle à classes latentes avec au moins 2 classes et avec le modèle à effets aléatoires partagés. Et globalement ce dernier se comporte bien.

Conclusion

A travers cette illustration, nous avons vu que la modélisation conjointe d'un biomarqueur longitudinal et d'un temps d'événement associé améliorerait considérablement l'ajustement du modèle comparé à une analyse séparée des deux processus. L'évaluation de la qualité prédictive a également démontré une réduction de l'erreur de prédiction lors de l'utilisation de ce type de modélisation.

Les différents modèles conjoints à effets aléatoires partagés ont tous révélé des effets significatifs du niveau courant et de la pente courante du PSA. En effet, une augmentation du niveau ou de la pente est associée à une augmentation du risque de rechute clinique, ce qui confirme l'existence d'un lien entre l'évolution du PSA et le risque de rechute clinique. Ces résultats confortent également la nécessité de modéliser conjointement ces deux processus.

Pour finir, notons que les modèles conjoints à effets aléatoires partagés présentent un ajustement aux données nettement moins bon que celui des modèles conjoints à classes latentes. Cette différence est grandement atténuée en ce qui concerne la qualité prédictive des modèles même si sur cet exemple le modèle à classes latentes donnait encore une meilleure qualité prédictive.

4 Conclusion

L'objectif de ce chapitre était de comparer différentes approches de modélisation conjointe dans le cadre de la prédiction de risque de rechute clinique. Pour ce faire, un premier travail a consisté à comparer différentes formes de dépendance dans un modèle conjoint à effets aléatoires partagés. Dans un second travail, nous avons comparé les modèles conjoints à classes latentes et les modèles conjoints à effets aléatoires partagés. Ces comparaisons ont été faites en terme de qualité d'ajustement mais surtout en terme de pouvoir prédictif.

Au cours de ces travaux, nous avons observé que le modèle présentant le meilleur ajustement n'était pas forcément celui ayant la meilleure qualité prédictive. En effet, la qualité d'ajustement est évaluée sur l'ensemble des mesures des sujets à la fois dans le sous-modèle longitudinal et celui de survie, tandis que la qualité prédictive est évaluée sur la partie survie spécifiquement et à chaque temps de prédiction uniquement chez les sujets encore à risque, et à partir des mesures collectées jusqu'à ce temps. Dans le cas où l'objectif principal d'une étude est de faire de la prédiction, la qualité prédictive devrait donc être systématiquement évaluée voire privilégiée par rapport à la qualité d'ajustement.

Nous avons également vu, à travers les résidus de martingale, que certains types de dépendance du modèle conjoint à effets aléatoires partagés ne respectaient pas forcément certaines hypothèses du modèle, notamment l'hypothèse de log-linéarité du modèle à risques proportionnels. Souvent les analyses sur les modèles conjoints à effets aléatoires partagés se limitent à une dépendance à travers le niveau courant du biomarqueur. Alors que nous avons vu dans notre analyse que cette hypothèse de log-linéarité n'était pas respectée lorsque l'on considérait une dépendance à travers la pente courante du PSA uniquement ou pas respectée pour le niveau courant pour une dépendance à travers le niveau courant et la pente courante. Donc quelle que soit la forme de dépendance utilisée il est important de vérifier que les hypothèses du modèle sont bien respectées.

Chapitre III. Prédiction dynamique du risque de rechute clinique de cancer de la prostate : comparaison d'approches

A travers l'entropie pronostique croisée appliquée aux modèles conjoints à effets aléatoires partagés, nous avons pu constater que la qualité prédictive d'un modèle peut être différente suivant le type de dépendance considéré. Cela implique, dans le cadre d'une prédiction dynamique, la nécessité de considérer plusieurs types de dépendance afin de pouvoir choisir le modèle présentant la meilleure qualité prédictive.

Les deux analyses présentées dans ce chapitre représentent des illustrations sur l'utilisation des modèles conjoints pour le développement d'outils pronostiques dynamiques. Pour faciliter la pédagogie, nous nous sommes limitées aux sujets n'ayant pas reçu de second traitement au cours au temps et à une petite cohorte (459 sujets). Donc les résultats cliniques ne sont pas nécessairement généralisés. Afin d'augmenter la puissance des analyses, il serait nécessaire de considérer un plus grand échantillon de données surtout pour des modèles aussi complexes que ceux présentés dans ce chapitre. Dans le chapitre [IV](#), nous proposons une analyse et une application sur un échantillon de données plus important (2386 sujets) incluant des sujets ayant reçu un second traitement au cours de leur suivi.

Chapitre IV

Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

Dans ce chapitre, nous étendons les outils pronostiques dynamiques issus des modèles conjoints à l'évaluation du risque d'événement selon l'initiation ou non d'un second traitement.

1 Introduction

En prédiction dynamique, des outils pronostiques ont déjà été proposés ([Rizopoulos, 2011b](#); [Proust-Lima and Taylor, 2009a](#)). Ils permettent de quantifier, à tout moment, le risque d'un événement d'intérêt à partir de l'histoire du sujet mais uniquement sous l'hypothèse que le suivi du sujet ne change pas. Pourtant dans de nombreux cancers, après un premier traitement, le suivi du patient peut évoluer. Notamment, certains patients initient un second traitement dans le but de contrôler la progression de la maladie. C'est le cas de l'hormonothérapie dans le cancer de la prostate qui permet de retarder la survenue d'une rechute.

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

Même si des outils pronostiques existent pour prédire une éventuelle rechute du cancer, ils ne prennent pas en compte actuellement ce possible changement de traitement au cours du temps. Ces outils ont été développés et surtout validés sous l'hypothèse que le patient ne recevra pas de second traitement (Proust-Lima and Taylor, 2009a ; Taylor et al., 2013a). En pratique, l'initiation d'un second traitement change la dynamique du biomarqueur (voir Figure I.2 pour l'exemple du PSA) et le risque d'événement. Il est donc nécessaire de prendre en compte cette information dans la modélisation et dans les outils pronostiques dérivés.

C'est dans ce contexte que nous avons souhaité proposer des outils pronostiques différentiels de rechute de cancer de la prostate suivant l'initiation ou non d'un second traitement, l'hormonothérapie. Le principe est de calculer la probabilité d'événement sous différentes hypothèses, deux exactement : le cas où le patient initie un traitement aujourd'hui et le cas où le patient n'initie aucun traitement dans les x prochaines années.

Ces outils s'inscrivent donc parfaitement dans le cadre de la médecine personnalisée, avec l'idée de proposer le bon traitement à la bonne personne et au bon moment. Ils permettent d'apprécier le bénéfice/risque pour chaque patient d'initier ce second traitement à différentes étapes de son suivi. En effet, un traitement comme l'hormonothérapie est en général le traitement de référence après une radiothérapie chez les patients atteints de cancer localisé. Mais il a aussi de nombreux effets indésirables sur la qualité de vie des patients.

L'objectif de ce chapitre est donc de développer et valider un outil pronostique dynamique différentiel de rechute de cancer de la prostate en fonction de l'initiation ou non d'un second traitement, l'hormonothérapie et de l'histoire des PSA. Dans ce cadre, les modèles conjoints à effets aléatoires partagés étaient une approche naturelle car ils permettent d'investiguer différentes natures de dépendance entre la dynamique des PSA, l'initiation de l'hormonothérapie et le risque de rechute clinique.

Ce travail a fait l'objet d'un article actuellement soumis dans le journal "Statistical Methods in Medical Research" et présenté dans la section suivante.

2 Article

Résumé de l'article

Avec l'émergence des informations collectées sur les biomarqueurs après un traitement, de nouveaux types d'outils pronostiques ont vu le jour : il s'agit des outils pronostiques dynamiques qui peuvent être mis à jour à chaque nouvelle mesure disponible du biomarqueur. De tels outils pronostiques sont d'un grand intérêt en oncologie où, après un traitement initial, les patients sont surveillés via l'évolution des données répétées du biomarqueur. Au cours de leur suivi, si l'évolution du biomarqueur suggère un possible retour de la maladie, les patients peuvent initier un second traitement dans le but de ralentir la progression de la maladie.

L'objectif de cet article est de développer et valider des outils dynamiques de prédiction individuelle permettant de prendre en compte un possible changement de traitement au cours du temps. Ces outils permettent de prédire un événement d'intérêt à partir d'un temps de prédiction s dans les x prochaines années sous deux scénarios : (1) le patient initie immédiatement d'un second traitement, (2) le patient n'initie pas de traitement dans les x prochaines années. Ils présentent donc l'avantage d'évaluer le bénéfice de l'initiation du second traitement. Les prédictions obtenues par ces outils sont dérivées des modèles conjoints à effets aléatoires partagés, modèles qui permettent de modéliser conjointement l'évolution longitudinale d'un biomarqueur et un risque d'événement associé aux caractéristiques individuelles du biomarqueur.

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

Notre application est basée sur des données de patients atteints de cancer de la prostate localisé et traités initialement par radiothérapie externe. Pour tenir compte du traitement hormonal que certains patients ont reçu au cours de leur suivi, nous avons investigué et comparé différentes spécifications de la dépendance entre les mesures répétées du PSA, l'initiation d'une hormonothérapie et le risque de rechute clinique. La qualité prédictive des outils dynamiques issus de ces différents modèles a été évaluée à travers deux mesures, le score de Brier et l'entropie croisée pronostique. Cette validation a été réalisée séparément pour les deux scénarios. Systématiquement, des estimateurs obtenus par validation croisée approximée ont été proposés afin de pouvoir corriger le sur-optimisme présent lorsque l'on prédit sur les mêmes données que celles utilisées dans la phase d'estimation.

Personalized dynamic prediction of prostate
cancer recurrence according to the initiation
of a second treatment : development
and validation

Mbéry Sène^{1,2,*}, Jeremy M. G. Taylor^{3,4}, James J. Dignam⁵,
Hélène Jacqmin-Gadda^{1,2} and Cécile Proust-Lima^{1,2}

¹ INSERM, U897, Epidemiology and Biostatistics Research Center, 33076 Bordeaux,
France

² Université Bordeaux Segalen, 33076 Bordeaux, France

³ Department of Biostatistics, University of Michigan, Ann Arbor, MI

⁴ Department of Radiation Oncology, University of Michigan, Ann Arbor, MI

⁵ Department of Health Studies, University of Chicago and Radiation Therapy Oncology
Group, American College of Radiology, Philadelphia

Correspondence to : Mbéry Sène, INSERM U897, ISPED, Université de Bordeaux Se-
galen, 146 rue Léo Saignat, 33076 Bordeaux Cedex, FRANCE

Mbery.Sene@isped.u-bordeaux2.fr

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

Abstract : With the emergence of rich information on biomarkers after treatments, a new type of prognostic tools has come out : dynamic prognostic tools that can be updated at each new biomarker measurement. Such predictions are of interest in oncology where after an initial treatment patients are monitored with repeated biomarker data. During such follow-up, patients may receive a second treatment to slow down the progression of the disease. This paper aims to develop and validate dynamic individual predictions allowing for a change in treatment. The prediction of the event in the next x years is done under two scenarios : (1) the patient initiates immediately a second treatment, (2) the patient does not initiate any treatment in the next x years. Such predictions are useful to understand the benefit of initiating a second treatment. Predictions are derived from shared random-effect models. Applied to prostate cancer data, different specifications in the model for the dependence between the PSA repeated measures, the initiation of a second treatment (hormonal therapy) and the risk of clinical recurrence are investigated and compared. The predictive accuracy of the dynamic predictions is evaluated with two measures (Brier score and prognostic cross-entropy) for which approximated cross-validated estimators are proposed.

Keywords : Brier score, Dynamic predictions, Hormonal treatment, Joint model, Prognostic cross-entropy, Prostate cancer, Shared random-effect models

2.1 Introduction

The increase of rich longitudinal information in health studies ([Faucett and Thomas, 1996](#); [Wulfsohn and Tsiatis, 1997](#); [Henderson et al., 2000](#); [Lin et al., 2002](#); [Tsiatis and Davidian, 2004](#); [Rizopoulos, 2012](#)) has motivated the development of joint models of longitudinal and time-to-event data. In addition to providing an efficient framework to correctly model correlated longitudinal markers and clinical events and to better understand their interrelationship, these models have more recently offered a new approach for monitoring the patients after the diagnosis of a chronic disease.

Indeed, dynamic individual predictions of an event of interest can be easily derived from joint models ([Proust-Lima and Taylor, 2009a](#); [Rizopoulos, 2011b](#); [Sène et al., 2013](#)). They consist of the predicted probability of having the event in a certain window of time conditional on the longitudinal marker history. As such, in contrast with standard predictive tools that only use information at diagnosis, these predictions allow one to dynamically adapt the risk of the event of interest to the individual trajectory of the biomarker of progression.

Such predictions are of interest in oncology where after a first treatment, patients are monitored to detect early recurrence of the cancer. In particular, after prostate cancer diagnosis and initial treatment by radiation therapy, some dynamic tools that are based on repeated measures of PSA (prostate specific antigen), in addition to the standard prognostic factors were found to more accurately predict the risk of clinical recurrence than standard prognostic tools ([Proust-Lima and Taylor, 2009a](#); [Proust-Lima et al., 2012](#)).

In practice, after a first treatment of the cancer, patients may receive a second treatment (ST) to slow down the progression of the disease, and prevent or delay the recurrence of the cancer. In particular for prostate cancer, after a first treatment by radiation therapy, the patient may initiate hormonal therapy (HT) when he has a high risk of clinical recur-

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

rence. The optimal time to initiate such HT is unknown. The timing is mainly determined by the clinician based on experience, the knowledge of the disease and the observation of the PSA trajectory. Yet, HT has consequences for the patient's personal life, so that following the principle of personalized medicine, defined as the application of the right treatment for the right patient at the right dose and at the right time ([Welsh and Powis, 2009](#)), assessing the benefit for the patient (in terms of reduction of his risk of recurrence) of initiating a ST is of great importance.

Dynamic predictive tools can be used for this purpose as they provide an up-to-date quantification of the risk of recurrence. However, the dynamic predictions currently developed do not take into account a possible change of treatment ([Yu et al., 2008](#); [Proust-Lima and Taylor, 2009a](#); [Rizopoulos, 2011b](#); [Taylor et al., 2013a](#)): they are computed by assuming the patient will not initiate any ST in the window of prediction. As ST's usually change the dynamics of the biomarker and of the time-to-event, such a joint model would be more complex to define ([Kennedy et al., 2010](#); [Taylor et al., 2013a](#)) and differential dynamic predictions would be required to distinguish the risk of event according to the initiation of ST.

In this context, the objective of this paper is to develop and validate different dynamic individual predictions of event in the next x years for two conditions: (1) whether the patient initiates a ST today or (2) whether the patient does not initiate any ST in the next x years. To reach this goal, we utilized a joint model and focused on the prostate cancer example with initiation of hormonal therapy. We explored different specifications of the interrelationships between the PSA repeated measures, the initiation of HT and the risk of clinical recurrence. Dynamic individual predictions were computed in situations (1) and (2), and different validation strategies were considered.

Validation of dynamic predictive tools is a new area of research. Measures, such as the AUC are appropriate for discrimination, but for predictive accuracy other measures have

been proposed : including an adaptation to the dynamic setting of the quadratic error of prediction (Brier Score) (Schoop et al., 2011; Proust-Lima et al., 2012) and the expected prognostic observed cross-entropy (EPOCE) (Proust-Lima et al., 2012; Commenges et al., 2012) estimated by the prognostic log-likelihood. When no external data are available, the Brier score estimate has to be computed using a cross-validation technique to avoid any problem of overoptimism, but this is usually computationally too demanding with joint models. In contrast, thanks to an approximated cross-validation estimator, the EPOCE can be computed directly and easily from the data on which the joint model was estimated. This approximated cross-validation technique was recently generalized to any regular problem (Commenges et al., 2013) so that in the present work, we propose to use an approximated cross-validated Brier Score estimate along with the EPOCE measure to validate the dynamic predictive tools. In situation (1) that focuses on the subsequent years after initiation of a ST, integrated versions of these predictive accuracy measures over the times of initiation of the ST are computed, again correcting for overoptimism.

The paper is organized as follows. Section 2 describes the joint model methodology. Section 3 includes the development and validation of the differential individual dynamic predictions. Section 4 describes the application to Prostate Cancer with the evaluation of the risk of clinical recurrence based on the PSA trajectory after initial radiation therapy, and on the immediate initiation of hormonal therapy. Finally, the model and the results of the application are discussed in section 5.

2.2 Joint models

The shared random-effect approach was chosen to jointly model the biomarker repeated measures and the time-to-event. It consists in two submodels : a longitudinal model and a time-to-event model. A general description is given below. More specific models will be described in the application section.

2.2.1 Notation

Let T_i^* be the time to the event of interest and C_i the censoring time for subject i , $i = 1, \dots, N$. We observe the time $T_i = \min(T_i^*, C_i)$ and $E_i = \mathbb{1}_{\{T_i^* \leq C_i\}}$ is the indicator of event. Let τ_i be the time to ST (unobserved if $\tau_i > T_i$). Let $\mathbb{1}_{\{t \geq \tau_i\}}$ be the indicator of ST status at any time t . For each subject we also observe n_i repeated measures of the biomarker $Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{in_i}))$ collected intermittently at times $(t_{i1}, \dots, t_{in_i})$. We denote by X_{Li} the matrix of (possibly time-dependent) covariates involved in the longitudinal submodel and by X_{Si} the vector of time-independent covariates involved in the survival submodel.

2.2.2 Longitudinal submodel

The biomarker trajectory is described using a linear mixed model (Laird and Ware, 1982). We assume that for $j = 1, \dots, n_i$, the repeated measures $Y_i(t_{ij})$ are noisy measures of $Y_i^*(t_{ij})$ the true unobserved biomarker value. The mean change over time of $Y_i^*(t_{ij})$ can depend on covariates and the within-subject correlation of the biomarker repeated measures is captured using subject-specific random effects :

$$\begin{aligned} Y_i(t_{ij}) &= Y_i^*(t_{ij}) + \epsilon_i(t_{ij}) \\ &= X_{Li}(t_{ij})^T \beta + Z_i(t_{ij})^T b_i + \epsilon_i(t_{ij}) \end{aligned} \quad (\text{IV.1})$$

where $X_{Li}(t_{ij})$ and $Z_i(t_{ij})$ are a p-vector and a q-vector of time-dependent covariates associated respectively with the p-vector of fixed effects β and the q-vector of random effects b_i , $b_i \sim \mathcal{N}(0, B)$. The vector of independent errors of measurement $\epsilon_i = (\epsilon_i(t_{i1}), \dots, \epsilon_i(t_{in_i})) \sim \mathcal{N}(0, \Sigma_i = \sigma^2 I_{n_i})$; ϵ_i and b_i are independent.

2.2.3 Survival submodel

To model the risk of the event and to quantify the effects of the biomarker dynamics and the initiation of ST on this risk adjusted for other covariates, we define a proportional hazard model as follows :

$$\lambda_i(t|X_{Si}, b_i, \tau_i) = \lambda_0(t) \exp \left[X_{Si}^T \gamma + W_i(b_i, \tau_i, t)^T \varphi \right] \quad (\text{IV.2})$$

where $\lambda_0(t)$ is the baseline hazard function, γ is the r-vector of coefficients associated with the r-vector of time-independent covariates X_{Si} , and $W_i(b_i, \tau_i, t)$ is a multivariate function of the random effects b_i from model (IV.1) and the initiation of ST when ST is considered. It is associated with the vector of parameters φ . In the standard framework without any ST, the vector φ measures the association between the longitudinal and the survival processes while $W_i(b_i, \tau_i, t)$ defines the nature of the dependence between the two processes. In the presence of a ST, $W_i(b_i, \tau_i, t)$ captures both the dependence between the two processes and the implication of ST in this dependence.

2.2.4 Maximum likelihood estimation

Maximum likelihood estimates were obtained using the JM R package (Rizopoulos, 2010), with modifications of the JM source code when necessary. The maximum likelihood estimates, denoted $\hat{\theta}$, were obtained by a quasi-Newton algorithm with a convergence criterion on the log-likelihood (Rizopoulos, 2010). The two integrals involved in the log-likelihood computation were approximated using Gaussian quadrature (Rizopoulos, 2010). Estimates of the variance-covariance matrix of the estimated parameters $\widehat{V}(\hat{\theta})$ were provided by the inverse of the Hessian matrix.

2.3 Individual dynamic predictions

Individual dynamic predictions derived from joint models (Proust-Lima and Taylor, 2009a; Rizopoulos, 2011b; Proust-Lima et al., 2012; Sène et al., 2013) consist in the individual predicted probability of event, $p_i(s, \mathcal{T}; \theta)$, between times s and $s + \mathcal{T}$ computed for a new subject given $Y_i^{(s)} = \{Y_i(t_{ij}), j = 1, \dots, n_i, \text{ such as } t_{ij} \leq s\}$ his biomarker history until s , the covariates history $X_{Li}^{(s)} = \{X_{Li}(t_{ij}), j = 1, \dots, n_i, \text{ such as } t_{ij} \leq s\}$ and time-independent covariates X_{Si} . It is defined as :

$$p_i(s, \mathcal{T}; \theta) = \mathbb{P}\left(T_i \leq s + \mathcal{T} | T_i \geq s, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta\right) \quad (\text{IV.3})$$

With a change in risk due to the initiation of a ST, two individual dynamic predictions can be distinguished from a time of prediction s in patients free of ST :

- The patient initiates ST at time s :

$$\begin{aligned} p_i(s, \mathcal{T}; \theta) &= \mathbb{P}\left(T_i \leq s + \mathcal{T} | T_i \geq s, \tau_i = s, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta\right) \\ &= 1 - \frac{\int_{b_i} f_{Y^{(s)}}\left(Y_i^{(s)} | X_{Li}^{(s)}, b_i; \theta\right) S_i(s + \mathcal{T} | X_{Si}, \tau_i = s, b_i; \theta) f_b(b_i; \theta) db_i}{\int_{b_i} f_{Y^{(s)}}\left(Y_i^{(s)} | X_{Li}^{(s)}, b_i; \theta\right) S_i(s | X_{Si}, \tau_i = s, b_i; \theta) f_b(b_i; \theta) db_i} \end{aligned} \quad (\text{IV.4})$$

- The patient does not initiate any ST before $s + \mathcal{T}$:

$$\begin{aligned} p_i(s, \mathcal{T}; \theta) &= \mathbb{P}\left(T_i \leq s + \mathcal{T} | T_i \geq s, \tau_i > \min(T_i, s + \mathcal{T}), Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta\right) \\ &= 1 - \frac{\int_{b_i} f_{Y^{(s)}}\left(Y_i^{(s)} | X_{Li}^{(s)}, b_i; \theta\right) S_i(s + \mathcal{T} | X_{Si}, \tau_i > \min(T_i, s + \mathcal{T}), b_i; \theta) f_b(b_i; \theta) db_i}{\int_{b_i} f_{Y^{(s)}}\left(Y_i^{(s)} | X_{Li}^{(s)}, b_i; \theta\right) S_i(s | X_{Si}, \tau_i > \min(T_i, s + \mathcal{T}), b_i; \theta) f_b(b_i; \theta) db_i} \end{aligned} \quad (\text{IV.5})$$

In (IV.4) and (IV.5), $f_{Y^{(s)}}$ and f_b are multivariate Gaussian density functions with respectively means $X_{Li}^{(s)} \beta + Z_i^{(s)} b_i$ and 0, and variance-covariance matrices $\Sigma_i^{(s)}$ and B ; S_i is the survival function. $X_{Li}^{(s)}$ and $Z^{(s)}$ are design matrices with respectively $X_{Li}^{(s)}(t_{ij})^T$ and $Z^{(s)}(t_{ij})^T$ as row vectors with $t_{ij} \leq s$, and $\Sigma_i^{(s)}$ is the submatrix of Σ_i with $t_{ij} \leq s$.

A point estimate of these individual dynamic predictions can be obtained with $p_i(s, \mathcal{T}; \hat{\theta})$. The posterior distribution of $p_i(s, \mathcal{T}; \theta)$ can be also approximated by a Monte Carlo method (Proust-Lima and Taylor, 2009a) : a large set of $\theta^{(d)}$ ($d = 1, \dots, D$) is generated from the asymptotic distribution of the estimates $\mathcal{N}\left(\hat{\theta}, \widehat{V}(\hat{\theta})\right)$ and $p_i(s, \mathcal{T}; \theta^{(d)})$ is computed. The median value of $p_i(s, \mathcal{T}; \theta^{(d)})$ provides the point estimate and the 2.5% and 97.5% percentiles give a 95% confidence band. Instead of computing the probabilities using (IV.4) and (IV.5), $b_i^{(d)}$ can be sampled from its posterior distribution, and the probabilities computed given $b_i^{(d)}$ (Rizopoulos, 2011b).

2.4 Evaluation of predictive accuracy

Validation of the prognostic tools is done in terms of predictive accuracy. Two measures adapted to the dynamic setting were considered : the prognostic cross-entropy and the Brier score. The measures and simple estimators of them are described. The formula for the approximate cross-validation version (Commenges et al., 2013) is applied to provide estimators of the measures that can be used on training data to correct overoptimism. Finally, integrated versions over the times of initiation of ST are proposed and confidence intervals are given.

In the following, N_s is the number of subjects at risk at time of prediction s . However, as predictive performances are evaluated differently in the absence of ST and after immediate initiation of ST, in the first case $N_s = \sum_{i=1}^N \mathbb{1}_{\{T_i > s \text{ \& } \tau_i > \min(T_i, s + \mathcal{T})\}}$ while in the second case $N_s = \sum_{i=1}^N \mathbb{1}_{\{T_i > s \text{ \& } \tau_i = s\}}$.

2.4.1 Measures of predictive accuracy

a) Expected Prognostic Observed Cross-Entropy

The expected prognostic observed cross-entropy (EPOCE) is defined as the expectation of the log of the conditional density $f_{T|Y^{(s)}, T^* \geq s}$ of the time to event given the history of the biomarker $Y^{(s)}$ until the time of prediction s , $E \left[-\log \left(f_{T|Y^{(s)}, T^* \geq s} | T^* \geq s \right) \right]$. The simple estimator of EPOCE, called POL for prognostic observed log-likelihood, in a window of time $[s, s + \mathcal{T}]$ is defined by :

$$\text{POL}(\hat{\theta}, s, \mathcal{T}) = \frac{1}{N_s} \sum_{i=1}^{N_s} F_i(\hat{\theta}, s, \mathcal{T}) \quad (\text{IV.6})$$

where $F_i(\hat{\theta}, s, \mathcal{T})$ is minus the observed individual contribution to the conditional log-likelihood defined below in (IV.7) given that the subject is still at risk at time s and given his covariates history until s . For $i = 1, \dots, N_s$, F_i is defined as :

$$F_i(\hat{\theta}, s, \mathcal{T}) = -\mathbb{1}_{\{T_i \geq s\}} \log \left(\frac{\int_{b_i} f_Y(Y_i^{(s)} | X_{L_i}^{(s)}, b_i; \theta) \lambda(\tilde{T}_i | X_{S_i}, b_i; \theta)^{\tilde{E}_i} S_i(\tilde{T}_i | X_{S_i}, b_i; \theta) f_b(b_i; \theta) db_i}{\int_{b_i} f_Y(Y_i^{(s)} | X_{L_i}^{(s)}, b_i; \theta) S_i(s | X_{S_i}, b_i; \theta) f_b(b_i; \theta) db_i} \right) \quad (\text{IV.7})$$

where $\tilde{T}_i = \min(T_i, s + \mathcal{T})$ and $\tilde{E}_i = \mathbb{1}_{\{T_i \leq s + \mathcal{T}\}}$ which means that subjects are artificially censored at $s + \mathcal{T}$.

b) Brier Score and Integrated Brier Score

The Brier Score (BS) developed in survival models (Gerds and Schumacher, 2006; Gerds and Schumacher, 2007) was extended to joint models (Henderson et al., 2002; Proust-Lima and Taylor, 2009a; Proust-Lima et al., 2012; Schoop et al., 2011).

It consists in $E \left[\left(\eta(s + \mathcal{T}) - \hat{S}(s + \mathcal{T}|s; \hat{\theta}) \right)^2 \right]$, the expectation of the squared difference between the observed survival status η and the predicted survival at a specific time : $\hat{S}(s + \mathcal{T}|s; \hat{\theta}) = 1 - p_i(s, \mathcal{T}; \hat{\theta})$ with $p_i(s, \mathcal{T}; \hat{\theta})$ defined in (IV.5) in the absence of HT and in (IV.4) after immediate initiation of HT. The simple estimator of BS at time $s + \mathcal{T}$ is :

$$\text{BS}(\hat{\theta}, s, \mathcal{T}) = \frac{1}{N_s} \sum_{i=1}^{N_s} w_i \left(\eta_i(s + \mathcal{T}) - \hat{S}(s + \mathcal{T}|s; \hat{\theta}) \right)^2 \quad (\text{IV.8})$$

where w_i is a weight that compensates for the loss of information due to censoring. The weights are defined according to the inverse probability of censoring (Graf et al., 1999; Proust-Lima et al., 2012) as $w_i = \frac{\mathbb{1}_{\{T_i > s + \mathcal{T}\}}}{\hat{G}(s + \mathcal{T}) / \hat{G}(s)} + \frac{E_i \mathbb{1}_{\{T_i \leq s + \mathcal{T}\}}}{\hat{G}(T_i) / \hat{G}(s)}$ with \hat{G} the Kaplan-Meier estimate of the survival function for the censoring.

An average prediction accuracy is derived from BS in a window $[s, s + \mathcal{T}]$ by integrating the quantity over the horizon times (Henderson et al., 2002; Proust-Lima et al., 2012). Again, to account for the loss of events due to censoring, a weighted mean can be used (Henderson et al., 2002) and estimated by :

$$\text{IBS}(\hat{\theta}, s, \mathcal{T}) = \left[\sum_k^{n_s^{\mathcal{T}}} d_k^{(s)} \left(\hat{G}(s) / \hat{G}(t_k) \right) \text{BS}(\hat{\theta}, s, t_k - s) \right] / \left[\sum_k^{n_s^{\mathcal{T}}} d_k^{(s)} \left(\hat{G}(s) / \hat{G}(t_k) \right) \right] \quad (\text{IV.9})$$

where t_k ($k = 1, \dots, n_s^{\mathcal{T}}$) are the distinct $n_s^{\mathcal{T}}$ times of events in the window $[s, s + \mathcal{T}]$ and $d_k^{(s)}$ is the number of events at each time t_k among subjects at risk at time s .

2.4.2 Approximated cross-validated estimators

To provide a valid assessment of predictive accuracy, these measures should be computed for independent data. On the model estimation data, a cross-validation technique is

required to correct for overoptimism. In complex situations, cross-validation is numerically too expensive to be used, so we applied the approximate leave-one-out cross-validation formula for regular problems proposed by [Commenges et al., \(2013\)](#). It is defined as :

$$\text{CV } \mathcal{M}_a(\hat{\theta}, s, \mathcal{T}) = \mathcal{M}(\hat{\theta}, s, \mathcal{T}) + N \text{Trace} \left(H^{-1} K_{s, \mathcal{T}} \right) \quad (\text{IV.10})$$

where $\mathcal{M}(\hat{\theta}, s, \mathcal{T})$ is the simple estimator (POL, BS or IBS) and the second term is the correction term for overoptimism obtained from the approximated leave-one-out cross-validation. H is the Hessian matrix of the joint log-likelihood, and $K_{s, \mathcal{T}} = \frac{1}{N_s(N-1)} \sum_{i=1}^N \mathbb{1}_{\{T_i \geq s\}} \hat{v}_i(s, \mathcal{T}) \hat{d}_i^T$ is the product of the gradients $\hat{v}_i(s, \mathcal{T})$ and \hat{d}_i of the individual contributions respectively to the simple estimator $\mathcal{M}(\hat{\theta}, s, \mathcal{T})$ and the maximized joint log-likelihood. The gradients are computed using finite differences.

Applied to POL, BS and IBS respectively, the approximate cross-validation estimators are respectively called CVPOL_a , CVBS_a and CVIBS_a ; H and \hat{d}_i are the same in the three

approximate cross-validation measures and $\hat{v}_i(s, \mathcal{T}) = \frac{\partial F_i(\theta, s, \mathcal{T})}{\partial \theta} \Big|_{\hat{\theta}}$ for CVPOL_a ,

$$\hat{v}_i(s, \mathcal{T}) = -2 w_i \frac{\partial S(s + \mathcal{T} | s; \theta)}{\partial \theta} \Big|_{\hat{\theta}} \left(\eta_i(s + \mathcal{T}) - \hat{S}(s + \mathcal{T} | s; \hat{\theta}) \right) \text{ for } \text{CVBS}_a \text{ and}$$

$$\hat{v}_i(s, \mathcal{T}) = \sum_k^{n_s^{\mathcal{T}}} d_k^{(s)} \frac{(\hat{G}(s) / \hat{G}(t_k))}{\sum_k^{n_s^{\mathcal{T}}} d_k^{(s)} (\hat{G}(s) / \hat{G}(t_k))} \left[-2 w_i \frac{\partial S(t_k | s; \theta)}{\partial \theta} \Big|_{\hat{\theta}} \left(\eta_i(t_k) - \hat{S}(t_k | s; \hat{\theta}) \right) \right] \text{ for } \text{CVIBS}_a.$$

2.4.3 Averaged predictive accuracy

When predicted probabilities are computed for the case of immediate initiation of a ST, the predictive accuracy is evaluated from the time of initiation of the ST. In practice, this time is different for each subject.

So instead of an evaluation at fixed times of prediction s , predictive accuracy is evaluated using an integrated measure over the τ_i -distribution with the weighted average :

$$\overline{\mathcal{M}(\hat{\theta}, \mathcal{T})} = \left[\sum_{i=1}^{n^{ST}} \frac{d_i^{ST}}{\hat{G}_\tau(\tau_i)} \mathcal{M}(\hat{\theta}, \tau_i, \mathcal{T}) \right] / \left[\sum_{i=1}^{n^{ST}} \frac{d_i^{ST}}{\hat{G}_\tau(\tau_i)} \right] \quad (\text{IV.11})$$

where \mathcal{M} can be BS or POL (simple or cross-validated) computed in (IV.6), (IV.8) and (IV.10); n^{ST} is the number of distinct times of ST initiations, d_i^{ST} is the number of ST initiations at time τ_i and \hat{G}_τ is the Kaplan-Meier estimate of the survival function of censoring related to times of ST initiation τ_i .

2.4.4 Confidence interval

Predictive accuracy measures between two models can be compared by computing the difference of their approximated cross-validation estimators with a 95% confidence interval (C.I.) (Commenges et al., 2013). It is computed from the asymptotic distribution of the predictive accuracy difference and its empirical variance.

Let $\hat{\theta}_A$ and $\hat{\theta}_B$ the vectors of parameter estimates for the two models A and B . Let $\Delta(A(\hat{\theta}_A), B(\hat{\theta}_B))$ the difference of predictive accuracy between the two models and $\mathcal{D}(A(\hat{\theta}_A), B(\hat{\theta}_B))$ its estimator. Let m be the number of subjects on which the predictive accuracy is computed. It is shown (Commenges et al., 2013) that the difference between $\mathcal{D}(A(\hat{\theta}_A), B(\hat{\theta}_B))$ and $\Delta(A(\hat{\theta}_A), B(\hat{\theta}_B))$ is asymptotically normal :

$$m^{1/2} \left[\mathcal{D}(A(\hat{\theta}_A), B(\hat{\theta}_B)) - \Delta(A(\hat{\theta}_A), B(\hat{\theta}_B)) \right] \rightarrow \mathcal{N}(0, w_*^2) \quad (\text{IV.12})$$

where w_*^2 can be estimated by the empirical variance \hat{w}^2 of the difference of the simple estimators. With z_u the u^{th} quantile of a standard normal variable, the confidence interval is then $\left[\mathcal{D}(A(\hat{\theta}_A), B(\hat{\theta}_B)) - z_{\alpha/2} m^{-1/2} \hat{w}; \mathcal{D}(A(\hat{\theta}_A), B(\hat{\theta}_B)) + z_{\alpha/2} m^{-1/2} \hat{w} \right]$.

In absence of ST, the predictive accuracy is evaluated at different times s with $m = N_s$ and \mathcal{D} is the approximated cross-validation estimate of POL, BS or IBS at time s . After initiation of HT, predictive accuracy is evaluated once with $m = \sum_{i=1}^{n^{ST}} d_i^{ST}$ and \mathcal{D} is the approximated cross-validation estimate of the average measures defined in (IV.11).

2.5 Application to the prediction of prostate cancer recurrence

2.5.1 Datasets

Data used in this application consist of 2386 men treated for localized prostate cancer by external beam radiation therapy (EBRT) in three different studies : 503 patients come from the cohort of the University of Michigan (UM) with a period of recruitment from 1998 to 2004; 1268 patients come from the cohort of Beaumont Hospital, in Michigan (BM), recruited between 1987 and 2003; 615 patients come from the multicenter clinical trial RTOG9406 recruited from 1994 to 2001. Among them, 261 (10.9%) received a ST that is hormonal therapy (HT) during their follow-up. The definition of clinical recurrence was as any kind of recurrence (local, regional, distant) or death from prostate cancer and only the first clinical recurrence was considered. 312 (13.1%) patients had a clinical recurrence among which 53 received HT. The four baseline prognostic factors considered in this application are the pre-radiation therapy level of PSA, the T-stage which indicates how large the tumor is and how far it has spread (in three categories, 2 and 3-4 versus 1), the Gleason score which quantifies the aggressiveness of the cancer (in three categories, 7 and 8-10 versus 2-6) and the corrected total dose of radiation therapy (Proust-Lima et al., 2011); full descriptions of these covariates have been previously done (Proust-Lima et al., 2008; Proust-Lima and Taylor, 2009a).

Figure IV.1 shows 8 random individual observed trajectories of PSA after the end of EBRT for patients who recurred or were censored and patients who did or did not receive HT. After EBRT a drop in PSA is observed in the first year. Then a subsequent rise of PSA

indicates a higher risk of recurrence and in some cases HT is initiated to reduce the risk and postpone the recurrence. Initiation of HT induces an immediate change in the PSA dynamics. Since the objective was to predict the risk of recurrence based on PSA dynamics prior to HT we chose to censor the post-HT PSA data. We call the PSA dynamics as if the person were not treated with HT the base PSA dynamics.

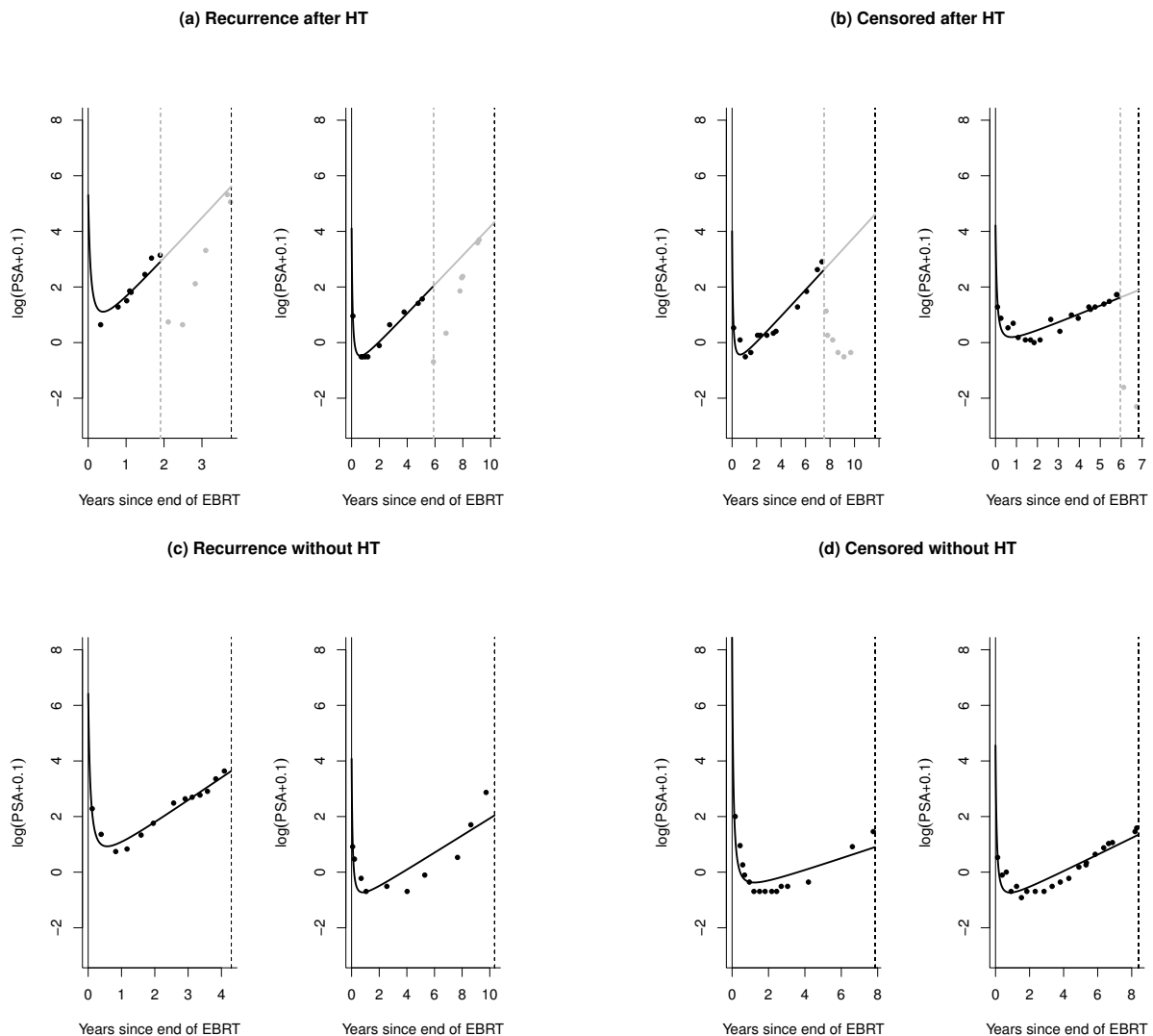


Figure IV.1 – Individual observed trajectories of $\log(\text{PSA} + 0.1)$ after the end of EBRT until the observed survival time (at the vertical black dash line) : (a) for two patients who received HT (at vertical grey dashed line) and who subsequently recurred, (b) for two patients who received HT and were subsequently censored, (c) for two patients who recurred without initiating any HT and (d) for two patients who were censored without initiating any HT. Black dots represent observed values of PSA before HT and the black curve represents the subject-specific PSA predictions from the linear mixed model. Grey dots are the observed PSA values after HT and the grey curve represents the extrapolated subject-specific PSA predictions from the mixed model based only on pre-HT data. It gives the expected PSA trajectory assuming the patient did not receive any HT.

2.5.2 Specification of the joint models

PSA repeated measures were analyzed in the logarithm scale. As previously proposed (Proust-Lima et al., 2008) we used the two phases trajectory of PSA defined as :

$$\begin{aligned} Y_i(t) &= \log(\text{PSA}_i(t) + 0.1) \\ &= \left(X_{0i}^T \beta_0 + b_{0i}\right) + \left(X_{1i}^T \beta_1 + b_{1i}\right) f(t) + \left(X_{2i}^T \beta_2 + b_{2i}\right) t + \epsilon_i(t), \forall t \in \mathbb{R}^+ \quad (\text{IV.13}) \end{aligned}$$

where X_{0i} included 1, the value of PSA pre-EBRT and the cohort indicators; X_{1i} included, in addition to X_{0i} , binary indicators for the T-stage; and X_{2i} included, in addition to X_{1i} , binary indicators for the Gleason score. The time function $f(t) = ((1+t)^{-1.5} - 1)$ captured the PSA dynamics in the decline phase while t modeled the long-term trend (Proust-Lima et al., 2008).

In the survival model, the baseline hazard function was approximated by splines and the four baseline prognostic factors were included in X_s . In addition, different specifications of $W_i(b_i, \tau_i, t)$ were explored. $W_i(b_i, \tau_i, t)$ includes two components : the multivariate function $h(b_i, t)$ of the random effects derived from (IV.1) that models the dependency between the PSA dynamics and the time-to-clinical-recurrence, and information about initiation of HT. In the following, we propose five specifications of $W_i(b_i, \tau_i, t)$ that differ in the way the initiation of HT enters into the model, and different variants that correspond to different functions of the PSA dynamics $h(b_i, t)$. The five specifications are :

1. $W_i(b_i, \tau_i, t)^T \varphi = \varphi_1 \mathbb{1}_{\{t \geq \tau_i\}}$. This is the standard survival model assuming there is no association between the PSA dynamics and the risk of event, but considering a change of risk of recurrence after initiation of HT.
2. $W_i(b_i, \tau_i, t)^T \varphi = \varphi_1 \mathbb{1}_{\{t \geq \tau_i\}} + h(b_i, t)^T \varphi_2$. This is the standard joint model for describing PSA dynamics and risk of recurrence (Proust-Lima et al., 2008; Kennedy et al., 2010; Taylor et al., 2013a; Taylor et al., 2013b). This model assumes that the

characteristics of the PSA dynamics have the same role before and after HT. After HT, these characteristics are extrapolated as if the patient did not initiate HT (see Figure 1). So that the change in risk after HT captured by parameter φ_1 summarizes the effect of HT adjusted for base PSA dynamics.

3. $W_i(b_i, \tau_i, t)^T \varphi = \varphi_1 \mathbb{1}_{\{t \geq \tau_i\}} + h(b_i, t)^T \varphi_2 \mathbb{1}_{\{t < \tau_i\}} + h(b_i, t)^T \varphi_3 \mathbb{1}_{\{t \geq \tau_i\}}$. Rather than assuming that the effect of PSA dynamics is the same before and after HT, this model considers a different effect of the base PSA dynamics before and after HT.
4. $W_i(b_i, \tau_i, t)^T \varphi = \varphi_1 \mathbb{1}_{\{t \geq \tau_i\}} + h(b_i, t)^T \varphi_2 \mathbb{1}_{\{t < \tau_i\}} + h(b_i, \tau_i)^T \varphi_3 \mathbb{1}_{\{t \geq \tau_i\}}$. This model is similar to model 3, except that after HT the extrapolated value of PSA is fixed at the current value at time τ_i rather than being allowed to change according to the base PSA dynamics model.
5. $W_i(b_i, \tau_i, t)^T \varphi = \mathbb{1}_{\{t \geq \tau_i\}} (\varphi_1 + \alpha g(t - \tau_i)) + h(b_i, t)^T \varphi_2 \mathbb{1}_{\{t < \tau_i\}} + h(b_i, \tau_i)^T \varphi_3 \mathbb{1}_{\{t \geq \tau_i\}}$. For specification, the risk of recurrence after HT may change with time after the initiation of HT according to a function $g(t - \tau_i)$. We considered both $g(t - \tau_i) = \log(t - \tau_i)$ and $g(t - \tau_i) = t - \tau_i$.

In specifications 2 to 5, up to three variants of $h(b_i, t)$ were considered (Sène et al., 2013).

- (a) $h_a(b_i, t) = (Y_i^*(t), \partial Y_i^*(t)/\partial t)^T$: the level and slope of PSA at time t are independent predictors of the time to clinical recurrence.
- (b) $h_b(b_i, t) = (\Gamma(Y_i^*(t)), \partial Y_i^*(t)/\partial t)^T$: instead of the crude PSA level, a transformed PSA level $\Gamma(Y_i^*(t))$ and the slope at time t are independent predictors of the time to clinical recurrence, with $\Gamma(Y_i^*(t)) = \text{logit}^{-1}((Y_i^*(t) - 0.71)/0.44)$ (Proust-Lima et al., 2008).
- (c) $h_c(b_i, t) = (b_{0i}, b_{1i}, b_{2i})^T$: the individual deviations from the mean PSA dynamics, that are the random effects, are independent predictors of the time to clinical recurrence. This variant was only considered with specification 2.

2.5.3 Estimation and goodness-of-fit of the joint models

Estimation of the different joint models is summarized in Table IV.1 and parameter estimates that measure both the effect of HT and the association between the PSA dynamics and the risk of clinical recurrence are shown in Table IV.2.

Whatever the assumed nature of the dependence between the PSA dynamics and the time-to-clinical recurrence, the joint models provided a substantial gain in fit compared to model M1 which assumes independence between the two processes (minimum gain of 435.9 points of AIC for the joint model M2_a).

Among the different joint models, considering a logistic transformation of the current level of PSA (models b) rather than the crude current level (models a) improved the fit. In previous work, a residual analysis (Sène et al., 2013) had noted a departure of the log-linearity assumption when considering the crude PSA level in the survival model, and the correction of this departure when considering the logistic transformation. This transformation that makes the effect of the PSA level increase in the range 0 to 4ng/ml and become maximal around 4ng/ml is particularly of importance in M2 where after initiation of HT, very high levels of PSA can be extrapolated from the longitudinal model (as illustrated in Figure IV.1), which may artificially increase the subsequent risk of recurrence.

Assuming that the effects of the crude current PSA level and the current slope differed before and after HT in M3_a greatly improved the fit (121.1 points of AIC) compared to M2_a. In contrast, when considering the logistic transformation instead of the crude PSA value, assuming different effects of PSA dynamics before and after HT in M3_b provided only a small gain in fit (5.4 points of the AIC) compared to M2_b. Indeed, after HT, most of the extrapolated PSA levels are very high so that they drive the estimate to a smaller overall impact of the current PSA level. When separating the effects pre and post-HT in M3_a, the pre-HT crude effect (defined from relatively standard PSA levels) was four

times bigger than the overall crude effect estimated in M2_a, and the effect post-HT was no longer significant. In contrast, when assuming a transformation of the PSA level, the overall effect in M2_b was similar to the effect pre-HT in M3_b. We noted the same things for models M4 and M5 compared to the models M2.

We observed from Table IV.2 that only the slope of PSA was significantly predictive of the risk of recurrence after HT with relatively stable estimates ranging from 0.94 to 1.29 in models M3 through M5. Neither the extrapolated current level in models M3 (with $p = 0.38$ for M3_a and $p = 0.42$ for M3_b) or the level reached at the time of initiation of HT in models M4 and M5 (with $p = 0.31$ and $p = 0.09$ for M4_a and M4_b; $p = 0.35$ and $p = 0.11$ for M5.1_a and M5.1_b; $p = 0.40$ and $p = 0.12$ for M5.2_a and M5.2_b) were associated with the risk of recurrence post-HT after adjustment for the slope of PSA.

Assuming a dependence through the random effects (M2_c) rather than the PSA level or slope provided a goodness-of-fit in between models M2_a and M2_b even though the dependence was summarized by three parameters (all significant) instead of two. Finally, assuming a non constant change in the baseline risk function after HT (in models M5.1 and M5.2) did not improve substantially the fit of the models. In summary, the model M4_b assuming an association with the transformed PSA level, separating effects of PSA prior and after HT, and focusing on characteristics at the time of HT after the initiation, provided the best fit of the data.

Regarding the specific effect of initiation of HT, the interpretation differs between models. Model M2_a aims at capturing the actual protective effect of HT after adjustment for the base PSA trajectory ($\varphi_1 = -1.89$, $p < 0.0001$). But as explained before, this model may suffer from the very high extrapolated PSA values after HT so that M2_b may be more appropriated to accurately evaluate the effect of HT with an estimate $\varphi_1 = -1.39$ ($p < 0.0001$). This corresponds to a relative reduction by 4 in the risk of recurrence when initiating HT and adjusted for the PSA characteristics.

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

In models M3 to M5, no single parameter represents the effect of HT, and particularly parameter φ_1 no longer represent the effect of HT and should not be interpreted as such. Indeed, distinct effects of PSA dynamics before and after HT are modeled so that (except for standard prognostic factors) the model is stratified on the initiation of HT and parameter φ_1 associated with the initiation of HT only represents a change in the baseline risk at HT initiation. This baseline risk appears to be substantially increased (e.g. $\varphi_1=1.33$ in M3_a and $\varphi_1=2.74$ in M3_b) but this has to be put in balance with the different effects of PSA level and slope before and after HT, PSA level being highly significant before initiation of HT and no longer significant after initiation of HT.

Table IV.1 – Goodness-of-fit statistics.

Model	L	AIC	# param.
1	-13549.4	27184.7	43
2.a	-13329.4	26748.8	45
2.b	-13222.9	26535.8	45
2.c	-13261.6	26615.1	46
3.a	-13266.8	26627.7	47
3.b	-13218.2	26530.4	47
4.a	-13265.5	26625.1	47
4.b	-13214.8	26523.5	47
5.1.a [†]	-13264.0	26624.0	48
5.1.b [†]	-13213.9	26523.7	48
5.2.a [‡]	-13263.2	26622.4	48
5.2.b [‡]	-13213.4	26522.7	48

[†] For this model, we assume a change in baseline risk after HT with the function $g(t - \tau_i) = t - \tau_i$ which corresponds to a Gompertz hazard function.

[‡] For this model, we assume that $g(t - \tau_i) = \log(t - \tau_i)$ which corresponds to a Weibull hazard function.

Table IV.2 – Parameters estimates (and standard error (se)) of the HT and the association between the PSA dynamics and the risk of clinical recurrence adjusted on the prognostic factors.

Parameters Model	HT		Before HT					After HT	
	$\hat{\varphi}_1$	(se)	Level $\hat{\varphi}_{21}$	Slope $\hat{\varphi}_{22}$	Random Effects $\hat{\varphi}_{21}$	Random Effects $\hat{\varphi}_{22}$	Random Effects $\hat{\varphi}_{23}$	Level $\hat{\varphi}_{31}$	Slope $\hat{\varphi}_{32}$
1.	0.16	(0.17)							
2.a	<u>-1.89</u>	<u>(0.25)</u>	<u>0.13</u>	<u>2.44</u>					
2.b	<u>-1.39</u>	<u>(0.17)</u>	<u>4.82</u>	<u>1.10</u>					
2.c	<u>-2.56</u>	<u>(0.22)</u>			<u>0.92</u>	<u>-0.31</u>	<u>3.70</u>		
3.a	1.33	(0.45)	<u>0.62</u>	<u>1.56</u>				-0.05	<u>1.29</u>
3.b	2.74	(1.28)	<u>4.77</u>	<u>1.19</u>				1.13	<u>0.95</u>
4.a	1.20	(0.46)	<u>0.64</u>	<u>1.50</u>				0.15	<u>1.10</u>
4.b	2.17	(1.01)	<u>4.77</u>	<u>1.18</u>				1.90	<u>0.94</u>
5.1.a	1.33	(0.47)	<u>0.62</u>	<u>1.55</u>				0.14	<u>1.20</u>
5.1.b	2.31	(1.01)	<u>4.77</u>	<u>1.20</u>				1.84	<u>0.96</u>
5.2.a	1.35	(0.47)	<u>0.62</u>	<u>1.56</u>				0.12	<u>1.23</u>
5.2.b	2.36	(1.01)	<u>4.71</u>	<u>1.22</u>				1.74	<u>1.00</u>

bold underlined : highly significant ($p < 0.001$) ; **bold** : significant ($0.001 \leq p \leq 0.05$) ; nonbold : not significant ($p > 0.05$).

2.5.4 Predictive accuracy of the joint models

For the comparison of the joint models in terms of predictive accuracy, we focused on 6 models : the model assuming independence between the PSA dynamics and the risk of clinical recurrence (M1), the standard joint models in PSA studies (M2_a and M2_b) (Taylor et al., 2013a), the joint models in which the extrapolated PSA current level and slope after HT are replaced by the PSA level and slope at initiation of HT (M4_a and M4_b) and the model assuming directly a dependence on the random effects (M2_c).

The predictive accuracy was only evaluated on the estimation data (2386 men) using the approximated cross-validated estimates. We chose to assess the ability of the joint models to predict the risk of clinical recurrence in a window of 3 years ($\mathcal{T}=3$) which was a clinically reasonable window. For all the measures, the lower the better.

a) Average predictive accuracy after immediate initiation of ST

Among men who initiated a HT during the follow-up, the average predictive accuracy POL and BS defined in section 2.4.3 is shown in Figures IV.2 (a) and IV.2 (c) for the 6 predictive models. The differences between pairs of models and their 95% confidence intervals were also computed and shown in Figures IV.2 (b) and IV.2 (d).

First, BS and POL measures mostly agreed even if a few differences were observed between the three or four most predictive models.

Whatever the nature of the dependency between the PSA dynamics and the risk of recurrence, the predictive accuracies of joint models were significantly better than those of model M1, which assumes independence between PSA dynamics and risk of recurrence.

In accordance with the goodness-of-fit measures, considering different effects prior to HT and after HT improved a lot more the predictive accuracy when the crude PSA level was considered (model M4_a compared to M2_a) than when considering a

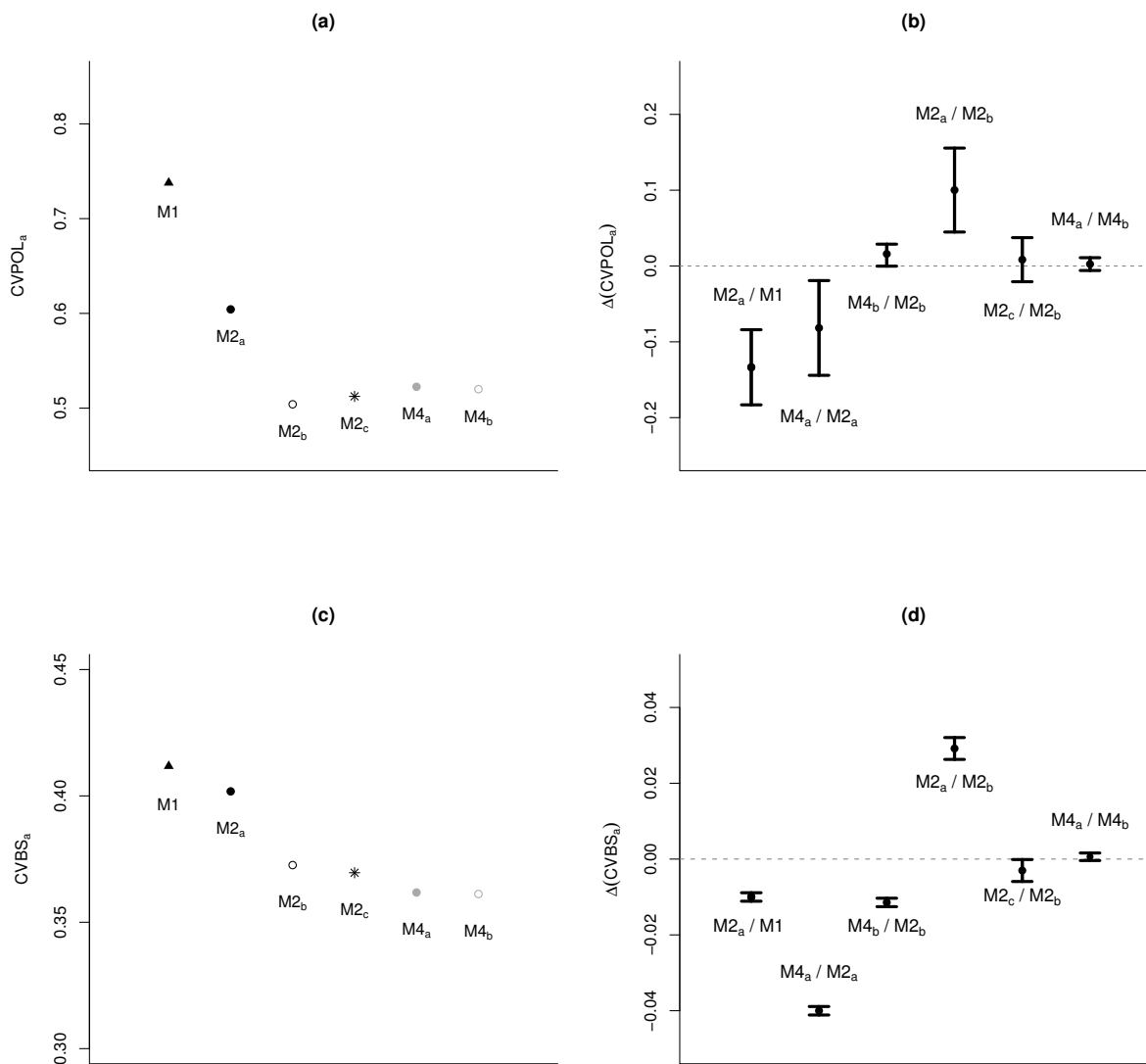


Figure IV.2 – Predictive accuracy measures after an immediate initiation of ST averaged over the times of ST initiation for 6 joint models : with (a) POL estimate, (b) difference in POL and 95% C.I., (c) BS estimate, (d) difference in BS and 95% C.I. Negative (respectively positive) differences indicate the first model has a better (respectively worse) predictive ability.

transformed PSA level (model $M4_b$ compared to $M2_b$). The latter comparison is the only one with discordance between POL and BS results : BS concluded that predictive ability of $M4_b$ was significantly better than the one of $M2_b$ while no difference was found with POL.

Among models M2, considering a transformation of the PSA current level rather than the crude PSA level improved significantly the predictive accuracy ($M2_b$ compared to $M2_a$) while among models M4, this did not induce any significant difference for either measure between $M4_b$ and $M4_a$. Finally, assuming a dependence on the random effects ($M2_c$) rather than on the PSA transformed level and slope ($M2_b$) did not alter much the ability to predict the risk of recurrence.

In summary, BS tended to favor model $M4_b$ while POL tended to slightly favor model $M2_b$. However, as the difference in POL between $M4_b$ and $M2_b$ was not significant, we chose $M4_b$ as the final best model to predict clinical recurrence in the next 3 years after immediate initiation of HT.

b) Predictive accuracy in absence of ST

To evaluate the predictive accuracy of the joint models in absence of HT initiation in the next 3 years, predictive accuracy measures were computed at different times of prediction s (from 1 to 6 years after end of EBRT) among men who did not initiate any HT in the window $[s, s + 3]$ years. These curves are displayed in Figures [IV.3 \(a\)](#), [IV.3 \(c\)](#) and [IV.3 \(e\)](#) for the approximated cross-validation estimates of prognostic cross-entropy, Brier score and Integrated Brier score. The corresponding differences between pairs of models and their 95% confidence bands were computed at the same times of prediction and shown in Figures [IV.3 \(b\)](#), [IV.3 \(d\)](#) and [IV.3 \(f\)](#).

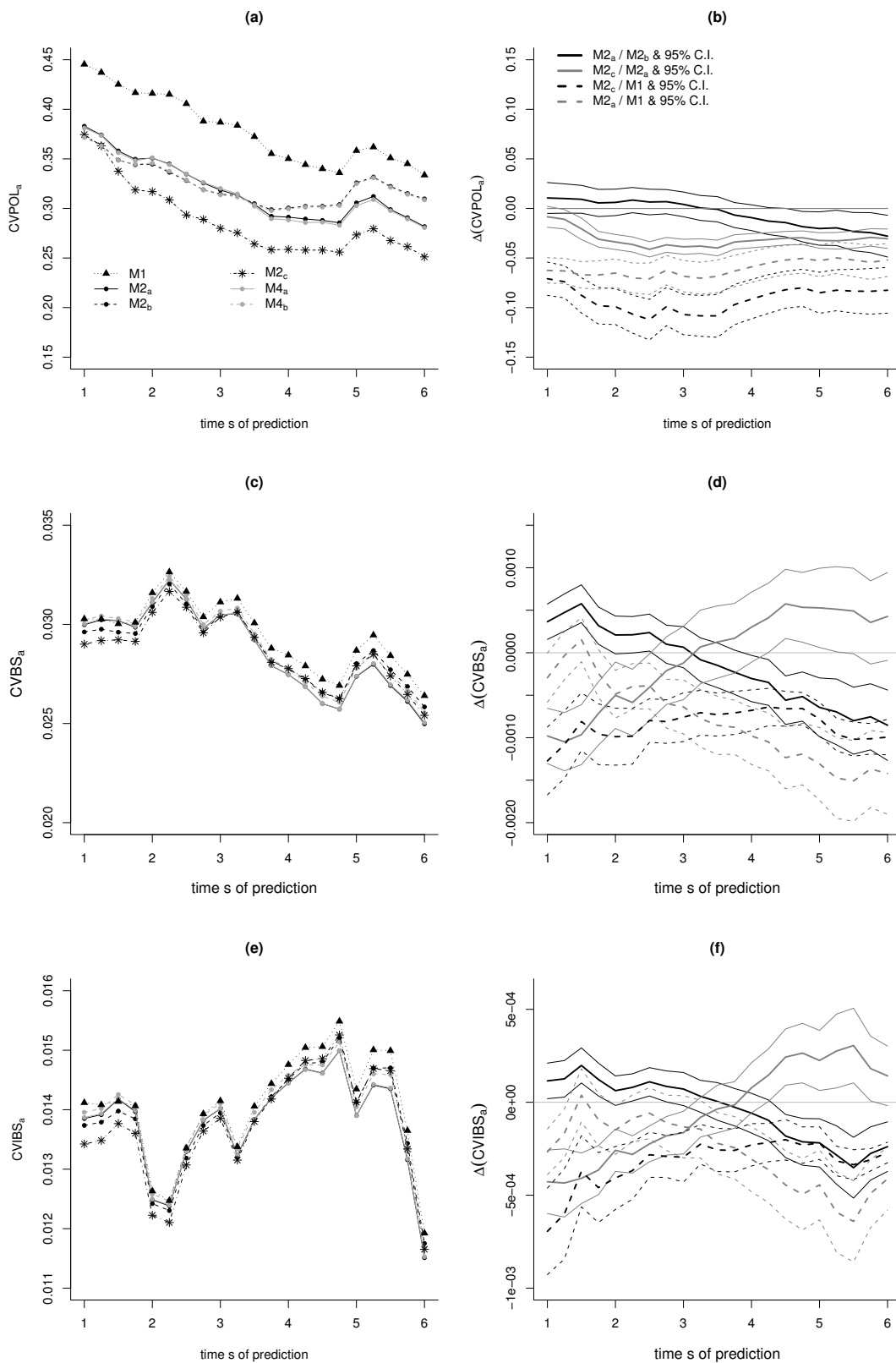


Figure IV.3 – Predictive accuracy measures in absence of ST for 6 joint models at times from 1 to 6 after EBRT with (a) EPOCE estimate, (b) difference in EPOCE and 95% C.I., (c) BS estimate, (d) difference in BS and 95% C.I., (e) IBS estimate and (f) difference in IBS and 95% C.I. Negative (positive) differences indicate the first model has a better (worse) predictive ability.

Whatever the predictive accuracy measure and the nature of the dependency between the PSA dynamics and the risk of recurrence, the joint models provided globally a significantly better predictive accuracy compared to model M1, which assumes independence between the two processes (with the surprising exception for M4_b in the first years according to the BS and IBS measures).

Whatever the predictive accuracy measure, in the absence of HT, models M4 and M2 had very similar predictive performances (differences not shown). Indeed, the overall estimates in M2 are mostly driven by the high proportion of subjects who did not initiate HT.

Whatever the measure, considering in models M2 a logistic transformation of the PSA level (M2_b) instead of the crude PSA level (M2_a) did not really improve the predictive performances in the short-term for $s \in [1, 4]$. This was expected as the transformation of PSA level is supposed to mainly correct very high extrapolated PSA values not observed among subjects who will not initiate any HT. However, in the long-term ($s \geq 4$ years) joint models considering the crude PSA level (M2_a) provided even a significantly better predictive accuracy. This contrasted with conclusions in terms of goodness-of-fit or after HT initiation where specification b was systematically better.

When considering a dependence on the random effects (M2_c) rather than on the PSA crude level and slope (M2_a), conclusions based on BS, IBS and POL measures differed : M2_c was found largely better than M2_a at times of prediction greater than 1.5years with POL and was also found better with BS and IBS but only for shorter times of prediction. At longer times of prediction, model M2_a was even slightly better with BS and IBS.

Although results differed substantially depending on the type of measure we used

(POL or BS/IBS), among men who did not undergo any HT, the joint model assuming directly a dependence between the PSA and the risk of clinical recurrence through the random effects (M2_c) provided a nice alternative to the more standard (M2_a) joint model. We note that these two models that are the most predictive in absence of HT are also the ones in which the effects of the PSA long-term slope are the highest. This was previously observed in (Sène et al., 2013), where among patients who did not initiate any HT, joint models having the largest effects of the slope of log PSA were also the ones having the best predictive ability suggesting that after a few years, the slope of log PSA would be the major predictor of the risk of recurrence in the absence of HT.

In summary, while the best model to predict the risk of clinical recurrence assuming an immediate initiation of HT was M4_b (based on transformed PSA level and crude PSA slope), the best model we chose to predict the risk of recurrence assuming the patient will not initiate any HT within the next 3 years was M2_c that directly focused on the random-effects as shared PSA characteristics.

c) Example of differential dynamic prediction of prostate cancer recurrence

To illustrate the use of these dynamic predictions of prostate cancer recurrence, we considered a subject who had a T-stage of 2, a Gleason of 6, an initial PSA of 12.7 ng/ml a corrected dose of radiation of 65.7 Gy and who recurred at 2.7 years after the end of EBRT. After each PSA measurement, we computed his individual predicted probability of clinical recurrence in the next 3 years under the two assumptions : whether he initiates HT immediately (probabilities computed according to model M4_b) or whether he does not initiate any HT in the next 3 years (probabilities computed according to model M2_c). These two individual predictions are shown in Figure IV.4. This shows the reduction of the risk of clinical recurrence that would be obtained by initiating HT.

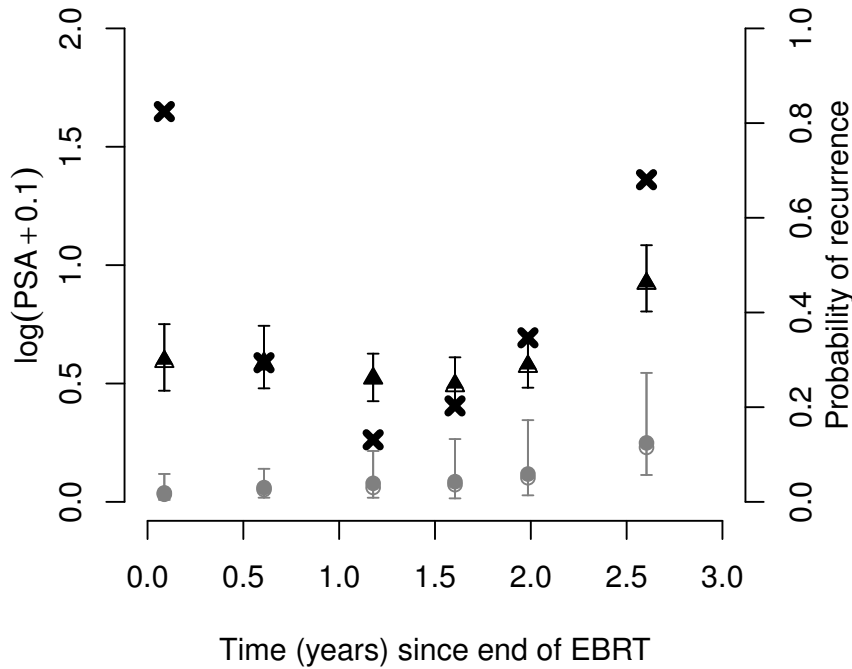


Figure IV.4 – Individual predicted probabilities of clinical recurrence within 3 years assuming either immediate initiation of HT (grey) or no initiation of HT in the next 3 years (black). Repeated PSA measurement are denoted \times . After each new measurement, the prediction within 3 years is computed at the vector of parameters estimates $\hat{\theta}$ (unfilled black triangle and unfilled grey circle), and the distribution is approximated by a Monte Carlo method (solid black triangle and solid grey circle indicate the median and the intervals indicate the 95% bands).

2.6 Discussion

We proposed to use joint models to develop and validate personalized dynamic prognostic tools. They allow prediction of any event of interest according to a biomarker trajectory by assuming different scenarios for the initiation of treatment. We focused on two scenarios : the prediction of the event assuming no change in the treatment, and the prediction of the event assuming the immediate initiation of a second treatment, which is very common in the monitoring of chronic diseases such as cancers.

Although very useful in practice, such differential dynamic predictive tools were never developed or validated. Usually, dynamic predictive tools only predict the risk of event based on a biomarker value ([Rizopoulos, 2011b](#)) or a biomarker trajectory ([Proust-Lima and Taylor, 2009a](#); [Taylor et al., 2013a](#)) by assuming that there is no change in treatment that might impact the subsequent risk of event. This was for example the case until now in prostate cancer where predictions were developed and validated by neglecting the possible initiation of HT ([Proust-Lima and Taylor, 2009a](#); [Taylor et al., 2013a](#)). Yet evaluating formally the gain in terms of risk reduction of a second treatment initiation could be of great interest for monitored patients and their clinicians to aid in the application of personalized medicine.

Even in absence of HT, the prognostic tool validation is not straightforward. We chose to focus here on patients who did not initiate any HT in the window of prediction. However these patients may not be a representative sample of the patients free of HT at the time of prediction. [Taylor et al. \(2013a\)](#) proposed instead to validate the prognostic tools by focusing on the sample of subjects free of HT at the time of prediction and by considering either all the HT initiations during the window of prediction as recurrences or as censoring. The results concerning the relative performance of the models did not change when using this technique as shown in Web supplementary materials.

We defined different dynamic predictions for two distinct scenarios and validated them separately. Indeed, it is perhaps unrealistic to expect that the same model provides the best predictions in different situations. In the application, we did find that two different joint models that differed in the way that initiation of HT and association with the biomarker were modeled provided the best predictive tools among patients who did not initiate any HT and among patients who immediately initiated HT. Even if in our application, all the joint models had a relatively good predictive accuracy in both situations, this highlights the importance of validating the predictive tools strictly for what they are aimed to quantify in practice.

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

Throughout the paper, we based the dynamic predictive tools development on joint models with shared random-effect and compared different specifications for the nature of the dependency between the biomarker dynamics and the risk of event. We note however that exactly the same methodology could have been applied with other approaches like joint latent class models or landmark analyses in the spirit of Proust-Lima and Taylor ([Proust-Lima and Taylor, 2009a](#)).

Evaluation of the joint models predictiveness was done according to two measures of predictive accuracy, the prognostic cross-entropy and the Brier score, on the same data as used for estimating the models. We did not consider any external validation. We favored using as large a sample as possible for estimation because of the relative complexity of the models and the differential validation procedure.

The prognostic cross-entropy derived from information theory and estimated by the conditional log-likelihood ([Commenges et al., 2012](#)) was developed to assess the prognostic value of joint models, and an approximated cross-validation estimate made it possible to directly evaluate such prognostic value on the same data as used for parameter estimation. Originally this measure evaluated the prognostic value on the remaining (infinite) window of horizons. We chose here to target a finite window of time (3 years in the application).

The second measure, the Brier Score, was developed for standard survival models ([Graf et al., 1999](#)) and extended ([Schoop et al., 2011](#)) to evaluate the predictiveness of dynamic predictive tools. Yet, its use was mostly limited to external validation. By using the same approximated leave-one-out cross-validation formula, we were able to provide an estimate of the Brier score which was very easy to compute on the estimation data. This estimate assumes that the inverse probability weights involved in the BS computation and derived from the data were independent which was not strictly evaluated. This will be done in a future work. However, it should not have much impact on the estimates.

We found in the application differences between the conclusions given by the two measures. This may be explained by the fact that the measures do not tackle the predictive accuracy in the same way. The Brier Score directly measures the Mean Square Error between the event process and the prediction of the model while the prognostic cross-entropy measures the distance between the conditional density of the time to event assumed in the model with the true one. We still chose to select the best models as a balance between the results given by these two measures rather than one only.

Another type of measure may be used in predictive model validation, the AUC ([Heagerty and Zheng, 2005](#)). We chose to not consider this. Indeed, first it focuses on discrimination while our focus was really on predictiveness since we wanted to quantify individual probabilities of recurrence. Second, its use in dynamic settings has been rather limited ([Rizopoulos, 2011b](#); [Zheng and Heagerty, 2007](#)). Third, providing an approximated cross-validation estimate was not straightforward.

In the application, the models providing the best goodness-of-fit did not necessarily have the best predictive ability. This illustrates the difference between these two types of assessment. While the goodness-of-fit measures use all the information available, the predictive accuracy measures focus on a part of the sample (the subjects at risk at time s) and use only the history of the biomarker up to the time of prediction. This is why when interested in dynamic predictions, the predictive ability of the models should be assessed ([Sène et al., 2013](#)).

In this work we compared many joint models in order to obtain the best dynamic prognostic tool. An alternative approach ([Rizopoulos et al., 2013](#)) consists in combining the predictions from different joint models using Bayesian model averaging in order to improve the predictive accuracy of predictions and obtain a more powerful model. This idea could also be investigated with our different dynamic predictions.

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

Finally, in prostate cancer, the initiation of a second treatment, namely hormonal therapy, has raised many questions about how to take it into account in the model for risk of clinical recurrence or how to evaluate its causal effect in the presence of indication bias (Taylor et al., 2013b). In the present paper, our focus was only on the dynamic individual predictions. As such, we chose to compare descriptive joint models that treated HT intuitively as a time-dependent covariate, possibly in interaction with other characteristics. However, using the same strategy, more causal or mechanistic models could also be investigated.

3 Analyse supplémentaire

Dans ce travail, nous avons comme objectif de proposer un outil pronostique dynamique dans deux situations : (1) si le sujet a une initiation immédiate d'un traitement hormonal au temps de prédiction s , (2) si le sujet n'initie pas de traitement dans les 3 années suivant le temps de prédiction s (entre s et $s+3$). Nous avons donc évalué les outils pronostiques dans ces deux cas de figure. Cependant, en l'absence de traitement hormonal dans la fenêtre de prédiction décrite dans (2), l'échantillon des sujets qui n'initient pas d'hormonothérapie pendant toute la fenêtre de prédiction n'est pas nécessairement représentatif de l'ensemble des sujets pour qui une initiation d'hormonothérapie n'est pas prévue au temps de prédiction s . Afin de pallier ce possible problème, nous avons réévalué les performances prédictives des outils pronostiques dans le scénario (2) sur l'échantillon de tous les sujets n'ayant pas reçu de second traitement au temps de prédiction s . Le problème est que parmi eux, certains ont reçu une hormonothérapie avant la fin de la fenêtre de prédiction et que cela a changé leur risque de rechute. Dans cette analyse de sensibilité, nous avons donc traité les initiations d'hormonothérapie après le temps s de façon extrême, soit comme des censures systématiquement soit comme des rechutes cliniques systématiquement, comme cela avait été proposé par [Taylor et al. \(2013a\)](#).

Pour ces deux scénarios extrêmes, l'entropie croisée pronostique (voir Figure IV.5 (a) et Figure IV.6 (a)) et le score de Brier (voir Figure IV.5 (c) et Figure IV.6 (c)) ont été recalculés à des temps de prédiction s allant de 1 à 6 années après la fin de la radiothérapie externe et avec les mêmes modèles conjoints présentés dans la Figure 3 de l'article. Des différences de pouvoir prédictif entre deux modèles et leurs intervalles de confiance à 95% ont également été calculés pour l'entropie ((voir Figure IV.5 (b) et Figure IV.6 (b))) et pour le score de Brier (voir Figure IV.5 (d) et Figure IV.6 (d)).

Que l'initiation de l'hormonothérapie soit considérée comme une censure ou comme un événement, la hiérarchie des différents modèles conjoints est restée globalement inchangée

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

en terme de qualité prédictive ce qui permet de conforter les résultats de l'article.

Quelle que soit la mesure de qualité prédictive utilisée, on remarquera que les valeurs des deux mesures sont plus proches de celles présentées dans l'article lorsqu'on considère l'hormonothérapie comme une censure. En effet, lorsque l'initiation de l'hormonothérapie est considérée comme une censure, le nombre d'événements dans la population à risque ne change pas et à chaque temps de prédiction la prévalence de la population ne bouge pas beaucoup par rapport à la prévalence de la population de l'article. Au contraire, dans le cas où l'initiation de l'hormonothérapie est considérée comme un événement, la prévalence diffère beaucoup de celle de l'article ce qui fait que l'entropie croisée et le score de Brier diffèrent également puisque, ces deux mesures sont liées à la prévalence de l'événement dans l'échantillon .

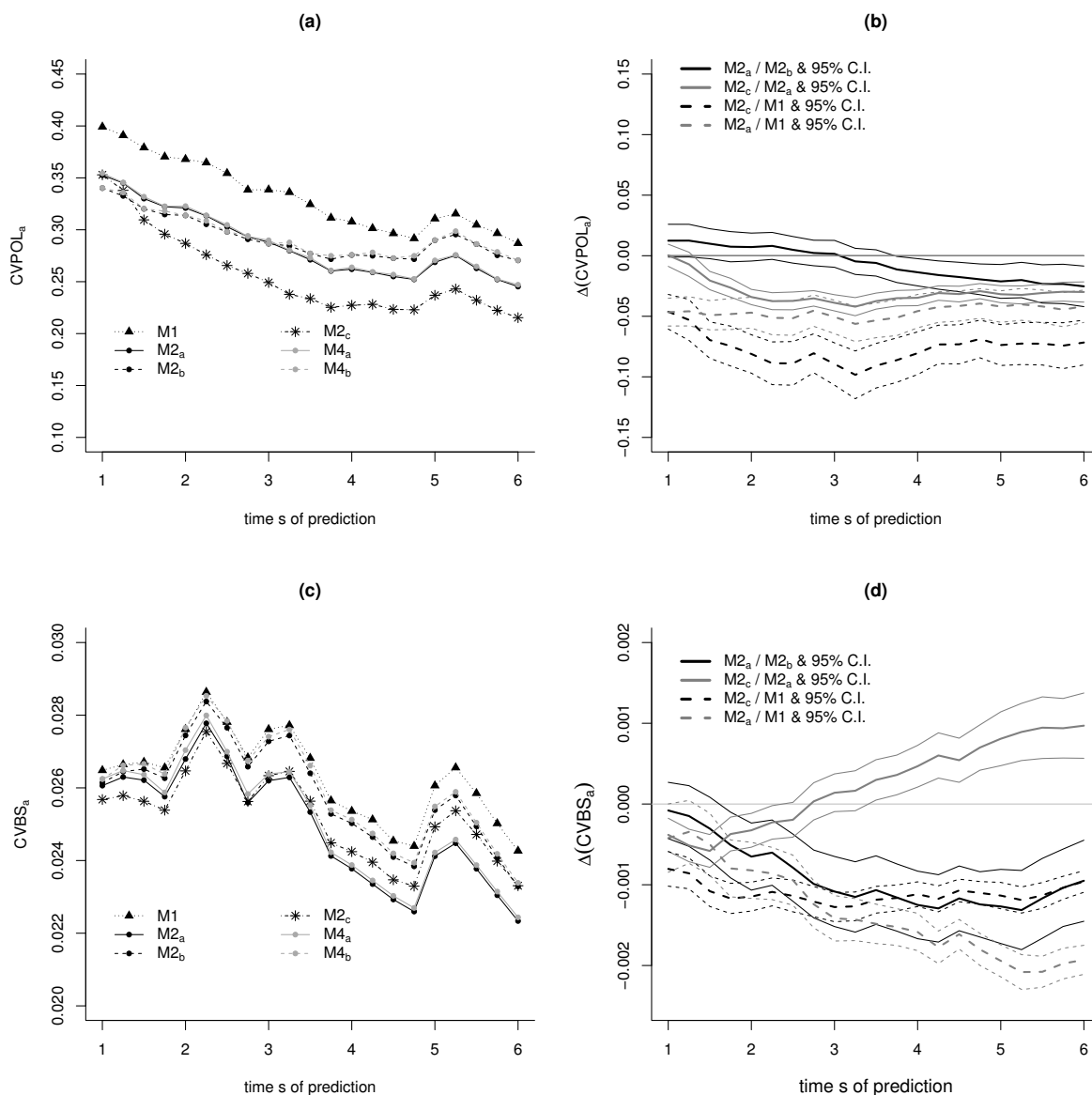


Figure IV.5 – Mesures de pouvoir prédictif si l'initiation de l'hormonothérapie est considérée comme une censure.

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

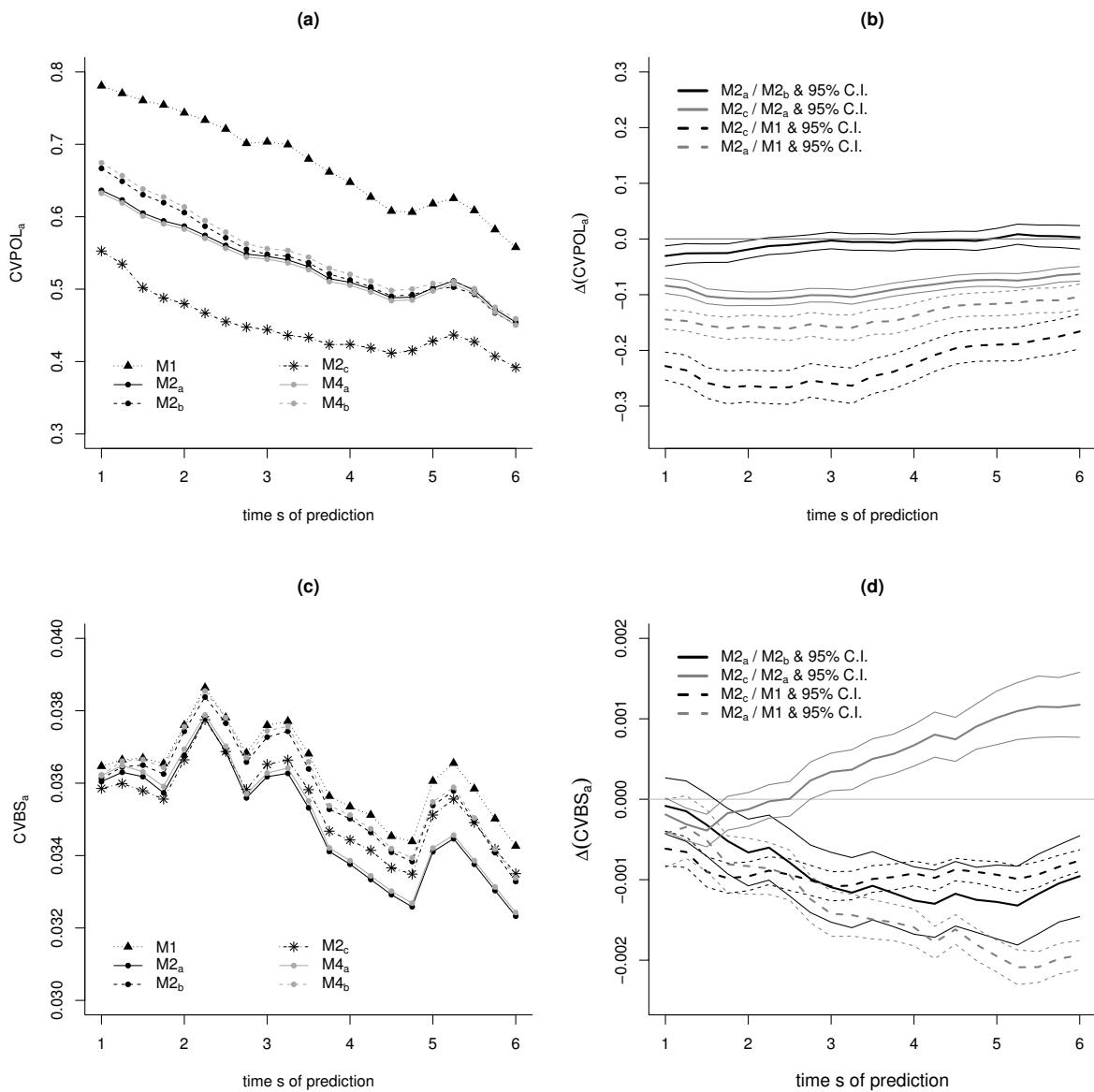


Figure IV.6 – Mesures de pouvoir prédictif si l'initiation de l'hormonothérapie est considérée comme une rechute clinique.

4 Exemples de prédiction individuelle dynamique avec tous les modèles conjoints

Nous avons présenté, dans cet article, un exemple de prédiction individuelle dynamique obtenue avec le meilleur modèle conjoint en terme de pouvoir prédictif dans le cas d'une initiation immédiate de l'hormonothérapie ($M4_b$) et dans le cas où le patient n'initie aucun traitement dans les 3 prochaines années ($M2_c$). Nous montrons ici les prédictions individuelles obtenues par les autres modèles conjoints présentés dans l'article. La Figure IV.7 (a) présente les prédictions, après chaque mesure de PSA, si le patient reçoit un traitement hormonal et s'il ne le reçoit pas (Figure IV.7 (b)).

Après une initiation immédiate de l'hormonothérapie (Figure IV.7 (a)), toutes les prédictions sont extrêmement proches les unes des autres. Dans le cas d'aucune initiation de traitement, les prédictions issues du modèles $M2_c$ sont très différentes des autres en début de suivi. Nous avons néanmoins décider de considérer ces prédictions puisque ce modèle donnerait aussi de meilleure performance prédictive que les autres modèles en tout début de suivi. Cette différence peut aussi s'expliquer par le fait que, comparée aux autres spécifications, la dépendance à travers les effets aléatoires était très différente dans la conception et l'interprétation, prenant directement en compte les déviations individuelles sur le déclin post-radiothérapie pour prédire le risque de rechute.

Chapitre IV. Prédiction du risque de rechute de cancer de la prostate après un premier traitement par radiothérapie selon l'initiation d'un second traitement au cours du suivi

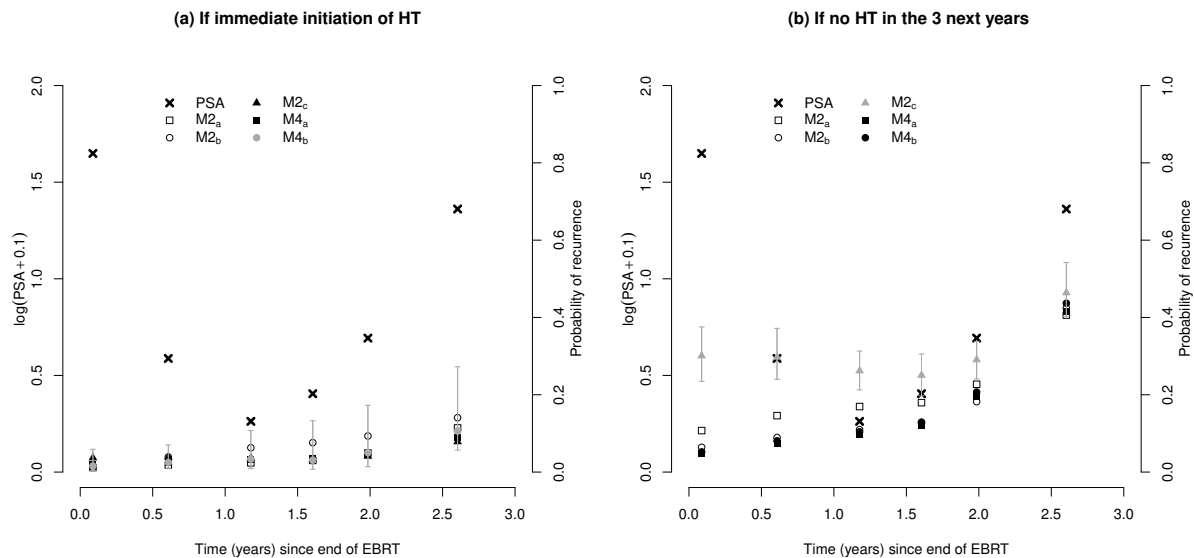


Figure IV.7 – Probabilités prédites de rechute clinique avec les différents modèles conjoints si le patient a une initiation immédiate de l'hormonothérapie (a) et s'il ne reçoit aucun second traitement dans les 3 prochaines années (b).

5 Conclusion

L'objectif de ce chapitre était de proposer un outil pronostique dynamique différentiel. Cet outil permet de prédire le risque de rechute clinique, à tout temps, si le patient reçoit ou non un second traitement. Nous avons évalué la qualité prédictive des différents modèles suivant deux situations, ce qui n'a jamais été fait auparavant. Lors de la validation de l'outil chez des sujets qui n'ont pas reçu de second traitement dans la fenêtre de prédiction, les sujets ne sont pas forcément représentatifs de l'ensemble des sujets pour qui l'initiation de l'hormonothérapie n'est pas prévue au temps de prédiction s, ce qui nous a amené à faire une analyse de sensibilité en considérant l'initiation de l'hormonothérapie comme une censure ou comme un événement.

L'évaluation de la qualité prédictive à travers le score de Brier ou avec l'entropie croisée pronostique a montré des résultats qui vont globalement dans le même sens que notre

analyse principale. Ces résultats montrent que la validation de l'outil chez les sujets n'ayant pas reçu de second traitement dans la fenêtre de prédiction représente une bonne stratégie.

Chez les sujets ayant une initiation immédiate de l'hormonothérapie, le fait de stratifier à l'initiation du traitement, dans le modèle conjoint, sur avant et après le traitement améliore la qualité prédictive du modèle.

Chez les sujets qui n'ont pas reçu le traitement hormonal, le modèle conjoint avec une dépendance directement à travers les effets aléatoires représente le meilleur modèle prédictif. Ce résultat a également été trouvé dans le premier article de cette thèse présenté dans le chapitre III dans lequel l'application concernait des sujets ayant été traités uniquement par radiothérapie et n'ayant pas initié de traitement hormonal au cours du suivi.

Chapitre V

Dynamique des PSA après l'initiation d'une hormonothérapie chez des patients traités initialement par radiothérapie pour un cancer de la prostate localisé

Dans ce chapitre, nous ne nous sommes plus directement intéressés à la prédiction dynamique de rechute de cancer de la prostate. Nous avons voulu décrire la dynamique des PSA après un traitement hormonal initié au cours du suivi chez des patients atteints de cancer de la prostate localisé et traités initialement par radiothérapie externe. En effet, cette dynamique est extrêmement peu décrite dans la littérature.

1 Introduction

Comme évoqué dans le chapitre précédent, après un traitement par radiothérapie chez des patients atteints de cancer de la prostate, les patients sont suivis régulièrement avec des tests sanguins pour doser le taux de PSA dans le sang. Après le traitement, on observe une chute du taux de PSA dans la première année suivie d'une phase de stabilisation ou de croissance sur le long-terme. Une augmentation importante du taux de PSA sur le long-terme peut être indicatrice d'une rechute biochimique ([Roach III et al., 2006](#) ; [Bellera et al., 2009](#)).

Chapitre V. Dynamique des PSA après l'initiation d'une hormonothérapie chez des patients traités initialement par radiothérapie pour un cancer de la prostate localisé

Si la croissance est suggestive d'une possible rechute, le patient peut être mis sous traitement hormonal pour prévenir ou retarder la rechute. La dynamique du PSA après la radiothérapie est bien connue et a été décrite par [Pauler and Finkelstein \(2002\)](#), [Proust-Lima et al. \(2008\)](#) et [Bellera et al. \(2008a\)](#). [Proust-Lima et al. \(2008\)](#) ont montré que la dose de radiothérapie ainsi que les facteurs pronostiques standards tels que le PSA pré-radiothérapie, le stade du cancer et le score de Gleason avaient un effet sur cette dynamique post-radiothérapie.

Si la dynamique après la radiothérapie (RT) est bien connue, celle après une hormonothérapie (HT) initiée au cours du suivi, n'a jamais été décrite et reste inconnue même si le PSA est encore mesuré après ce second traitement. Comme illustré sur la Figure I.2 du chapitre I, le nouveau traitement entraîne un changement dans la dynamique du PSA si bien que dans les études longitudinales sur la progression du cancer après une RT, les mesures de PSA faites après le traitement hormonal sont censurées ([Proust-Lima et al., 2008](#); [Proust-Lima and Taylor, 2009a](#); [Kennedy et al., 2010](#); [Taylor et al., 2013a](#)). Dans l'analyse présentée dans le chapitre précédent (Chapitre IV), nous avons également censuré les données de PSA post-HT puisque la prédiction dynamique n'était envisagée qu'avant la possible initiation de l'hormonothérapie.

Dans la littérature même si quelques études ont été faites autour de l'initiation de ce second traitement hormonal, aucune n'a été consacrée à décrire la dynamique des PSA. On peut citer en exemple [Miralbell et al. \(2009\)](#) et [Valicenti et al. \(2011\)](#) qui ont évalué l'efficacité de la combinaison de la radiothérapie et de l'hormonothérapie sur la survie du cancer. [D'Amico et al. \(2004b\)](#) ont montré que le ratio de la pente post-HT sur la pente pre-HT était associé au temps jusqu'au décès par cancer chez les patients traités initialement par radiothérapie ou par chirurgie. [Payne et al. \(2012\)](#) ont étudié le "timing" pour l'initiation de HT sur la réussite du traitement et recommandent une initiation immédiate du traitement chez les patients avec une rechute métastatique ou une cinétique défavorable

à la rechute. Et pour finir, on peut citer [Crook et al. \(2012\)](#) qui évaluent la survie globale avec une HT intermittente versus continue après un traitement initial par radiothérapie. Ils montrent que le traitement intermittent n'est pas inférieur au traitement continu et que certains facteurs de la qualité de vie des patients sont améliorés avec le traitement intermittent.

Dans ce contexte, l'objectif de cette analyse était d'abord de décrire la dynamique du PSA post-HT chez des patients traités initialement par radiothérapie externe puis d'évaluer l'impact des facteurs pronostiques standards sur cette dynamique. Cette analyse s'est donc focalisée sur les données de PSA qui avaient été censurées dans les chapitres précédents.

2 Analyses statistiques

Les données répétées du PSA ont été analysées à travers un modèle linéaire mixte (voir section 1.1 du chapitre II). Le logarithme des données de PSA a été analysé afin de satisfaire l'hypothèse de normalité du modèle.

Pour mieux comprendre la dynamique des PSA après l'initiation de l'hormonothérapie, nous avons considéré plusieurs formes paramétriques de trajectoire :

- Une trajectoire en 2 phases, comme pour la dynamique des PSA post-RT, qui considère un déclin sur le court-terme suivant une fonction de type $f(t; \alpha) = (1 + t)^\alpha - 1$ (dans laquelle α est estimé par la technique du profil de maximum de vraisemblance) et une tendance linéaire sur le long-terme. Le modèle est alors donné par :

$$\log(\text{PSA}_i(t) + 0.1) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) f(t; \alpha) + (\beta_2 + b_{i2}) t + w_i(t) + \epsilon_i(t) \quad (\text{V.1})$$

où $(\beta_0, \beta_1, \beta_2)^T$ est le vecteur des effets fixes, $(b_{i0}, b_{i1}, b_{i2})^T$ est le vecteur des effets

Chapitre V. Dynamique des PSA après l'initiation d'une hormonothérapie chez des patients traités initialement par radiothérapie pour un cancer de la prostate localisé

aléatoires Gaussiens centré et de matrice de variance-covariance non structurée; les $\epsilon_i(t)$ sont des erreurs de mesure Gaussiennes et indépendantes et $(w_i(t))$ est un processus Gaussien auto-régressif d'ordre 1 ($\text{corr}(w_i(t_{ij}), w_i(t_{ik})) = \sigma_w^2 e^{\rho|t_{ij}-t_{ik}|}$).

- Plusieurs alternatives au modèle (V.1) ont été investiguées. Au lieu d'avoir une évolution linéaire sur le long-terme, on a supposé un possible ralentissement de la croissance du PSA en considérant les fonctions $t/\log(t+1)$ ou $(t/\log(t+1), t^2)$ à la place de t , $t/\log(t+1)$ en plus de t ou t^2 en plus de t .
- Des modèles paramétriques plus flexibles utilisant les B-splines et qui pourraient potentiellement capturer les variations locales des PSA ont été aussi investigués. Ils consistent à remplacer $f(t)$ et t dans (V.1) par une base de splines. Nous avons considéré des B-splines quadratiques et cubiques avec des nombre de nœuds compris entre 1 et 3.
- Des polynômes fractionnaires ont été également investigués comme des alternatives plus parcimonieuses aux splines lorsque l'on veut modéliser des trajectoires non linéaires (Long and Ryoo, 2010). Nous avons choisi de nous limiter à des polynômes fractionnaires d'ordre 2.

Ces différentes formes paramétriques ont été considérées dans un modèle linéaire mixte incluant des effets aléatoires sur chacune de ces fonctions du temps afin de prendre en compte la variabilité inter-individuelle. Pour chaque modèle, nous avons aussi considéré soit des erreurs de mesure indépendantes, donc $w_i(t) = 0$, soit un processus auto-régressif en plus des erreurs indépendantes avec $w_i(t)$ afin de permettre plus de souplesse dans la modélisation de la corrélation intra-sujet.

Dans une première étape, aucune covariable n'a été incluse dans les modèles pour définir la forme globale de la dynamique du PSA. Les différents modèles ont été comparés en terme de leur qualité d'ajustement mais également en terme de complexité d'interprétation. Le test de rapport de vraisemblances a été utilisé pour la comparaison de deux modèles imbriqués et le critère AIC a été utilisé pour comparer les modèles non imbriqués.

A partir de la meilleure structure de modèle trouvée dans la première étape, nous avons ensuite investigué l'impact de différents facteurs sur la dynamique du PSA après l'initiation d'une hormonothérapie. Deux types de facteurs ont été considérés. Tout d'abord, nous avons considéré les facteurs pronostiques standards tels que le niveau de PSA pré-radiothérapie, le stade de la tumeur (en 3 catégories : 1, 2, 3-4), le score de Gleason (en 3 catégories : 2-6, 7, 8-10) et la dose totale corrigée de radiothérapie (Proust-Lima et al., 2011). Ensuite, nous avons considéré les caractéristiques de l'initiation de l'hormonothérapie telles que le temps d'initiation, le niveau de PSA prédit ainsi que la pente prédite juste avant l'hormonothérapie. Ces dernières valeurs ont été prédites à partir d'un modèle linéaire mixte ajusté sur les données de PSA pré-hormonothérapie uniquement ; il s'agissait des prédictions individuelles spécifiques aux sujets dans l'idée de celles introduites en chapitre II, section 1.3, calculées à partir des BLUPs des effets aléatoires au temps d'initiation de l'hormonothérapie (Proust-Lima et al., 2008). Comme cette analyse a été réalisée sur les données combinées des trois cohortes de patients, nous avons également mis les indicatrices des cohortes dans le modèle pour prendre en compte les potentielles différences entre les cohortes.

Outre les indicatrices des cohortes qui ont été contraintes dans le modèle, les autres prédicteurs ont été sélectionnés par une procédure de sélection de modèles pas à pas descendante. Le test de Wald a été utilisé pour tester la significativité des covariables binaires et continues, et le test de rapport de vraisemblances pour la significativité des covariables catégorielles avec au moins trois modalités. Nous avons considéré des tests bilatéraux et une valeur de $\alpha = 0,05$ a été choisie pour le risque de première espèce.

Toutes les analyses ont été faites avec le logiciel R et l'estimation des modèles a été faite par la méthode du maximum de vraisemblance avec la fonction *hlme* du package *lcmm*.

3 Sélection des données

Les cohortes utilisées dans ce travail sont les mêmes que celles utilisées pour l'article présenté dans le chapitre IV. Nous avons la cohorte hospitalière UM venant de l'Université de Michigan, l'essai clinique RTOG qui est un essai multi-centrique et la cohorte BM de l'hôpital Beaumont dans le Michigan. Les données ainsi que les critères d'inclusion des patients sont présentés dans (Proust-Lima et al., 2008). Parmi les 2386 patients, 261 (10,9%) ont initié une hormonothérapie au cours de leur suivi mais uniquement 241 sujets ont des mesures de PSA collectées après l'initiation du traitement. Le problème majeur de cette analyse est que les sujets ont pu initier un second traitement hormonal à la suite du premier qui nous intéressait. En effet, ce traitement hormonal n'était souvent pas définitif. Dans la perspective de décrire la dynamique de PSA après une première initiation, il nous a fallu éliminer ces données ultérieures à une seconde initiation. Dans les cohortes UM et RTOG, pour les patients ayant initié un deuxième traitement hormonal au cours du suivi, les mesures de PSA faites après la deuxième HT ont donc été éliminées. Mais dans BM, l'information sur une deuxième HT était manquante, pourtant certains patients avaient des profils d'évolution ressemblant à ceux de sujets ayant reçu une deuxième HT (voir Figure V.1). Donc pour éviter des problèmes de biais de sélection par rapport à ces mesures de PSA après la deuxième HT et pour gérer cette chute du PSA due à une possible seconde HT, nous avons décidé de réduire notre fenêtre d'étude. Nous avons conservé :

- les mesures de PSA faites avant 5 ans. Car au-delà, il était plus vraisemblable que le patient ait initié un second traitement et nous étions principalement plus intéressés par la progression juste après l'initiation du traitement.
- les mesures de PSA précédant un passage au dessus de 3,5 ng/ml (en log). Car au-delà, il était vraisemblable que le patient initie un second traitement.
- les mesures de PSA précédant 2 diminutions successives de PSA ou une unique

diminution de plus de 1 ng/ml (en log) au-delà de 12 mois (c'est-à-dire dans la phase de stabilisation ou de croissance sur le long-terme). Les données à partir de 2 diminutions successives ou après la diminution de plus de 1 ng/ml étaient donc toutes éliminées.

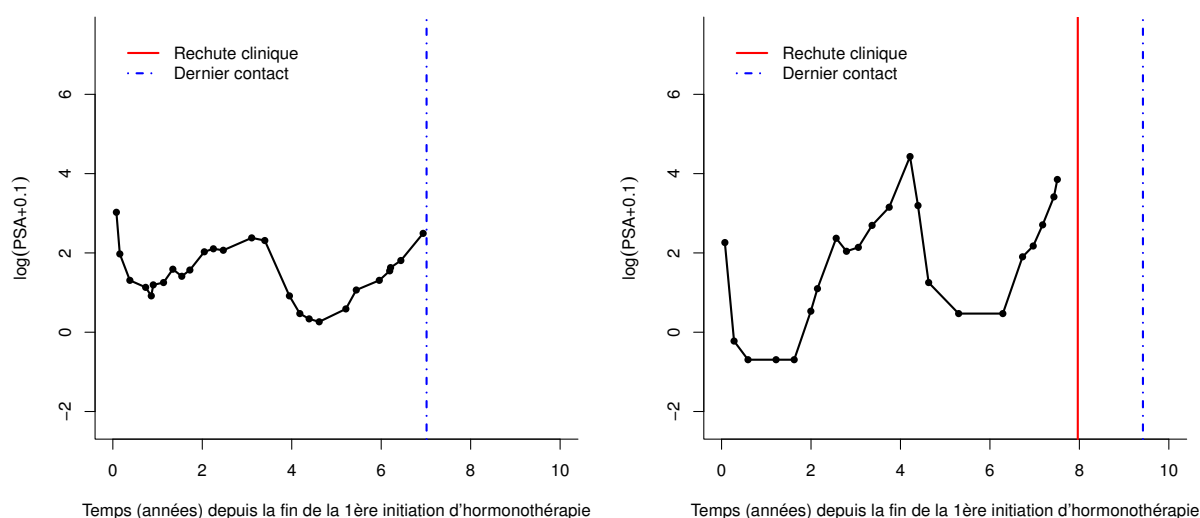


Figure V.1 – Exemples de deux patients de BM dont les profils d'évolution du PSA ressemblent à ceux de patients ayant reçu une seconde hormonothérapie.

4 Résultats

4.1 Description de l'échantillon

Cette sélection a abouti à un échantillon de 238 sujets et la description des données est donnée dans la Table V.1. Au total, 1103 mesures répétées de PSA ont été analysées à partir de l'initiation de l'hormonothérapie avec un nombre médian de 4 mesures par patient et un maximum de 23 mesures. Le traitement hormonal a été initié à un temps médian de 4,2 années après la fin du traitement initial par RT. Les patients avaient un niveau médian de PSA à l'initiation de l'HT de 2,1 (en log ng/ml) et une pente médiane

Chapitre V. Dynamique des PSA après l'initiation d'une hormonothérapie chez des patients traités initialement par radiothérapie pour un cancer de la prostate localisé

de 0,6 (en log ng/ml/année) correspondant à un temps de doublement de PSA égal à 1,2 années. Le temps médian de suivi était de 3,9 années. Des différences significatives ont été notées entre les cohortes. Les sujets issus de BM étaient plus âgés ($p < 0,001$), avaient un niveau de PSA pre-RT plus élevé ($p = 0,003$), et avaient initié l'hormonothérapie plus tardivement avec un temps médian de 4,5 années. Les patients issus de RTOG avaient reçu des doses de radiothérapie plus élevées ($p < 0,001$), avaient un stade clinique et un score de Gleason plus faibles ($p < 0,001$).

Table V.1 – Description des données sélectionnées des différents cohortes.

Cohorte	UM	RTOG	BM	Combinées
Période de recrutement	1988-2004	1994-2001	1987-2003	
Nombre de sujets	39	39	160	238
Nombre de mesures	227	96	780	1103
Nombre de mesures/patient	4 (1; 14)	2 (1; 5)	4 (1; 11)	4 (1; 12)
Age (Années)	66,6 (60,2; 76,2)	71,0 (57,8; 77,2)	74,0 (63,0; 90,0)	73,0 (61,0; 81,0)
log(PSA pre-RT + 0.1) (log ng/ml)	3,0 (1,1; 4,3)	2,2 (1,2; 3,2)	2,7 (1,5; 4,5)	2,6 (1,3; 4,4)
Stade clinique # (%)				
1	6 (15,4)	18 (46,2)	28 (17,5)	52 (21,8)
2	26 (66,7)	18 (46,2)	116 (72,5)	160 (67,2)
3-4	7 (17,9)	3 (7,6)	16 (10,0)	26 (10,9)
Score de Gleason # (%)				
2 - 6	20 (51,3)	26 (66,7)	94 (58,7)	140 (58,8)
7	14 (35,9)	9 (23,1)	36 (22,5)	59 (24,8)
8 - 10	5 (12,8)	4 (10,2)	30 (18,8)	39 (16,4)
Dose totale RT (Gy)	69,0 (65,7; 77,1)	77,0 (69,5; 83,0)	66,6 (66,0; 70,2)	68,4 (66,0; 80,0)
Dose corrigée RT avec un ratio $\alpha/\beta = 3$ (Gy)	67,2 (63,6; 75,2)	75,5 (67,1; 83,8)	63,9 (63,9; 67,4)	65,7 (63,9; 79,8)
Niveau de PSA pre-HT prédit (log ng/mL) [†]	2,1 (-0,2; 3,9)	1,6 (0,1; 2,9)	2,2 (0,5; 4,4)	2,1 (0,1; 4,0)
Pente de PSA pre-HT prédite (log ng/mL/année) [†]	0,7 (0,1; 1,3)	0,6 (0,1; 1,6)	0,6 (0,2; 1,4)	0,6 (0,1; 1,4)
Temps jusqu'à l'initiation de HT	3,7 (1,3; 9,3)	3,3 (1,5; 7,1)	4,5 (1,8; 8,3)	4,2 (1,6; 8,3)
Nb. rechutes # (%)	8 (20,5)	6 (15,4)	28 (17,5)	42 (17,6)
Temps de rechute clinique depuis la fin de HT (Années)	1,4 (0,8; 4,7)	1,6 (0,5; 4,3)	2,7 (0,8; 7,0)	2,2 (0,6; 7,0)
Temps de dernier contact (Années) [‡]	5,2 (1,2; 8,9)	3,4 (0,8; 6,5)	3,9 (0,6; 8,9)	3,9 (0,6; 8,8)

Les covariables continues sont présentées de la manière suivante : médiane (5th ; 95th) percentiles ; les covariables catégorielles : nombre (pourcentage).

[†] valeurs prédites obtenues à partir des données de PSA pré-hormonothérapie dans un modèle linéaire mixte incluant les facteurs pronostiques ayant un impact sur la dynamique des PSA pré-hormonothérapie (Proust-Lima et al., 2008).

[‡] calculé chez les patients qui n'ont pas observé la rechute clinique.

Chapitre V. Dynamique des PSA après l'initiation d'une hormonothérapie chez des patients traités initialement par radiothérapie pour un cancer de la prostate localisé

La Figure V.2 présente les trajectoires du PSA pour tous les sujets. Nous pouvons remarquer que les trajectoires individuelles sont relativement variables, mais une dynamique en 2 phases apparaît, comme après la RT, avec un déclin sur le court-terme et une stabilisation ou croissance sur le long-terme.

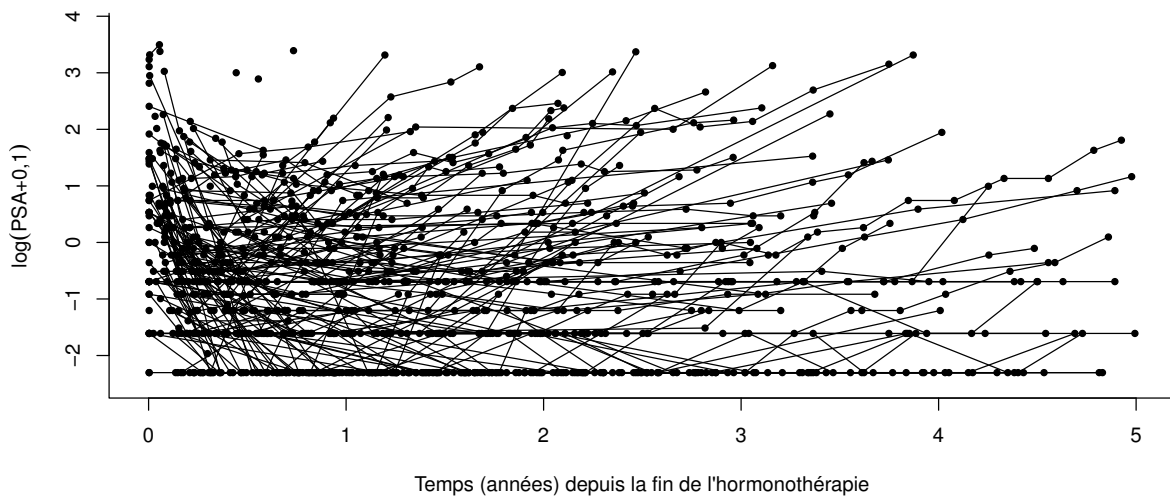


Figure V.2 – Trajectoires individuelles du $\log(\text{PSA} + 0,1)$ après l'hormonothérapie

4.2 Dynamique globale du PSA

Pour déterminer la dynamique globale du PSA, nous avons estimé les différents modèles sans aucune variable d'ajustement. Nous avons estimé chaque modèle soit avec des erreurs de mesures indépendantes soit avec un processus d'auto-corrélation en plus. Les résultats sont présentés dans la Table [V.2](#).

Tout d'abord, quelle que soit la forme supposée de la trajectoire du PSA, l'inclusion d'un processus d'auto-corrélation a considérablement amélioré la qualité d'ajustement du modèle. Ce résultat était attendu étant donnée la fréquence des mesures.

Ensuite, lorsqu'on suppose une trajectoire en 2 phases avec $f(t; \alpha)$ et t et des erreurs indépendantes, la technique du profil de vraisemblance donne $\alpha = -2,5$ avec un intervalle de confiance à 95% de $[-3,1; -2,0]$; ce qui est inférieur à la valeur de α trouvée après la radiothérapie ($\alpha = -1,5$). Lorsque l'on suppose un processus d'auto-corrélation en plus, on obtient $\alpha = 5,5$ avec un intervalle de confiance de $[-6,9; -4,3]$ ce qui indique une décroissance plus rapide sur le court-terme.

Lorsque l'on relâche la tendance linéaire sur le long-terme pour prendre en compte une possible stabilisation du PSA en utilisant $t/\log(t+1)$, (t, t^2) , $(t/\log(t+1), t^2)$ ou $(t, t/\log(t+1))$, la qualité d'ajustement, l'AIC, est améliorée par rapport au modèle avec 2 phases $\{f(t; \alpha), t\}$. En considérant des polynômes fractionnaires, l'ajustement n'a pas été amélioré comparé au modèle $\{f(t; \alpha), t\}$ avec ou sans le processus d'auto-corrélation. En supposant une trajectoire approximée par des fonctions splines, comparé au modèle $\{f(t; -5,5), t\}$ avec le processus d'auto-corrélation, l'ajustement a été amélioré avec les splines cubiques mais ne l'a pas été avec des splines quadratiques. Comparé au modèle $\{f(t; -2,5), t\}$ avec les erreurs indépendantes, l'ajustement des modèles avec les splines cubiques ou quadratiques est amélioré. Mais, ces trajectoires nécessitent beaucoup de paramètres à estimer et sont très sensibles aux variations locales, ce qui les rend difficiles

Chapitre V. Dynamique des PSA après l'initiation d'une hormonothérapie chez des patients traités initialement par radiothérapie pour un cancer de la prostate localisé

à utiliser pour interpréter la dynamique des PSA de façon générale.

Table V.2 – Qualité d'ajustement des différents modèles.

Forme de la trajectoire	Erreurs indépendantes			Erreurs ind. + AR(1)			
	p	L	AIC	p	L	AIC	
2 phases :		$\alpha = -2, 5$			$\alpha = -5, 5$		
$\{f(t; \alpha), t\}$	10	-1176,22	2372,44	12	-1121,52	2267,03	
$\{f(t; \alpha), t/\log(t+1)\}$	10	-1174,28	2368,56	12	-1119,69	2263,38	
$\{f(t; \alpha), (t, t^2)\}$	15	-1158,74	2347,48	17	-1110,24	2254,48	
$\{f(t; \alpha), (t/\log(t+1), t^2)\}$	15	-1158,19	2346,38	17	-1109,45	2252,90	
$\{f(t; \alpha), (t, t/\log(t+1))\}$	15	-1155,98	2341,96	17	-1106,21	2246,42	
Alternatives plus flexibles :							
Polynômes fractionnaires [‡]	10	-1185,50	2391,00	12	-1131,76	2287,52	
Splines quadratiques	28	-1133,76	2323,52	30	-1121,30	2302,59	
Splines cubiques	36	-1088,59	2249,18	38	-1079,56	2235,13	

p est le nombre de paramètres du modèle et L est la log-vraisemblance.

[‡] En supposant des polynômes fractionnaires d'ordre 2, nous avons obtenu les puissances 0,5 et 0,5 par profil de vraisemblance.

La Figure V.3 présente les trajectoires moyennes prédites pour chaque modèle. On remarque une évolution lente sur le long-terme, comme après RT. Si sur le long-terme les prédictions sont éloignées pour certains modèles, elles restent très proches sur le court-terme. Le nadir est atteint aux alentours de 6 à 8 mois.

A partir de là, nous avons choisi de conserver le modèle avec la trajectoire en 2 phases avec $f(t; -5, 5)$ sur le court-terme, une évolution linéaire sur le long-terme et avec le processus d'auto-corrélation comme le meilleur modèle. En effet, ce modèle constitue un très bon compromis entre la qualité d'ajustement aux données et une facilité d'interprétation de la dynamique du PSA, notamment auprès des cliniciens.

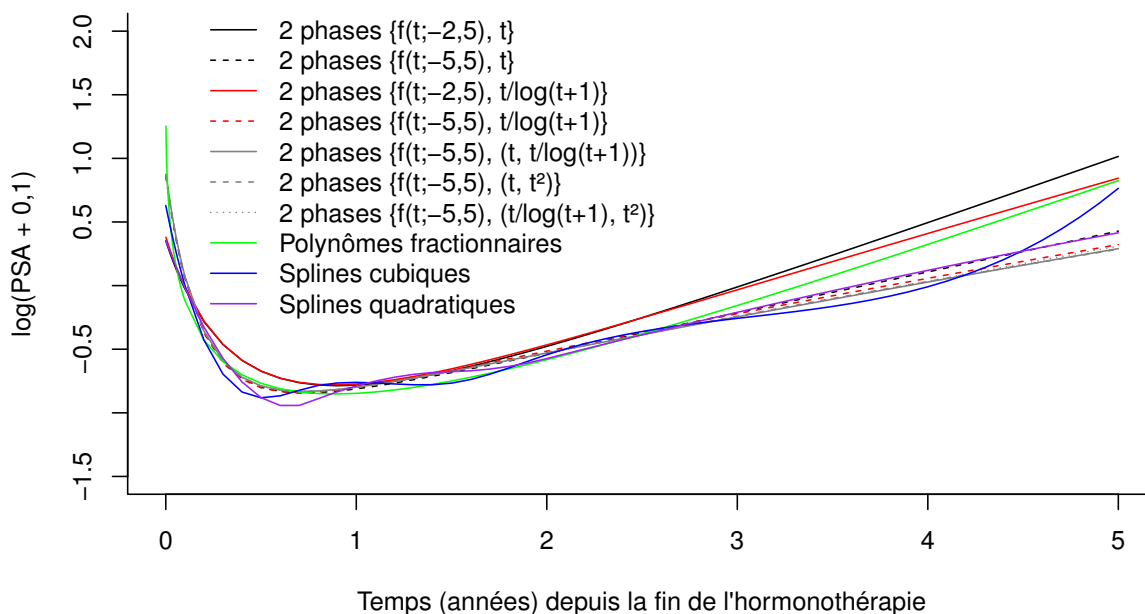


Figure V.3 – Évolutions moyennes prédites du $\log(\text{PSA} + 0,1)$ pour les trajectoires en 2 phases avec $\{f(t; -2, 5), t\}$ et $\{f(t; -2, 5), t/\log(t+1)\}$ et des erreurs indépendantes et pour les trajectoires avec des erreurs indépendantes en plus du processus d'auto-corrélation pour les autres courbes.

4.3 Facteurs associés à la dynamique du PSA

Les facteurs ayant un impact sur la dynamique du PSA post-HT ont été investigués à partir du modèle qui suppose une trajectoire en 2 phases, le processus d'auto-corrélation en plus des erreurs indépendantes. Cela correspond à un déclin sur le court-terme modélisé par $f(t, -6.2)$ et une tendance linéaire sur le long-terme. Ce modèle correspond au modèle défini en (V.1) dans lequel $\beta_0 = \alpha_0 X + \mu_0 Q$, $\beta_1 = \alpha_1 X + \mu_1 Q$ et $\beta_2 = \alpha_2 X + \mu_2 Q$ avec X les facteurs pronostiques investigués et Q les différentes cohortes. Le meilleur ajustement a été choisi en utilisant la procédure de sélection de modèle pas à pas descendante et les résultats sont présentés dans la Table V.3. Après ajustement sur la cohorte (effet forcé dans le modèle), seuls le niveau de PSA pré-radiothérapie, le niveau et la pente de PSA précédant l'initiation de l'HT avaient un impact sur la dynamique des PSA post-HT.

Chapitre V. Dynamique des PSA après l'initiation d'une hormonothérapie chez des patients traités initialement par radiothérapie pour un cancer de la prostate localisé

Table V.3 – Facteurs pronostiques sur les différentes phases d'évolution pour la modèle en 2 phases $f(t, 5, 5)$ et une tendance linéaire sur le long-terme ajusté sur les trois cohortes (UM, RTOG, BM).

Facteurs pronostiques	Estimation	Erreur-Standard	P-value
<u>Effet sur le niveau de PSA post-HT :</u>			
Intercept*	0,14	0,38	0,7154
PSA pre-RT	-0,23	0,09	0,0134
Niveau de PSA prédit à l'HT	0,27	0,09	0,0044
Pente de PSA prédite à l'HT	1,20	0,41	0,0037
<u>Effet sur la phase de déclin :</u>			
Intercept [†]	1,38	0,42	0,0011
Pente de PSA prédite à l'HT [‡]	1,48	0,50	0,0028
<u>Effet sur la phase de croissance :</u>			
Intercept	0,31	0,15	0,0297
Pente de PSA prédite à l'HT	0,42	0,17	0,0151

* correspond à la valeur moyenne du $\log(\text{PSA} + 0,1)$ post-HT pour un sujet avec un niveau de PSA pre-RT, un niveau de PSA précédant l'HT et une pente précédant l'HT nuls. L'effet est moyenné sur les cohortes.

[†] correspond à la valeur moyenne du $\log(\text{PSA} + 0,1)$ sur le court-terme pour un sujet avec une pente nulle du taux de PSA précédant l'initiation de l'HT ajusté sur les cohortes. L'effet est moyenné sur les cohortes.

[‡] correspond au déclin du niveau de PSA sur le court-terme pour une augmentation d'une unité de la pente de PSA précédant l'initiation de l'HT. L'effet est moyenné sur les cohortes.

Le niveau de PSA post-HT était entièrement prédit par ces trois facteurs avec une augmentation du niveau ou de la pente du PSA précédant l'HT logiquement associée à une augmentation du niveau de PSA post-HT, et une augmentation du niveau de PSA pré-radiothérapie associée étonnamment à une diminution du niveau de PSA post-HT (voir Figure V.4, Figure V.5 et Figure V.6). Néanmoins, ceci pourrait être expliqué par le fait qu'il s'agit de l'effet ajusté sur les autres caractéristiques pre-HT. Une autre explication pourrait être que les médecins vont avoir tendance à initier l'hormonothérapie plus rapidement chez les patients avec des niveaux de PSA pre-RT plus élevés.

Les phases de déclin du PSA sur le court-terme et de croissante ou de stabilisation sur le long-terme étaient prédites uniquement par la pente de PSA précédant l'HT avec une augmentation de la pente associée à un déclin plus important ou à une croissance plus importante des PSA post-HT (voir Figure V.6).

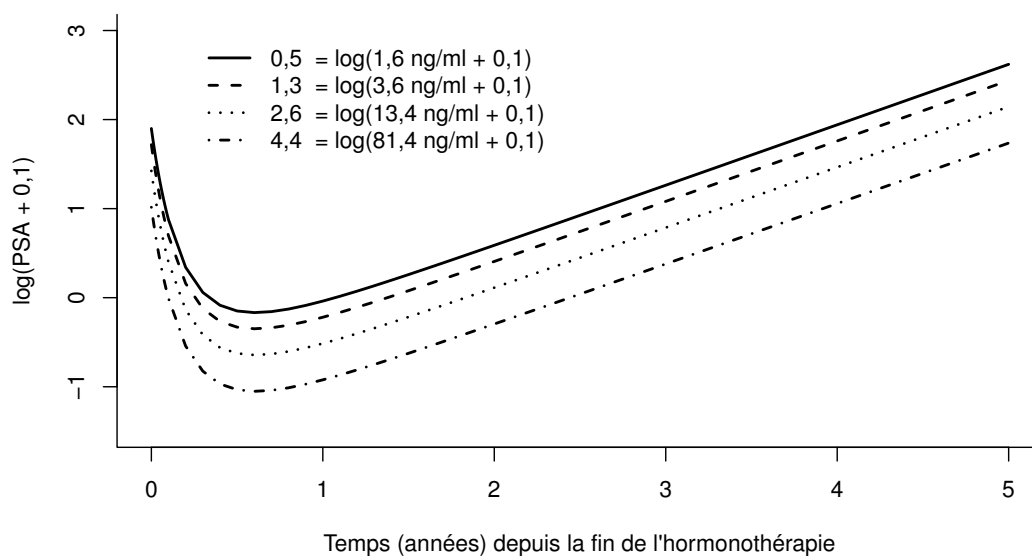


Figure V.4 – Evolution moyenne prédite du $\log(\text{PSA} + 0,1)$ après l'HT suivant le niveau de PSA initial, au diagnostic, pour un patient avec un niveau de PSA précédant l'initiation de l'HT égal à 3 ng/ml (en log) et une pente égale à 0,69.

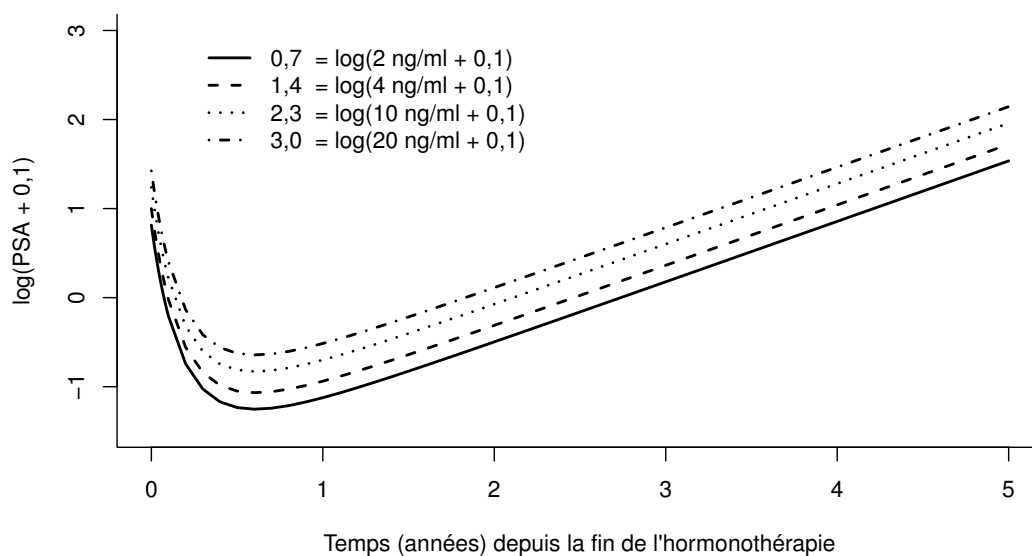


Figure V.5 – Evolution moyenne prédite du $\log(\text{PSA} + 0,1)$ après l'HT suivant le niveau de PSA précédant l'initiation de l'HT pour un patient avec un niveau de PSA au diagnostic égal à 2,6 ng/ml (en log) et une pente de PSA précédant l'initiation de l'HT égale à 0,69.

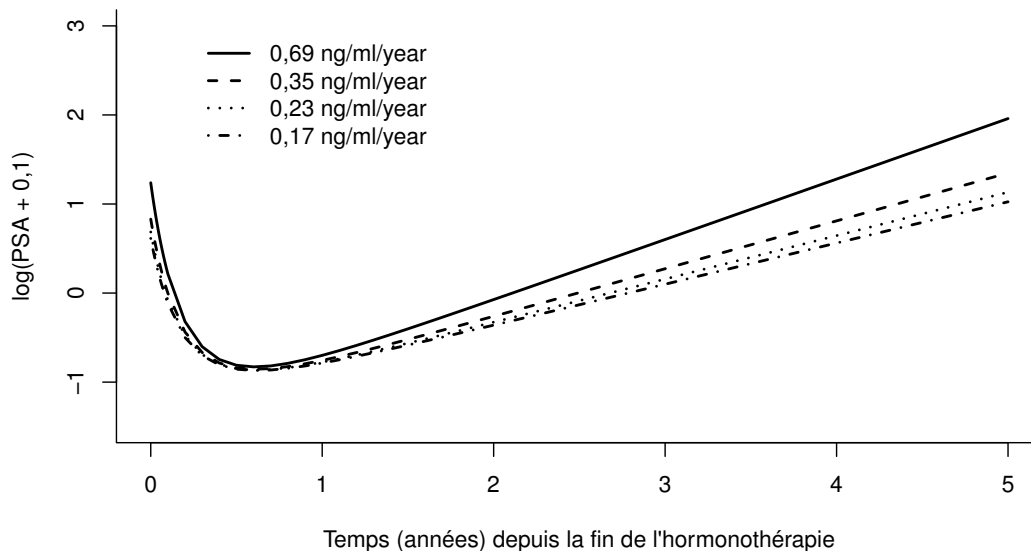


Figure V.6 – Evolution moyenne prédite du $\log(\text{PSA} + 0,1)$ suivant la pente de PSA précédant l'initiation de l'HT pour un patient avec un niveau de PSA au diagnostic égal à 2,6 ng/ml (en log) et un niveau de PSA précédant l'initiation de l'HT égal à 2,3 ng/ml (en log). Une pente de 0,69 correspond à un temps de doublement de PSA de 1 année et 0,17 correspond à 4 années.

4.4 Analyse de sensibilité

Une analyse de sensibilité a été réalisée afin d'évaluer la robustesse de nos résultats à la sélection de l'échantillon faite pour éviter les données postérieures à l'initiation non documentée d'une seconde hormonothérapie. Nous avons pour cela estimé le modèle final trouvé dans notre analyse principale et présenté dans la Table V.3 sur d'autres ensembles de données. Les différents ensembles de données considérés sont :

- (a) les données initiales sans sélection particulière exceptée la suppression des mesures de PSA obtenues après une indication d'une seconde hormonothérapie. Cette sélection comporte 241 sujets et 1526 mesures.
- (b) les mesures de PSA faites avant 5 ans et censurées au premier passage au dessus de 3,5 ng/ml. Cette sélection regroupe 238 sujets et 1312 mesures.

(c) la sélection dans (b) + une sélection visuelle réalisée dans la cohorte BM. Cette sélection visuelle consiste à retenir les mesures de PSA, avant l'initiation supposée d'une seconde hormonothérapie, de patients ayant un profil d'évolution de PSA similaire à ceux ayant reçu une seconde hormonothérapie dans les autres cohortes. On obtient 238 sujets et 1236 mesures suite à cette sélection.

(d) la sélection dans (b) + les mesures de PSA précédant 2 diminutions successives de PSA au-delà de 12 mois. On obtient 238 sujets avec 1142 mesures.

Les résultats sont présentés dans la Table V.4. La significativité des covariables va globalement dans le même sens que les résultats obtenus dans l'analyse principale. En effet, quelle que soit la sélection effectuée, la pente de PSA précédant l'HT a un impact sur les différentes phases d'évolution du PSA. De plus les estimations de l'effet de la pente pre-HT sont assez similaires d'un échantillon à l'autre, ce qui confirme nos résultats. Nous notons néanmoins que, le niveau de PSA pre-RT et le niveau de PSA précédant l'HT ne sont plus significatifs sur le niveau de PSA post-HT avec l'ensemble de données (a).

Nous avons aussi reproduit la sélection pas à pas descendante sur les échantillons (b) et (d) et le modèle final reste inchangé.

4.5 Facteurs prédicteurs de rechute clinique après une hormonothérapie

De la même façon que nous avons investigué les facteurs prédicteurs de la dynamique des PSA post-hormonothérapie, nous avons cherché à connaître quels étaient les facteurs associés à une rechute clinique après l'initiation d'une hormonothérapie. Pour cela, nous avons ajusté un modèle de Cox incluant tous les facteurs pronostiques considérés précédemment puis nous avons fait une sélection de variables pas à pas descendante. Seule la pente précédant l'initiation de l'hormonothérapie avait un effet significatif sur la rechute clinique. Stratifié sur les cohortes, l'effet de cette variable était égal à 1,01 (p-value =

Chapitre V. Dynamique des PSA après l'initiation d'une hormonothérapie chez des patients traités initialement par radiothérapie pour un cancer de la prostate localisé

Table V.4 – Estimation des paramètres du modèle final réalisée sur des ensembles de données différents.

Données	(a)			(b)			(c)			(d)		
	Est.	E-S	p	Est.	E-S	p	Est.	E-S	p	Est.	E-S	p
<u>Facteurs pronostiques</u>												
<u>Effet sur le niveau de PSA post-HT :</u>												
Intercept	-0,13	0,37	0,7239	0,13	0,36	0,7209	0,12	0,38	0,7508	0,14	0,37	0,7076
PSA pre-RT	-0,03	0,10	0,7393	-0,19	0,09	0,0286	-0,22	0,09	0,0151	-0,21	0,09	0,0184
Niveau de PSA prédit à l'HT	0,17	0,10	0,0901	0,24	0,09	0,0090	0,26	0,09	0,0045	0,25	0,09	0,0067
Pente de PSA prédite à l'HT	1,12	0,40	0,0053	1,15	0,40	0,0038	1,16	0,42	0,0054	1,15	0,41	0,0047
<u>Effet sur la phase de déclin :</u>												
Intercept	1,33	0,40	0,0009	1,31	0,40	0,0010	1,38	0,42	0,0009	1,35	0,41	0,0010
Pente de PSA prédite à l'HT	1,20	0,45	0,0081	1,35	0,46	0,0034	1,37	0,49	0,0055	1,32	0,48	0,0062
<u>Effet sur la phase de croissance :</u>												
Intercept	0,21	0,11	0,0504	0,21	0,11	0,0482	0,31	0,14	0,0245	0,25	0,13	0,0569
Pente de PSA prédite à l'HT	0,33	0,12	0,0044	0,30	0,12	0,0122	0,33	0,16	0,0429	0,31	0,15	0,0418

Est. = Estimation, E-S= Erreur-Standard et p = p-value.

0,004) tandis qu'il était de 1,07 (p-value = 0,002) dans le modèle non stratifié sur les cohortes.

Comme un facteur pronostique essentiel au risque de rechute clinique est la dynamique courante des PSA, nous avons réitéré l'analyse avec un modèle conjoint à effets aléatoires partagés pur évaluer si l'effet de la pente précédant l'hormonothérapie subsistait. Une fois ajustée sur la dynamique des PSA post-HT (niveau courant et pente courante de PSA) la pente précédant l'initiation de l'hormonothérapie n'était plus significative. Sa p-value est de 0,14 dans le modèle ajusté sur les cohortes et de 0,11 dans le modèle sans ajustement sur les cohortes. De plus, son effet était réduit de près de moitié, 0,5706 et 0,6178 respectivement dans le modèle ajusté ou non sur les cohortes.

5 Conclusion

Bien que l'hormonothérapie soit un traitement souvent donné en cas de suspicion de progression chez les patients atteints d'un cancer de la prostate localisé et traités initialement par radiothérapie, la dynamique du PSA après l'initiation de l'HT n'a jamais été étudiée à notre connaissance. L'analyse proposée dans ce chapitre avait pour objectif de décrire la dynamique du PSA et déterminer les facteurs impactant celle-ci. Nous avons montré que la trajectoire du PSA après l'HT était caractérisée par une évolution biphasique avec un déclin très rapide dans les premiers mois et une évolution linéaire sur le long-terme comme pour la trajectoire après une RT uniquement. Cependant, le déclin sur le court-terme est plus immédiat après l'initiation d'une hormonothérapie qu'après une radiothérapie.

Parmi les facteurs pronostiques standards connus au diagnostic, comme le PSA pre-RT, le stade clinique, le score de Gleason et la dose de radiothérapie, aucun n'a été associé à la dynamique post-HT des PSA à l'exception du niveau de PSA au diagnostic qui n'était cependant associé qu'au niveau de PSA post-HT. Le niveau de PSA et la pente de PSA précédant l'initiation de l'HT qui caractérisaient la dynamique des PSA avant la mise sous traitement étaient également associés au niveau de PSA post-HT. Mais l'unique prédicteur de la dynamique des PSA post-HT s'est révélé être la pente de PSA précédant l'initiation de l'hormonothérapie. Plus la pente de PSA était grande au moment de l'initiation, plus les PSA évoluaient vers une augmentation importante sur le long-terme.

La pente précédant l'initiation de l'HT était aussi l'unique prédicteur de la rechute clinique, une augmentation de cette pente était associée à une augmentation du risque de rechute clinique. Une fois ajustée sur la dynamique des PSA post-HT, la pente précédant l'HT n'était plus significative, la dynamique courante des PSA capture toute l'association avec le risque de rechute clinique.

Ces résultats sont extrêmement intéressants. Ils suggèrent que la réponse à l'initiation de l'HT pourrait donc peut-être améliorée en étant initiée au bon moment.

Au cours de ce travail, diverses difficultés ont été rencontrées. Principalement, nous ne disposons pas des mêmes informations dans les différentes cohortes combinées sur l'initiation ou non d'une seconde HT. Les résultats présentés auraient donc pu être biaisés par une sélection initiale des données. Mais l'analyse de sensibilité a systématiquement conclu globalement aux mêmes résultats, ce qui est extrêmement rassurant. De plus, nous ne connaissons ni le type d'HT (intermittente, continue), ni la durée de ce traitement. Nous ne savions donc pas si les sujets avaient tous reçu le même type d'HT ou non. Enfin, concernant les sujets ayant reçu une seconde HT, le caractère intermittent non connu de ce second traitement a pu nous amener à exclure dans l'analyse des mesures collectées après l'initiation d'un second traitement qui en réalité n'en était pas un. La méconnaissance de toutes ces informations a pu aussi entraîner des biais dans les estimations des modèles. Malgré une analyse sur trois cohortes de patients, le nombre de sujets est relativement limité et la puissance de l'analyse aussi. En particulier, après ajustement sur la dynamique courante des PSA, la pente pre-HT s'est révélée non associée à la survenue de rechute clinique. Etant donné les p-values trouvées, il serait plus vraisemblable que son impact soit largement diminué mais pas nécessairement inexistant. De même, le manque de puissance a pu amener à éliminer des facteurs pronostiques dans la phase de sélection. Cependant, il est très vraisemblable que ces facteurs restent secondaires par rapport à la pente précédant la mise sous hormonothérapie.

Malgré les limites des données, ce travail est unique en son genre. Il propose pour la première fois une description de la dynamique du PSA post-HT et de ces déterminants. De plus, il met à nouveau en évidence l'importance de la vélocité des PSA sur la progression du cancer de la prostate, ce qui dans le contexte de l'initiation d'hormonothérapie, pourrait aider à déterminer le meilleur "timing" pour une hormonothérapie optimale. Pour ces raisons, ce travail fera l'objet d'une publication dans une revue clinique prochainement.

Chapitre VI

Discussion, conclusion et perspectives

Dans ce travail de thèse, nous avons développé et validé des outils pronostiques dynamiques de rechute clinique chez des patients atteints de cancer de la prostate localisé et traités initialement par radiothérapie externe. Dans un premier travail, nous avons comparé différentes approches de modélisation pour prédire le risque de rechute clinique. Par la suite, nous avons proposé un outil pronostique dynamique différentiel permettant de prédire le risque de rechute clinique selon l'initiation ou non d'un second traitement au cours du suivi. Enfin, un dernier travail nous a permis de décrire la dynamique des PSA après l'initiation d'un traitement hormonal chez les patients traités initialement par radiothérapie.

Nous aborderons, dans ce chapitre, une discussion sur ces travaux réalisés en présentant leurs forces et leurs faiblesses. Nous clôturerons cette thèse par une conclusion générale et présenterons quelques perspectives envisageables.

1 Discussion

1.1 Prédiction dynamique : comparaison d'approches

Dans le chapitre [III](#), nous avons tout d'abord réalisé une étude comparative des modèles conjoints à effets aléatoires partagés suivant la structure de dépendance qui était supposée

entre la dynamique des PSA et le risque de rechute clinique. Cette comparaison s'est effectuée suivant différents critères comme leur ajustement aux données, leur adéquation aux hypothèses du modèle et leur pouvoir prédictif. Le même type d'étude comparative a été réalisé entre les modèles conjoints à effets aléatoires partagés et les modèles conjoints à classes latentes.

Dans le premier travail, nous avons constaté que certains types de dépendance dans les modèles conjoints à effets aléatoires partagés ne respectaient pas nécessairement les hypothèses du modèle. C'est le cas notamment avec une dépendance sur le niveau courant du PSA qui ne respecte pas l'hypothèse de log-linéarité du modèle de survie. Pourtant le modèle conjoint à effets aléatoires partagés le plus répandu en pratique est justement celui-ci, avec une dépendance sur le niveau courant du biomarqueur uniquement. Il est vrai qu'il est facile à interpréter et constitue l'extension directe des modèles de survie avec variables dépendant du temps. Mais les utilisateurs de modèles conjoints oublient souvent que cette dépendance n'est qu'une des innombrables structures de dépendance possibles et que comme dans tout modèle, les hypothèses (notamment de log-linéarité) devraient être évaluées afin de se prémunir de toute mauvaise spécification qui pourrait entraîner des conclusions erronées.

Lorsque l'on s'intéresse au pouvoir prédictif des modèles, il en est de même. Le modèle incluant uniquement le niveau courant s'est révélé, sur notre exemple du cancer de la prostate être un des plus mauvais candidats. En effet, le prédicteur principal semble dans le cancer de la prostate être la pente, c'est-à-dire la vitesse des PSA si bien que ce sont les modèles mettant le plus d'impact sur la vitesse qui se sont révélés avoir le meilleur pouvoir prédictif.

Ce travail aussi a pu mettre en évidence le fait qu'il y a une différence majeure entre ajustement aux données jointes et pouvoir prédictif. Par exemple, les modèles conjoints supposant une dépendance sur les effets aléatoires directement n'étaient pas ceux que l'on

aurait choisis pour leur ajustement aux données. Pourtant, leur pouvoir prédictif était largement meilleur. Cela s'explique assez simplement. Les mesures d'ajustement se basent à la fois sur l'ajustement des données répétées et l'ajustement de la partie survie. Il est vrai que dans le premier travail, le modèle pour les données répétées était toujours le même et que finalement, les différences en terme d'ajustement se sont basées principalement sur l'ajustement aux données de survie. Cependant, par rapport aux mesures de pouvoir prédictif, les mesures d'ajustement prennent en compte toute l'information, quels que soient les délais de survenue d'événement. Au contraire, les mesures de pouvoir prédictif se focalisent sur une fenêtre spécifique de survie, et à partir d'un temps de prédiction donné. C'est ainsi que le modèle basé directement sur les effets aléatoires s'est révélé être meilleur que les autres modèles mais uniquement à partir d'un point de prédiction 2.25 années après la fin de la radiothérapie alors que près de la moitié des événements avait déjà eu lieu.

Cette différence entre qualité d'ajustement et pouvoir prédictif a été encore plus nette dans le deuxième travail de comparaison entre les modèles conjoints à effets aléatoires partagés et les modèles conjoints à classes latentes. En effet, les modèles conjoints à classes latentes permettent un ajustement beaucoup plus précis des données répétées si bien que les mesures d'ajustement aux données ont été bien meilleures. Lorsque l'intérêt a été sur la survie et l'aspect pronostique spécifiquement, les différences avec les modèles conjoints à effets aléatoires partagés ont été plus limitées.

Ces deux travaux ont donc permis d'illustrer et comparer les différentes approches de modélisation conjointe que l'on peut considérer lorsque l'on veut modéliser un biomarqueur et un temps d'événement conjointement. Ils avaient pour vocation d'illustrer de façon pédagogique ces différences. Nous nous sommes donc limités ici à un petit échantillon de données. Les conclusions plus cliniques restent donc assez limitées. C'est la deuxième partie de la thèse discutée ci-dessous qui a permis de confirmer nombre des résultats de ce travail sur un grand échantillon de données.

1.2 Prédiction dynamique différentielle

Dans le chapitre IV, nous avons développé des outils pronostiques dynamiques dérivés de modèles conjoints à effets aléatoires partagés qui avaient la particularité d'être "différentiels" en ce sens qu'ils étaient calculés dans deux contextes différents vis-à-vis de l'initiation de l'hormonothérapie. Pour cela, nous avons développé plusieurs modèles conjoints à effets aléatoires partagés qui différaient de par leur façon de prendre en compte l'association entre la dynamique des PSA, l'initiation d'une éventuelle hormonothérapie au cours du suivi et le risque de rechute clinique. La comparaison de ces modèles s'est effectuée, comme dans les analyses précédentes, à la fois en terme de qualité d'ajustement aux données et en terme de pouvoir prédictif.

Ce travail a plusieurs originalités. Tout d'abord, nous avons pour la première fois proposé d'utiliser les modèles conjoints pour faire de la prédiction d'événement selon deux scénarios de traitement. En effet, jusqu'à présent, les outils pronostiques dynamiques étaient développés et validés sous l'hypothèse qu'aucun changement de suivi du patient n'avait lieu. C'est notamment ce que nous avons supposé dans le chapitre précédent. Or, en pratique, il n'est pas rare que les patients et leurs cliniciens aient à prendre la décision d'initier un nouveau traitement au cours du suivi, le bénéfice d'une telle mise sous traitement et le meilleur "timing" pour le faire étant difficile à apprécier. Ce travail a permis de proposer des quantifications du risque de rechute selon la mise immédiate sous hormonothérapie qui pourrait aider le clinicien et le patient à mieux apprécier la plus-value clinique de ce traitement. Bien sûr, cette décision se base aussi sur des critères de qualité de vie que notre outil différentiel ne prend pas en compte. Néanmoins, la quantification du risque de rechute reste un élément important dans ce type de contexte.

Comme ce travail s'est basé sur des données observationnelles, il est potentiellement sujet à un biais d'indication dont nous sommes conscients. Avant d'être utilisé en pratique, cet

outil différentiel devrait donc sans doute être d'abord recalculé sur des données dans lesquelles la mise sous hormonothérapie aurait été randomisée, puis évalué sur des données externes.

Notre travail a surtout consisté à donner la méthodologie statistique permettant ce type de développement. En effet, le développement et la validation d'outil pronostique dynamique n'est pas simple et s'est révélé encore plus difficile dans ce contexte de scénarios de traitement.

Dans ce travail comme dans le précédent, nous avons appuyé la validation des outils pronostiques sur deux mesures de pouvoir prédictif.

L'entropie croisée pronostique, basée sur la densité de l'événement conditionnellement à l'histoire du biomarqueur, avait été proposée par [Commenges et al. \(2012\)](#) et appliquée uniquement aux modèles conjoints à classes latentes. Nous l'avons adaptée pour l'évaluation de modèles conjoints à effets aléatoires partagés. De plus cette mesure avait été développée pour évaluer la valeur pronostique d'un modèle à partir d'un temps de prédiction s sur une fenêtre d'horizons infinie. Nous l'avons adaptée à l'évaluation sur une fenêtre d'horizons de t années.

La deuxième mesure considérée est le score de Brier. Elle correspond à l'erreur quadratique moyenne de prédiction. Elle est beaucoup considérée dans la validation sur des données externes d'outils pronostiques en présence de censure. Dans ces travaux de thèse, nous avons proposé un nouvel estimateur qui a le grand avantage de corriger le sur-optimisme présent lors de la validation sur les données d'estimation, et ainsi de permettre une évaluation objective d'outils pronostiques dynamiques sur ces données. Nous avons pour cela appliqué une formule de validation croisée approchée ([Commenges et al., 2013](#)). Par contre, nous ne l'avons pas validée formellement par des simulations.

Ces deux mesures ont aussi été implémentées sous R et seront prochainement disponibles sur internet pour le calcul des performances prédictives de modèles estimés par le

package R JM ou des versions modifiées du package pour des dépendances plus spécifiques traitées dans cette thèse.

Dans le contexte de scénarios de traitement, l'évaluation des performances prédictives doit être faite séparément dans les deux populations ciblées. Dans notre cas, il s'agissait de la population ne recevant pas de traitement au temps de prédiction et la population recevant le traitement à ce temps exactement. Cela a donné lieu dans le premier cas à une application classique des mesures de validation, à ceci près que la population d'étude pouvait être définie de plusieurs façons à cause des initiations ultérieures de l'hormonothérapie. Dans le deuxième cas, il a fallu définir une mesure moyenne de pouvoir prédictif puisqu'à chaque temps de prédiction (temps d'hormonothérapie individuel), nous ne disposions que des prédictions pour un ou deux sujets ayant reçu l'hormonothérapie à ce temps précis.

Enfin dans ce travail, nous avons choisi de nous focaliser sur les modèles conjoints à effets aléatoires partagés pour construire ces outils pronostiques. En effet ces modèles offrent la possibilité de spécifier de nombreuses structures de dépendance entre la dynamique du biomarqueur et le risque d'événement, et dans notre cas de figure, l'initiation d'un second traitement. Ce choix s'inscrivait également dans la continuité des travaux présentés dans la section 2 du chapitre III et dans le chapitre IV dans lesquels nous avons souhaité investiguer différentes formes de dépendances et leur impact sur les qualités prédictives des modèles.

1.3 Dynamique des PSA après l'hormonothérapie

L'analyse de la dynamique des PSA après un traitement hormonal chez des patients traités initialement par radiothérapie présentée dans le chapitre V constitue un travail en marge de l'objectif premier de cette thèse. Elle a été justifiée par le manque d'information dont nous disposions sur la dynamique des PSA après une mise sous traitement hormo-

nal au cours du suivi alors que cette initiation de second traitement au cours du suivi des patients était au cœur de nos préoccupations. En effet, à notre connaissance, cette dynamique et ses déterminants n'ont jamais été étudiés auparavant, et dans la plupart des analyses conjointes de données longitudinales de PSA et de risque de rechute clinique (Proust-Lima and Taylor, 2009a ; Taylor et al., 2013a) comme dans celles détaillées dans les chapitres III et IV de cette thèse, ces données post-HT étaient même supprimées.

Le résultat principal de cette étude est que le prédicteur majeur de la dynamique post-hormonothérapie (le seul trouvé dans ce travail) est la pente de PSA précédant l'initiation de ce traitement. Certes, les données de cette analyse étaient loin d'être parfaites comme cela a été discuté dans le chapitre V. Nous manquons de puissance avec les données de seulement 238 patients et nous manquons aussi de précisions sur la nature de l'hormonothérapie et l'initiation d'éventuelles secondes thérapies. Néanmoins, des analyses de sensibilité assez poussées nous ont systématiquement confirmé ce résultat.

De plus, dans le chapitre IV, nous avons déjà trouvé que le modèle prédisant le mieux le risque de rechute clinique après l'initiation immédiate d'une hormonothérapie était celui qui supposait un effet spécifique de la pente au moment de l'hormonothérapie. En particulier, nous trouvons déjà dans ce travail que le niveau de PSA au moment de l'hormonothérapie n'était plus prédicteur du risque de rechute clinique post hormonothérapie après ajustement sur la pente.

2 Conclusion

Dans ce travail de thèse, nous avons largement développé l'analyse conjointe des données répétées de PSA et le risque de rechute clinique associé à travers les données de plusieurs grandes cohortes. Nous nous sommes également spécifiquement intéressés au problème de la mise sous hormonothérapie au cours du suivi. En effet, la question du meilleur moment ("timing") pour cette initiation de traitement reste assez abstraite, les

cliniciens se reposant sur leur propre expérience le plus souvent.

Grâce d'une part à une modélisation de la dynamique post-hormonothérapie, et d'autre part au développement d'outils pronostiques différentiels, nous avons fourni des éléments qui pourraient contribuer à mieux définir le meilleur moment pour initier ce type de traitement qui a des implications potentiellement lourdes sur le devenir clinique face à la maladie et sur le quotidien du patient.

D'un point de vue statistique, nous avons aussi approfondi l'analyse des modèles conjoints à effets aléatoires partagés. D'une part, nous avons détaillé la façon d'implémenter ce type de modèles en pratique et les questions qu'il est nécessaire de se poser. D'autre part, nous avons appliqué ce type de développement à une problématique nouvelle de scénarios de traitement qui prend tout son sens à l'heure du développement de la médecine personnalisée.

3 Perspectives

A la suite de ce travail de thèse, plusieurs perspectives peuvent être envisagées.

Dans le travail présenté dans la section 3 du chapitre III, nous avons comparé des modèles conjoints à classes latentes et des modèles à effets aléatoires partagés. Dans la perspective de proposer des outils toujours plus prédictifs, il pourrait être intéressant de considérer des modèles conjoints combinant les classes latentes et les effets aléatoires partagés comme proposé par [Beunckens et al. \(2008\)](#) et [Garre et al. \(2008\)](#). Cela permettrait d'assouplir les hypothèses de ces deux familles de modèles, notamment les hypothèses d'indépendance conditionnelle.

Dans le cancer de la prostate, plusieurs types de rechute clinique peuvent exister : une rechute locale, régionale, distante ou le décès du patient par le cancer. Dans ce travail de thèse, nous avons considéré une définition commune pour la rechute clinique. [Elashoff et al. \(2008\)](#) et [Huang et al. \(2011\)](#) ont proposé des extensions des modèles conjoints

en présence de risques compétitifs. Appliqués au cancer de la prostate, cela permettrait d'évaluer l'association entre le PSA et les différents types de rechute clinique et de proposer des outils pronostiques dynamiques spécifiques à chaque type de rechute.

Dans le but de proposer un bon outil pronostique dynamique, nous avons souvent été amenés à comparer les performances prédictives de différents modèles conjoints. [Rizopoulos et al. \(2013\)](#) proposent, au lieu de comparer différents modèles, de les combiner en utilisant la méthode "Bayesian Model Averaging" afin d'obtenir un modèle plus puissant possédant un meilleur pouvoir prédictif. Cette méthodologie pourrait aussi être envisagée dans le cadre de la prédiction du risque de rechute clinique après un traitement.

Le PSA est le biomarqueur le plus utilisé pour le diagnostic ou le pronostic du cancer de la prostate. Cependant, il existe un nouveau biomarqueur, le PCA3, qui est de plus en plus utilisé pour le diagnostic du cancer. Ce biomarqueur pourrait également être intéressant à considérer dans l'étude du risque de rechute clinique et le développement d'outils pronostiques dynamiques. Plus généralement, une modélisation conjointe des deux biomarqueurs, le PSA et le PCA3, couplée avec différents types de rechute clinique fournirait potentiellement des outils pronostiques très intéressants, et possiblement une meilleure compréhension de la maladie.

Enfin, comme précédemment discuté, la méthodologie proposée dans le chapitre IV à propos du développement d'outils pronostiques dynamiques différentiels pourrait être appliquée sur des données d'essais cliniques dans lesquelles les dates d'initiation de l'hormonothérapie seraient définies plus formellement. Cela permettrait de proposer en pratique cet outil différentiel sans le problème de biais d'indication que nous avons rencontré. En outre, cela pourrait peut-être aussi permettre de réaliser une évaluation de type bénéfice/risque de l'initiation d'une hormonothérapie au cours du suivi.

Références bibliographiques

- [1] Z. D. Abdi, M. Essig, D. Rizopoulos, Y. Le Meur, A. Prémaud, J. Woillard, J. Rérolle, P. Marquet, and A. Rousseau. Impact of longitudinal exposure to mycophenolic acid on acute rejection in renal-transplant recipients using a joint modeling approach. *Pharmacological Research*, 72(0) :52–60, 2013.
- [2] P. S. Albert and J. H. Shih. On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics*, pages 983–991, Sept. 2010.
- [3] P. Albertsen, J. Hanley, D. Penson, and J. Fine. Validation of increasing prostate specific antigen as a predictor of prostate cancer death after treatment of localized prostate cancer with surgery or radiation. *The Journal of Urology*, 171(6, Part 1) : 2221–2225, 2004.
- [4] D. G. Altman and P. Royston. What do we mean by validating a prognostic model?. *Statistics in Medicine*, 19(4) :453–473, 2000.
- [5] G. L. Andriole, E. D. Crawford, R. L. Grubb, S. S. Buys, D. Chia, T. R. Church, M. N. Fouad, E. P. Gelmann, P. A. Kvale, D. J. Reding, J. L. Weissfeld, L. A. Yokochi, B. O’Brien, J. D. Clapp, J. M. Rathmell, T. L. Riley, R. B. Hayes, B. S. Kramer, G. Izmirlian, A. B. Miller, P. F. Pinsky, P. C. Prorok, J. K. Gohagan, and C. D. Berg. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine*, 360(13) :1310–1319, 2009.
- [6] ASTRO. American society for therapeutic radiology and oncology consensus panel. consensus statement : Guidelines for psa following radiation therapy. *International Journal of Radiation Oncology Biology Physics*, 37 :1035–1041, 1997.
- [7] A. J. Atkinson, W. A. Colburn, V. G. DeGruttola, D. L. DeMets, G. J. Downing, D. F. Hoth, J. A. Oates, C. C. Peck, R. T. Schooley, B. A. Spilker, J. Woodcock, and S. L. Zeger. Biomarkers and surrogate endpoints : Preferred definitions and conceptual framework. *Clinical Pharmacology Therapy*, 69 :89–95, 2001.

- [8] M. J. Barry. Prostate-specific-antigen testing for early diagnosis of prostate cancer. *New England Journal of Medicine*, 344(18) :1373–1377, 2001.
- [9] C. A. Bellera, J. A. Hanley, L. Joseph, and P. C. Albertsen. Hierarchical changepoint models for biochemical markers illustrated by tracking postradiotherapy prostate-specific antigen series in men with prostate cancer. *Annals of epidemiology*, 18(4) : 270–82, Apr. 2008a.
- [10] C. A. Bellera, J. A. Hanley, L. Joseph, and P. C. Albertsen. Detecting trends in noisy data series : Application to biomarker series. *American Journal of Epidemiology*, 167 (9) :1130–1139, 2008b.
- [11] C. A. Bellera, J. A. Hanley, L. Joseph, and P. C. Albertsen. A statistical evaluation of rules for biochemical failure after radiotherapy in men treated for prostate cancer. *Radiation Oncology*, 75(5) :1357–1363, 2009.
- [12] A. Berchtold. Optimisation of mixture models : Comparison of different strategies. *Computational Statistics*, 19(3) :385–406, 2004.
- [13] C. Beunckens, G. Molenberghs, G. Verbeke, and C. Mallinckrodt. A latent-class mixture model for incomplete longitudinal gaussian data. *Biometrics*, 64(1) :96–105, Mar. 2008.
- [14] F. J. Bianco, P. T. Scardino, and J. A. Eastham. Radical prostatectomy : long-term cancer control and recovery of sexual and urinary function ("trifecta"). *Urology*, 66 (5 Suppl) :83–94, Nov. 2005.
- [15] C. Biernacki, G. Celeux, and G. Govaert. Choosing starting values for the em algorithm for getting the highest likelihood in multivariate gaussian mixture models. *Computational Statistics & Data Analysis*, 41(3-4) :561–575, Jan. 2003.
- [16] A. Bill-Axelsson, L. Holmberg, M. Ruutu, H. Garmo, J. R. Stark, C. Busch, S. Nordling, M. Häggman, S.-O. Andersson, S. Bratell, A. Spangberg, J. Palmgren, G. Steineck, H.-O. Adami, and J.-E. Johansson. Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine*, 364(18) :1708–1717, 2011.
- [17] T. J. Bradford, S. A. Tomlins, X. Wang, and A. M. Chinnaiyan. Molecular markers of prostate cancer. *Urologic Oncology : Seminars and Original Investigations*, 24(6) : 538–551, 2006.
- [18] G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1950.
- [19] E. R. Brown and J. G. Ibrahim. Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics*, 59(3) :686–93, Sept. 2003.

-
- [20] E. R. Brown, J. G. Ibrahim, and V. DeGruttola. A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1) :64–73, Mar. 2005.
- [21] H. B. Carter. A psa threshold of 4.0 ng/ml for early detection of prostate cancer : the only rational approach for men 50 years old and older. *Urology*, 55(6) :796–799, 2000.
- [22] W. J. Catalona, D. S. Smith, T. L. Ratliff, K. M. Dodds, D. E. Coplen, J. J. Yuan, J. A. Petros, and G. L. Andriole. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *New England Journal of Medicine*, 324(17) : 1156–1161, 1991.
- [23] M. M. Center, A. Jemal, J. Lortet-Tieulent, E. Ward, J. Ferlay, O. Brawley, and F. Bray. International variation in prostate cancer incidence and mortality rates. *European urology*, 61(6) :1079–1092, June 2012.
- [24] Y. Y. Chi and J. G. Ibrahim. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2) :432–45, June 2006.
- [25] G. Chodak. Prostate cancer : epidemiology, screening, and biomarkers. *Reviews in urology*, 8 Suppl 2 :S3–8, Jan. 2006.
- [26] D. Commenges, B. Liqueur, and C. Proust-Lima. Choice of prognostic estimators in joint models by estimating differences of expected conditional kullback-leibler risks. *Biometrics*, 68(2) :380–7, June 2012.
- [27] D. Commenges, C. Proust-Lima, C. Samieri, and B. Liqueur. A universal approximate cross-validation criterion and its asymptotic distribution. *arXiv :1206.1753 [math.ST]*, 2013.
- [28] D. R. Cox. Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, 34 :187–220, 1972.
- [29] R. A. Critz, W. H. Williams, J. B. Benton, A. K. Levinson, C. T. Holladay, and D. A. Holladay. Prostate specific-antigen bounce after radioactive seed implantation followed by external beam radiation for prostate cancer. *The Journal of Urology*, 63 (4) :1085–1089, 2000.
- [30] J. M. Crook, C. J. O’Callaghan, G. Duncan, D. P. Dearnaley, C. S. Higano, E. M. Horwitz, E. Frymire, S. Malone, J. Chin, A. Nabid, P. Warde, T. Corbett, S. Angyalfi, S. L. Goldenberg, M. K. Gospodarowicz, F. Saad, J. P. Logue, E. Hall, P. F. Schellhammer, K. Ding, and L. Klotz. Intermittent androgen suppression for rising psa level after radiotherapy. *New England Journal of Medicine*, 367(10) :895–903, 2012.
- [31] M. J. Crowther, K. R. Abrams, and P. C. Lambert. Flexible parametric joint modelling of longitudinal and survival data. *Statistics in Medicine*, 31(30) :4456–4471, 2012.

- [32] U. G. Dafni and A. A. Tsiatis. Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, 54(4) :1445–62, Dec. 1998.
- [33] A. D’Amico, A. Renshaw, B. Sussman, and M. Chen. Pretreatment psa velocity and risk of death from prostate cancer following external beam radiation therapy. *Journal of the American Medical Association*, 294(4) :440–447, 2005.
- [34] A. D’Amico, M. Chen, A. Renshaw, M. Loffredo, and P. Kantoff. Androgen suppression and radiation vs radiation alone for prostate cancer : A randomized trial. *Journal of the American Medical Association*, 299(3) :289–295, 2008.
- [35] A. V. D’Amico, R. Whittington, S. B. Malkowicz, D. Schultz, K. Blank, G. A. Broderick, J. E. Tomaszewski, A. A. Renshaw, I. Kaplan, C. J. Beard, and A. Wein. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *Journal of the American Medical Association*, 280(11) :969–74, Sept. 1998.
- [36] A. V. D’Amico, M.-H. Chen, K. A. Roehl, and W. J. Catalona. Preoperative psa velocity and the risk of death from prostate cancer after radical prostatectomy. *New England Journal of Medicine*, 351(2) :125–135, 2004a.
- [37] A. V. D’Amico, J. W. Moul, P. R. Carroll, K. Cote, L. Sun, D. Lubeck, A. A. Renshaw, M. Loffredo, and M.-H. Chen. Intermediate end point for prostate cancer-specific mortality following salvage hormonal therapy for prostate-specific antigen failure. *Journal of the National Cancer Institute*, 96(7) :509–515, 2004b.
- [38] V. De Gruttola and X. M. Tu. Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, 50(4), Dec. 1994.
- [39] A. P. Dempster, M. N. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 39 :1–22, 1977.
- [40] I. L. Deras, S. M. Aubin, A. Blase, J. R. Day, S. Koo, A. W. Partin, W. J. Ellis, L. S. Marks, Y. Fradet, H. Rittenhouse, and J. Groskopf. Pca3 : A molecular urine assay for predicting prostate biopsy outcome. *The Journal of Urology*, 179(4) :1587–1592, 2008.
- [41] P. Diggle and M. Kenward. Informative dropout in longitudinal data analysis (with discussion). *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 43 :49–94, 1994.
- [42] J. Ding and J.-L. Wang. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*, 64(2) :546–56, June 2008.
- [43] S.-J. Drouin and M. Rouprêt. Epidemiology, diagnosis and prognosis of localized prostate cancer : what’s new ?. *Progrès en urologie : journal de l’Association française d’urologie et de la Société française d’urologie*, 19 Suppl 1 :S3–7, Apr. 2009.

-
- [44] 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*, 64(3) :762–71, Sept. 2008.
- [45] C. L. Faucett and D. C. Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates : a gibbs sampling approach. *Statistics in medicine*, 15(15) :1663–85, Aug. 1996.
- [46] G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs. *Longitudinal Data Analysis*. Chapman & Hall/CRC Press, 2008.
- [47] P. Fletcher. *Practical Methods of Optimization, second edition (Chapter 3)*. John Wiley & Sons, Ltd, 2000.
- [48] F. G. Garre, A. H. Zwinderman, R. B. Geskus, and Y. W. Sijpkens. A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of the Royal Statistical Society : Series A (Statistics in Society)*, 171(1) :299–308, 2008.
- [49] T. A. Gerds and M. Schumacher. Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6) :1029–1040, Dec. 2006.
- [50] T. A. Gerds and M. Schumacher. Efron-type measures of prediction error for survival analysis. *Biometrics*, 63(4) :1283–7, Dec. 2007.
- [51] P. Ghosh, K. Ghosh, and R. C. Tiwari. Joint modeling of longitudinal data and informative dropout time in the presence of multiple changepoints. *Statistics in medicine*, 30(6) :611–26, Mar. 2011.
- [52] Globocan. International agency for research on cancer : World health organization. *Globocan*, 2008.
- [53] C. Gosselaar, M. J. Roobol, S. Roemeling, and F. H. Schröder. The role of the digital rectal examination in subsequent screening visits in the european randomized study of screening for prostate cancer (erspc), rotterdam. *European Urology*, 54(3) :581–588, 2008.
- [54] E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18) :2529–45, 1999.
- [55] P. L. Graham, L. M. Ryan, and M. a. Luszcz. Joint modelling of survival and cognitive decline in the australian longitudinal study of ageing. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 60(2) :221–238, Mar. 2011.
- [56] X. Guo and B. P. Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1) :16–24, Feb. 2004.

- [57] G. P. Haas, N. B. Delongchamps, R. F. Jones, V. Chandan, A. M. Serio, A. J. Vickers, M. Jumbelic, G. Threatte, R. Korets, H. Lilja, and G. de la Roza. Needle biopsies on autopsy prostates : Sensitivity of cancer detection based on true prevalence. *Journal of the National Cancer Institute*, 99(19) :1484–1489, 2007.
- [58] G. P. Haas, N. Delongchamps, O. W. Brawley, C. Y. Wang, and G. de la Roza. The worldwide epidemiology of prostate cancer : perspectives from autopsy studies. *The Canadian journal of urology*, 15(1) :3866–71, Feb. 2008.
- [59] S. L. Hancock, R. S. Cox, and M. A. Bagshaw. Prostate specific antigen after radiotherapy for prostate cancer : a reevaluation of long-term biochemical control and the kinetics of recurrence in patients treated at stanford university. *The Journal of urology*, 154(4) :1412–7, Oct. 1995.
- [60] G. E. Hanks, A. L. Hanlon, B. Epstein, and E. M. Horwitz. Dose response in prostate cancer with 8-12 years' follow-up. *International journal of radiation oncology, biology, physics*, 54(2) :427–35, Oct. 2002.
- [61] A. L. Hanlon, H. Diratzouian, and G. E. Hanks. Posttreatment prostate-specific antigen nadir highly predictive of distant failure and death from prostate cancer. *International journal of radiation oncology, biology, physics*, 53(2) :297–303, June 2002.
- [62] F. E. J. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati. Evaluating the yield of medical tests. *Journal of the American Medical Association*, 247(18) : 2543–2546, 1982.
- [63] D. A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358) : 320–338, 1977.
- [64] D. Hawkins, D. Allen, and A. Stromberg. Determining the number of components in mixtures of linear models. *Computational Statistics & Data Analysis*, 38(1) :15–48, 2001.
- [65] B. He and S. Luo. Joint modeling of multivariate longitudinal measurements and survival data with applications to parkinson's disease. *Statistical Methods in Medical Research*, 2013.
- [66] P. Heagerty and Y. Zheng. Survival model predictive accuracy and roc curves. *Biometrics*, 61 :92–105, 2005.
- [67] J. Heckman. The common structure of statistical models of truncated, sample selection and limited dependent variables, and a simple estimator of such models. *Annals of Economic and Social Measurement*, 5 :475–492, 1976.
- [68] A. Heidenreich, P. J. Bastian, J. Bellmunt, M. Bolla, S. Joniau, M. D. Mason, V. Matveev, N. Mottet, T. H. V. D. Kwast, T. Wiegel, and F. Zattoni. Guidelines on prostate cancer. *EAU European Association of Urology*, 2013.

-
- [69] R. Henderson, P. Diggle, and A. Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics (Oxford, England)*, 1(4) :465–80, Dec. 2000.
- [70] R. Henderson, P. Diggle, and A. Dobson. Identification and efficacy of longitudinal markers for survival. *Biostatistics (Oxford, England)*, 3(1) :33–50, Mar. 2002.
- [71] J. W. Hogan and N. M. Laird. Mixture models for the joint distribution of repeated measures and event times. *Statistics in medicine*, 16(1-3) :239–57, 1997.
- [72] L. Holmberg, A. Bill-Axelsson, F. Helgesen, J. O. Salo, P. Folmerz, M. Häggman, S.-O. Andersson, A. Spangberg, C. Busch, S. Nordling, J. Palmgren, H.-O. Adami, J.-E. Johansson, and B. J. Norlén. A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer. *New England Journal of Medicine*, 347(11) :781–789, 2002.
- [73] E. Horwitz, H. Thames, D. Kuban, L. Levy, P. Kupelian, A. Martinez, J. Michalski, T. Pisansky, H. Sandler, W. Shipley, M. Zelefsky, and A. Hanks G.E. abd Zietman. Definitions of biochemical failure that best predict clinical failure in patients with prostate cancer treated with external beam radiation alone : a multi-institutional pooled analysis. *The Journal of Urology*, 173(3) :797–802, 2005.
- [74] E. M. Horwitz, L. B. Levy, H. D. Thames, P. A. Kupelian, A. A. Martinez, J. M. Michalski, T. M. Pisansky, H. M. Sandler, W. U. Shipley, M. J. Zelefsky, A. L. Zietman, and D. A. Kuban. Biochemical and clinical significance of the posttreatment prostate-specific antigen bounce for prostate cancer patients treated with external beam radiation therapy alone. *Cancer*, 107(7) :1496–1502, 2006.
- [75] P. Hu, A. A. Tsiatis, and M. Davidian. Estimating the parameters in the cox model when covariate variables are measured with error. *Biometrics*, 54(4) :1407–1419, 1998.
- [76] X. Huang, G. Li, R. Elashoff, and J. Pan. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Analysis*, 17 :80–100, 2011.
- [77] J. G. Ibrahim, M.-H. Chen, and D. Sinha. Bayesian survival analysis. *Springer*, 2001.
- [78] J. G. Ibrahim, H. Chu, and L. M. Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 28(16) :2796–801, June 2010.
- [79] H. Jacqmin-Gadda, D. Commenges, and J. Dartigues. Analyse de données longitudinales gaussiennes comportant des données manquantes sur la variable à expliquer. *Revue d'épidémiologie et Santé publique*, 47 :525–34, 1999.

- [80] H. Jacqmin-Gadda, R. Thiébaud, and J. F. Dartigues. Modélisation conjointe de données longitudinales quantitatives et délais censurés. *Revue d'épidémiologie et de santé publique*, 52(6) :502–10, 2004.
- [81] H. Jacqmin-Gadda, D. Commenges, and J.-F. Dartigues. Random change point model for joint modeling of cognitive decline and dementia. *Biometrics*, 62(1) : 254–60, Mar. 2006.
- [82] H. Jacqmin-Gadda, C. Proust-Lima, J. M. G. Taylor, and D. Commenges. Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. *Biometrics*, 66(1) :11–9, Mar. 2010.
- [83] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman. Global cancer statistics. *CA : a cancer journal for clinicians*, 61(2) :69–90, 2011.
- [84] J. Kalbfleisch and R. Prentice. *The statistical analysis of failure time data, Second Edition*. Wiley & Sons, 2002.
- [85] E. H. Kennedy, J. M. G. Taylor, D. E. Schaubel, and S. Williams. The effect of salvage therapy on survival in a longitudinal study with treatment by indication. *Statistics in medicine*, 29(25) :2569–80, Nov. 2010.
- [86] L. L. Kestin, F. A. Vicini, E. L. Ziaja, J. S. Stromberg, R. C. Frazier, and A. A. Martinez. Defining biochemical cure for prostate carcinoma patients treated with external beam radiation therapy. *Cancer*, 86(8) :1557–66, Oct. 1999.
- [87] D. A. Kuban, A. M. El-Mahdi, and P. F. Schellhammer. Prostate-specific antigen for pretreatment prediction and posttreatment evaluation of outcome after definitive irradiation for prostate cancer. *International Journal of Radiation Oncology Biology Physics*, 32(2) :307–316, 1995.
- [88] D. A. Kuban, H. D. Thames, L. B. Levy, E. M. Horwitz, P. A. Kupelian, A. A. Martinez, J. M. Michalski, T. M. Pisansky, H. M. Sandler, W. U. Shipley, M. J. Zelefsky, and A. L. Zietman. Long-term multi-institutional analysis of stage t1-t2 prostate cancer treated with radiotherapy in the {PSA} era. *International Journal of Radiation Oncology Biology Physics*, 57(4) :915–928, 2003.
- [89] D. a. Kuban, H. D. Thames, and W. U. Shipley. Defining recurrence after radiation for prostate cancer. *The Journal of urology*, 173(6) :1871–8, June 2005.
- [90] D. A. Kuban, L. B. Levy, L. Potters, D. C. Beyer, J. C. Blasko, B. J. Moran, J. P. Ciezki, A. L. Zietman, M. J. Zelefsky, T. M. Pisansky, M. Elshaikh, and E. M. Horwitz. Comparison of biochemical failure definitions for permanent prostate brachytherapy. *International Journal of Radiation Oncology Biology Physics*, 65(5) : 1487–1493, 2006.
- [91] P. Kupelian, H. Thames, L. Levy, E. Horwitz, A. Martinez, J. Michalski, T. Pisansky, H. Sandler, W. Shipley, M. Zelefsky, A. Zietman, and D. Kuban. Year of treatment as independent predictor of relapse-free survival in patients with localized prostate

- cancer treated with definitive radiotherapy in the psa era. *International journal of radiation oncology, biology, physics*, 63(3) :795–9, Nov. 2005.
- [92] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38 :963–974, 1982.
- [93] N. J. Law, J. M. G. Taylor, and H. Sandler. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics (Oxford, England)*, 3(4) :547–63, Dec. 2002.
- [94] W. R. Lee, G. E. Hanks, and A. Hanlon. Increasing prostate-specific antigen profile following definitive radiation therapy for localized prostate cancer : clinical observations. *Journal of Clinical Oncology*, 15(1) :230–8, 1997.
- [95] M. F. Leitzmann and S. Rohrmann. Risk factors for the onset of prostatic cancer : age, location, and behavioral correlates. *Clinical epidemiology*, 4 :1–11, Jan. 2012.
- [96] E. Lesaffre and B. Spiessens. On the effect of the number of quadrature points in a logistic random effects model : an example. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 50(3) :325–335, 2001.
- [97] H. Lin, C. E. McCulloch, B. W. Turnbull, E. H. Slate, and L. C. Clark. A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine*, 19(10) :1303–1318, 2000.
- [98] H. Lin, C. E. McCulloch, and S. T. Mayne. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*, 21(16) :2369–2382, 2002.
- [99] H. Lin, B. W. Turnbull, C. E. McCulloch, and E. H. Slate. Latent class models for joint analysis of longitudinal biomarker and event process data : Application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, 97(457) :53–65, mar 2002.
- [100] H. Lin, C. E. McCulloch, and R. A. Rosenheck. Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics*, 60(2) : 295–305, 2004.
- [101] M. J. Lindstrom and D. M. Bates. Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404) :1014–1022, 1988.
- [102] R. J. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. Wiley, New York, 1987.
- [103] R. J. Little and D. B. Rubin. *Statistical Analysis with Missing Data, Second edition*. Wiley, New York, 2002.

- [104] J. Long and J. Ryoo. Using fractional polynomials to model non-linear trends in longitudinal data. *British Journal of Mathematical and Statistical Psychology*, 63 (1) :177–203, 2010.
- [105] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2) :431–441, 1963.
- [106] M. May, P. Royston, M. Egger, A. C. Justice, and J. A. Sterne. Development and validation of a prognostic model for survival time data : application to prognosis of hiv positive patients treated with antiretroviral therapy. *Statistics in Medicine*, 23 (15) :2375–2398, 2004.
- [107] R. Miralbell, S. Pampallona, M. Rouzaud, A. Pugliesi, D. Weber, and C. Ares. External radiotherapy for prostate cancer with or without androgen deprivation : Geneva, 1991 to 2004. *Swiss medical weekly*, 139(35-36) :511–7, Sept. 2009.
- [108] B. Muthén and K. Shedden. Finite mixture modeling with mixture outcomes using the em algorithm. *Biometrics*, 55 :463–469, 1999.
- [109] M. F. O’Brien, A. M. Cronin, P. A. Fearn, B. Smith, J. Stasi, B. Guillonneau, P. T. Scardino, J. A. Eastham, A. J. Vickers, and H. Lilja. Pretreatment prostate-specific antigen (psa) velocity and doubling time are associated with outcome but neither improves prediction of outcome beyond pretreatment psa alone in patients treated with radical prostatectomy. *Journal of Clinical Oncology*, 27(22) :3591–3597, 2009.
- [110] O. T. Okotie, K. A. Roehl, M. Han, S. Loeb, S. N. Gashti, and W. J. Catalona. Characteristics of prostate cancer detected by digital rectal examination only. *Urology*, 70(6) :1117–1120, 2007.
- [111] L. Parast, S.-C. Cheng, and T. Cai. Landmark prediction of long-term survival incorporating short-term event time information. *Journal of the American Statistical Association*, 107(500) :1492–1501, Aug. 2012.
- [112] D. K. Pauler and D. M. Finkelstein. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in medicine*, 21(24) :3897–911, Dec. 2002.
- [113] H. Payne, A. Khan, S. Chowdhury, and R. Davda. Hormone therapy for radiorecurrent prostate cancer. *World journal of urology*, Sept. 2012.
- [114] S. T. Peeters, W. D. Heemsbergen, P. C. Koper, W. L. van Putten, A. Slot, M. F. Dielwart, J. M. Bonfrer, L. Incrocci, and J. V. Lebesque. Dose-response in radiotherapy for localized prostate cancer : Results of the dutch multicenter randomized phase iii trial comparing 68 gy of radiotherapy with 78 gy. *Journal of Clinical Oncology*, 24(13) :1990–1996, 2006.
- [115] M. J. Pencina, R. B. D’Agostino, R. B. D’Agostino, and R. S. Vasan. Evaluating the added predictive ability of a new marker : From area under the roc curve to reclassification and beyond. *Statistics in Medicine*, 27(2) :157–172, 2008.

-
- [116] A. V. Piccorelli and M. D. Schluchter. Jointly modeling the relationship between longitudinal and survival data subject to left truncation with applications to cystic fibrosis. *Statistics in Medicine*, 31(29) :3931–3945, 2012.
- [117] T. Pickles, C. Kim-Sing, W. Morris, S. Tyldesley, and C. Paltiel. Evaluation of the houston biochemical relapse definition in men treated with prolonged neoadjuvant and adjuvant androgen ablation and assessment of follow-up lead-time bias. *International Journal of Radiation Oncology Biology Physics*, 57(1) :11–18, 2003.
- [118] A. Pollack, G. K. Zagars, and V. S. Kavadi. Prostate specific antigen doubling time and disease relapse after radiotherapy for prostate cancer. *Cancer*, 74(2) :670–678, 1994.
- [119] C. Pound, A. Partin, M. Eisenberger, D. Chan, J. Pearson, and P. Walsh. Natural history of progression after psa elevation following radical prostatectomy. *Journal of the American Medical Association*, 281(17) :1591–1597, 1999.
- [120] R. L. Prentice. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2) :331–342, 1982.
- [121] 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 (Oxford, England)*, 10(3) :535–49, July 2009a.
- [122] C. Proust-Lima, L. Letenneur, and H. Jacqmin-Gadda. A nonlinear latent class model for joint analysis of multivariate longitudinal data and a binary outcome. *Statistics in Medicine*, 26(10) :2229–2245, 2007.
- [123] C. Proust-Lima, J. M. G. Taylor, W. Scott, D. Ankerst, N. Liu, L. Kestin, B. K. and S. Howard. Determinants of change in prostate-specific antigen over time and its association with recurrence after external beam radiation therapy for prostate cancer in five large cohorts. *International Journal of Radiation Oncology Biology Physics*, 72(3) :782–791, aug 2008.
- [124] C. Proust-Lima, P. Joly, J. F. Dartigues, and H. Jacqmin-Gadda. Joint modelling of multivariate longitudinal outcomes and a time-to-event : A nonlinear latent class approach. *Computational Statistics & Data Analysis*, 53(4) :1142–1154, Feb. 2009b.
- [125] C. Proust-Lima, J. M. G. Taylor, S. Sécher, H. Sandler, L. Kestin, T. Pickles, K. Bae, R. Allison, and S. Williams. Confirmation of a low α/β ratio for prostate cancer treated by external beam radiation therapy alone using a post-treatment repeated-measures model for psa dynamics. *International journal of radiation oncology, biology, physics*, 79(1) :195–201, Jan. 2011.
- [126] C. Proust-Lima, M. Sène, J. M. G. Taylor, and H. Jacqmin-Gadda. Joint latent class models for longitudinal and time-to-event data : A review. *Statistical methods in medical research*, Apr. 2012.

- [127] S. J. Ratcliffe, W. Guo, and T. R. Ten Have. Joint Modeling of Longitudinal and Survival Data via a Common Frailty. *Biometrics*, 60 :892–899, dec 2004.
- [128] J. P. Richie, W. J. Catalona, F. R. Ahmann, M. A. Hudson, P. T. Scardino, R. C. Flanigan, J. B. DeKernion, T. L. Ratliff, L. R. Kavoussi, and B. L. Dalkin. Effect of patient age on early detection of prostate cancer with serum prostate-specific antigen and digital rectal examination. *Urology*, 42(4) :365–74, Oct. 1993.
- [129] J. P. Richie, L. R. Kavoussi, G. T. Ho, M. A. Vickers, M. A. O’Donnell, D. St Laurent, A. Chen, D. S. Goldstein, and K. R. Loughlin. Prostate cancer screening : role of the digital rectal examination and prostate-specific antigen. *Annals of surgical oncology*, 1(2) :117–20, Mar. 1994.
- [130] D. Rizopoulos. JM : An R package for the joint modelling of longitudinal and time-to-event data. *Journal Of Statistical Software*, 35(9), 2010.
- [131] D. Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3) :819–29, Sept. 2011b.
- [132] D. Rizopoulos. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics and Data Analysis*, 56(3) :491–501, 2012.
- [133] D. Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data : With Applications in R*. Chapman & Hall/CRC Biostatistics Series, 2012.
- [134] D. Rizopoulos and P. Ghosh. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine*, 30 (12) :1366–80, May 2011a.
- [135] D. Rizopoulos, G. Verbeke, and E. Lesaffre. Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 71(3) :637–654, June 2009.
- [136] D. Rizopoulos, L. A. Hatfield, B. P. Carlin, and J. J. Takkenberg. Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging. *arXiv :1303.2797 [stat.AP]*, 2013.
- [137] M. Roach III, G. Hanks, H. Thames Jr., P. Schellhammer, W. U. Shipley, G. H. Sokol, and H. Sandler. Defining biochemical failure following radiotherapy with or without hormonal therapy in men with clinically localized prostate cancer : Recommendations of the rtog-astro phoenix consensus conference. *International Journal of Radiation Oncology Biology Physics*, 65(4) :965–974, 2006.
- [138] M. Roach III, V. Weinberg, H. Sandler, and I. Thompson. Staging for prostate cancer. *Cancer*, 109(2) :213–220, 2007.

- [139] C. J. Rosser, D. A. Kuba, L. B. Levy, A. Chichakli, Ramsey nd Pollack, A. K. Lee, and L. L. Pisters. Prostate specific antigen bounce phenomenon after external beam radiation for clinically localized prostate cancer. *The Journal of Urology*, 168(5) : 2001–2005, 2002.
- [140] 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. *Statistics in Medicine*, 23(6) :907–926, 2004.
- [141] D. B. Rubin. Inference and missing data. *Biometrika*, 63(3) :581–592, 1976.
- [142] A. Ruffion, X. Rebillard, and F. Grima. Temps de doublement du psa et son calcul. *Progrès en Urologie*, pages 1035–1041, 2005.
- [143] W. A. Sakr, G. P. Haas, B. F. Cassin, J. E. Pontes, and J. D. Crissman. The frequency of carcinoma and intraepithelial neoplasia of the prostate in young male patients. *The Journal of urology*, 150(2 Pt 1) :379–85, Aug. 1993.
- [144] L. Salomon, D. Azria, C. Bastide, P. Beuzeboc, L. Cormier, F. Cornud, D. Eiss, P. Eschwège, N. Gaschignard, C. Hennequin, V. Molinié, P. Mongiat Artus, J.-L. Moreau, M. Péneau, M. Peyromaure, V. Ravery, X. Rebillard, P. Richaud, P. Rischmann, F. Rozet, F. Staerman, A. Villers, and M. Souli. Recommandations en onco-urologie 2010 : Cancer de la prostate. *Progrès en Urologie*, 20, Supplement 4 (0) :S217–S251, 2010.
- [145] H. M. Sandler, R. L. Dunn, P. McLaughlin, J. A. Hayman, M. A. Sullivan, and J. M. Taylor. Overall survival after prostate-specific-antigen-detected recurrence following conformal radiation therapy. *International Journal of Radiation Oncology Biology Physics*, 48(3) :629–633, Oct. 2000.
- [146] C. I. Sartor, M. H. Strawderman, X.-H. Lin, K. E. Kish, P. W. McLaughlin, and H. M. Sandler. Rate of {PSA} rise predicts metastatic versus local recurrence after definitive radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 38(5) :941–947, 1997.
- [147] M. Schemper and R. Henderson. Predictive accuracy and explained variation in cox regression. *Biometrics*, 56(1) :249–55, Mar. 2000.
- [148] R. Schoop, E. Graf, and M. Schumacher. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics*, 64(2) :603–10, June 2008.
- [149] R. Schoop, M. Schumacher, and E. Graf. Measures of prediction error for survival data with longitudinal covariates. *Biometrical journal. Biometrische Zeitschrift*, 53 (2) :275–93, Mar. 2011.
- [150] F. H. Schröder, J. Hugosson, M. J. Roobol, T. L. Tammela, S. Ciatto, V. Nelen, M. Kwiatkowski, M. Lujan, H. Lilja, M. Zappa, L. J. Denis, F. Recker, A. Berenguer, L. MÃdÃdttÃdnen, C. H. Bangma, G. Aus, A. Villers, X. Rebillard, T. van der

- Kwast, B. G. Blijenberg, S. M. Moss, H. J. de Koning, and A. Auvinen. Screening and prostate-cancer mortality in a randomized european study. *New England Journal of Medicine*, 360(13) :1320–1328, 2009.
- [151] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6 :461–464, 1978.
- [152] M. Sène, C. A. Bellera, and C. Proust-Lima. Shared random-effect models for the joint analysis of longitudinal and time-to-event data : application to the prediction of prostate cancer recurrence. *Journal de la Société Française de Statistique*, 2013.
- [153] M. Sengoz, U. Abacioglu, I. Cetin, and L. Turkeri. {PSA} bouncing after external beam radiation for prostate cancer with or without hormonal treatment. *European Urology*, 43(5) :473–477, 2003.
- [154] Z. Shi, C. B. Pinnock, S. Kinsey-Trotman, M. Borg, K. L. Moretti, S. Walsh, and T. Kopsaftis. Prostate-specific antigen (psa) rate of decline post external beam radiotherapy predicts prostate cancer death. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*, 2013.
- [155] W. Shipley, H. Thames, H. Sandler, and et al. Radiation therapy for clinically localized prostate cancer : A multi-institutional pooled analysis. *Journal of the American Medical Association*, 281(17) :1598–1604, 1999.
- [156] M. Sobin, LH abd Gospodariwicz and C. e. Wittekind. Tnm classification of malignant tumors. uicc international union against cancer. 7th edition. *Wiley-Blackwell*, pages 243–8, 2009.
- [157] I. Sousa. A review on joint modeling of longitudinal measurements and time-to-event. *Revstat*, 9(1) :57–81, 2011.
- [158] G. D. Steinberg, G. T. Bales, and C. B. Brendler. An analysis of watchful waiting for clinically localized prostate cancer. *The Journal of Urology*, 159(5) :1431–1436, 1998.
- [159] M. J. Sweeting and S. G. Thompson. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5) :750–763, 2011.
- [160] J. d. D. Tapsoba, S.-M. Lee, and C. Y. Wang. Joint modeling of survival time and longitudinal data with subject-specific change-points in the covariates. *Statistics in Medicine*, 30(3) :232–249, 2011.
- [161] J. M. G. Taylor, M. Yu, and H. M. Sandler. Individualized predictions of disease progression following radiation therapy for prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 23(4) :816–25, Feb. 2005.

-
- [162] J. M. G. Taylor, Y. Park, D. P. Ankerst, C. Proust-Lima, S. Williams, L. Kestin, K. Bae, T. Pickles, and H. Sandler. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, 69(1) :206–213, 2013a.
- [163] J. M. G. Taylor, J. Shen, E. H. Kennedy, L. Wang, and D. E. Schaibel. Comparison of methods for estimating the effect of salvage therapy in prostate cancer when treatment is given by indication. *Statistics in Medicine*, July 2013b.
- [164] H. Thames, D. Kuban, L. Levy, E. M. Horwitz, P. Kupelian, A. Martinez, J. Michalski, T. Pisansky, H. Sandler, W. Shipley, M. Zelefsky, and A. Zietman. Comparison of alternative biochemical failure definitions based on clinical outcome in 4839 prostate cancer patients treated by external beam radiotherapy between 1986 and 1995. *International Journal of Radiation Oncology Biology Physics*, 57(4) :929–943, 2003.
- [165] R. Thiébaud, H. Jacqmin-Gadda, A. Babiker, and D. Commenges. Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of cd4+ cell count and hiv rna viral load in response to treatment of hiv infection. *Statistics in Medicine*, 24(1) :65–82, 2005.
- [166] I. Thompson, D. Ankerst, C. Chi, and et al. Operating characteristics of prostate-specific antigen in men with an initial psa level of 3.0 ng/ml or lower. *Journal of the American Medical Association*, 294(1) :66–70, 2005.
- [167] a. Toledano, R. Chiche, H. Lamallem, a. Kanoui, S. Beley, F. Thibault, and P. Sèbe. Elevation of psa after prostate radiotherapy : rebound or biochemical recurrence ?. *Progrès en urologie : journal de l'Association française d'urologie et de la Société française d'urologie*, 18(9) :557–61, Oct. 2008.
- [168] A. A. Tsiatis and M. Davidian. Joint modeling of longitudinal and time-to-event data : An overview. *Statistica Sinica*, 14 :809–834, 2004.
- [169] A. A. Tsiatis, D. Victor, and M. S. Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90(429), Mar. 1995.
- [170] R. K. Valicenti, K. Bae, J. Michalski, H. Sandler, W. Shipley, A. Lin, and J. Cox. Does hormone therapy reduce disease recurrence in prostate cancer patients receiving dose-escalated radiation therapy? an analysis of radiation therapy oncology group 94-06. *International journal of radiation oncology, biology, physics*, 79(5) :1323–9, Apr. 2011.
- [171] H. C. Van Houwelingen. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, 34(1) :70–85, 2007.
- [172] G. Verbeke and E. Lesaffre. A linear mixed effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association*, 91 : 217–221, 1996.

- [173] G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer, 2000.
- [174] A. J. Vickers, C. Savage, M. F. O'Brien, and H. Lilja. Systematic review of pretreatment prostate-specific antigen velocity and doubling time as predictors for prostate cancer. *Journal of Clinical Oncology*, 27(3) :398–403, 2009.
- [175] S. Viviani, M. Alfó, and D. Rizopoulos. Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing*, pages 1–11, 2013.
- [176] Y. Wang and J. M. G. Taylor. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96(455), Sept. 2001.
- [177] S. J. Welsh and G. Powis. *Personalized Cancer Medicine*. Schwab M. (Ed.) Encyclopedia of Cancer, 2009.
- [178] P. Williamson, P. Philipson, and A. Marson. Joint modelling of longitudinal and competing risks data. *Statistics in medicine*, 27(Aug.) :6426–6438, 2008.
- [179] L. Wu, W. Liu, G. Y. Yi, and Y. Huang. Analysis of longitudinal and survival data : Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012.
- [180] M. Wu and K. Bailey. Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, 7 :337–346, 1988.
- [181] M. Wu and K. Bailey. Estimation and comparison of changes in the presence of informative right censoring : conditional linear model. *Biometrics*, 45 :939–955, 1989.
- [182] M. C. Wu and R. J. Carrol. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1) : 175–188, 1988.
- [183] M. S. Wulfsohn and A. A. Tsiatis. A joint model of survival and longitudinal data measured with error. *Biometrics*, 53 :330–339, Mar. 1997.
- [184] J. Xu and S. L. Zeger. Joint analysis of longitudinal data comprising repeated measures and times to events. *Pharmaceutical Research*, pages 375–387, 2001.
- [185] W. Ye, X. Lin, and J. M. G. Taylor. Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach. *Biometrics*, 64(4) :1238–46, Dec. 2008.
- [186] M. Yu, N. J. Law, J. M. G. Taylor, and H. M. Sandler. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 14 :835–862, 2004.

- [187] M. Yu, J. M. G. Taylor, and H. M. Sandler. Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association*, 103(481) :178–187, Mar. 2008.
- [188] G. Zagars, A. Pollack, V. Kavadi, and A. von Eschenbach. Prostate-specific antigen and radiation therapy for clinically localized prostate cancer. *International Journal of Radiation Oncology Biology Physics*, 32(2) :293–306, 1995.
- [189] D. Zeng and J. Cai. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics*, 33(5) :2132–2163, Oct. 2005.
- [190] J. J. Zhang and M. Wang. Latent class joint model of ovarian function suppression and dfs for premenopausal breast cancer patients. *Statistics in Medicine*, 29(22) : 2310–2324, 2010.
- [191] M. Zhang and Q. Cheng. Determine the number of components in a mixture model by the extended KS test. *Pattern Recognition Letters*, 25(2) :211–216, 2004.
- [192] Y. Zheng and P. J. Heagerty. Prospective accuracy for longitudinal markers. *Biometrics*, 63(2) :332–341, June 2007.

Annexe

Joint latent class models for longitudinal and time-to-event data: A review

Cécile Proust-Lima,^{1,2} Mbéry Séne,^{1,2}
Jeremy MG Taylor^{3,4} and Hélène Jacqmin-Gadda^{1,2}

Statistical Methods in Medical Research
0(0) 1–17

© The Author(s) 2012

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0962280212445839

smm.sagepub.com



Abstract

Most statistical developments in the joint modelling area have focused on the shared random-effect models that include characteristics of the longitudinal marker as predictors in the model for the time-to-event. A less well-known approach is the joint latent class model which consists in assuming that a latent class structure entirely captures the correlation between the longitudinal marker trajectory and the risk of the event. Owing to its flexibility in modelling the dependency between the longitudinal marker and the event time, as well as its ability to include covariates, the joint latent class model may be particularly suited for prediction problems. This article aims at giving an overview of joint latent class modelling, especially in the prediction context. The authors introduce the model, discuss estimation and goodness-of-fit, and compare it with the shared random-effect model. Then, dynamic predictive tools derived from joint latent class models, as well as measures to evaluate their dynamic predictive accuracy, are presented. A detailed illustration of the methods is given in the context of the prediction of prostate cancer recurrence after radiation therapy based on repeated measures of Prostate Specific Antigen.

Keywords

Brier score, joint model, longitudinal data, mixture model, predictive accuracy, prognosis, prostate cancer

1 Introduction

It is frequent in longitudinal studies to collect both repeated measures of a longitudinal marker and the time to an event of interest. Common examples include trajectory of CD4 counts and time to AIDS in HIV studies¹ or trajectory of Prostate Specific Antigen (PSA) and risk of prostate cancer recurrence.^{2,3} In these examples, both quantities are linked so that their joint analysis is required for addressing different objectives. First, the interest can be on the prognostic value of the longitudinal

¹INSERM, U897, Epidemiology and Biostatistics Research Center, F-33076 Bordeaux, France

²Université Bordeaux Segalen, ISPED, F-33076 Bordeaux, France

³Department of Biostatistics, University of Michigan, Ann Arbor, MI

⁴Department of Radiation Oncology, University of Michigan, Ann Arbor, MI

Corresponding author:

Cécile Proust-Lima, INSERM U897, ISPED, Université de Bordeaux Segalen, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France
Email: cecile.proust-lima@inserm.fr

marker, as in prostate cancer with the prognostic value of PSA trajectory on the risk of recurrence. Joint modelling of the two quantities corrects for biases induced by the random measurement errors and the intermittent measurement of the marker.⁴ In other applications, the interest is on the marker trajectory during the course of a disease, and the joint model corrects for bias induced by the occurrence of the event. Finally, joint models are required when the interest is specifically in understanding how the repeated marker data and the risk of event are linked. For example, investigating what is the link between the PSA trajectory and the subsequent risk of prostate cancer recurrence is of importance to understand the dynamics of the disease and provide powerful dynamic prognostic tools.⁵

The joint modelling approach consists in defining: (1) a model for the time-to-event, usually a proportional hazard model, (2) a model for the marker trajectory, usually a mixed model, and (3) linking both models using a shared latent structure.⁶ In this context, most developments have focused on the shared random-effect model (SREM), also called a selection model in missing data problems, in which a characteristic of the longitudinal process defined as a function of the random-effects is included as a covariate in the survival model.¹ This function can be any function capturing the dynamics of the marker trajectory, such as the individual deviation from the mean trajectory or the expected individual current level of the marker. The latter extends directly the standard survival model with time-dependent covariate. This model has been used to evaluate the association between a marker trajectory and a time-to-event,⁵ to make dynamic predictions,⁷ and was extended in different ways: for example to multiple time-to-events,⁸ to multiple longitudinal markers,⁹ and to include a cured fraction.³

An alternative approach for joint modelling a marker trajectory and the time to an event is inspired by finite mixture (also called mixture-of-experts) modelling.¹⁰ This method, called a joint latent class model (JLCM), considers the population of subjects as heterogeneous, and assumes that it consists of homogeneous latent subgroups of subjects that share the same marker trajectory and the same risk of the event.^{2,11,12} This assumption of heterogeneity is frequently relevant in medical research where several differing profiles of patients are expected. For example, in prostate cancer progression after treatment, different profiles of PSA are observed. Compared to the shared random-effect model, the joint latent class model has received less attention with only a few applications, for the description of disease progression,^{11–13} for sensitivity analyses in missing data problems,^{14,15} and recently for dynamic prediction.² Yet, the JLCM offers a computationally attractive alternative to the SREM and it is based on different assumptions regarding the link between the longitudinal and event time components of the model. The link between the two components needs to be more precisely defined (through functions of the marker trajectory) in the SREM than in the JLCM. As a consequence, while the JLCM may not be suited to evaluate specific assumptions regarding the characteristics of the marker trajectory that are the most influential on the event risk, it may be of interest when: (1) developing predictive joint models or (2) investigating the link between the longitudinal marker and the time-to-event without specific assumptions, especially in a heterogeneous population.

Other special joint models can be found in the literature that we do not describe further in the present work, such as joint models based on pattern-mixture modelling¹⁶ or ‘simple transformation models’ assuming for example a multivariate Gaussian distribution for the longitudinal data and the logarithm of the time-to-event.¹⁷

The aim of this article is to introduce the joint latent class model and review methods to evaluate its goodness-of-fit and its predictive accuracy in the context of dynamic predictive tool development. Specific aspects of the JLCM as well as differences with the SREM are illustrated through an application in prostate cancer where the objective was to validate a dynamic prognostic tool of

prostate cancer recurrence based on the post-treatment PSA trajectory. Section 2 focuses on the joint model specification. Section 3 is dedicated to goodness-of-fit techniques while Section 4 describes dynamic predictive tools development and their predictive accuracy assessment both for the JLCM and the SREM. Section 5 illustrates the different methods through the prostate cancer example. Some concluding remarks are given in Section 6.

2 Joint latent class model

2.1 Latent class membership probability

Assume a population of N subjects that can be divided into a finite number G of latent homogeneous subgroups. The latent class membership for each subject i ($i = 1, \dots, N$) is defined using a categorical latent variable c_i , which equals g if subject i belongs to latent class g ($g = 1, \dots, G$). An individual has a probability π_{ig} of belonging to latent class g , which is modelled using a multinomial logistic regression according to covariates X_{pi} :

$$\pi_{ig} = P(c_i = g | X_{pi}) = \frac{e^{\xi_{0g} + X_{pi}^T \xi_{1g}}}{\sum_{l=1}^G e^{\xi_{0l} + X_{pi}^T \xi_{1l}}} \quad (1)$$

where ξ_{0g} is the intercept for class g and ξ_{1g} is the vector of class-specific parameters associated with the vector of time-independent covariates X_{pi} . For identifiability, $\xi_{0G} = 0$ and $\xi_{1G} = 0$.

Each latent class is characterised by a class-specific marker trajectory and a class-specific risk of the event, and the marker and the time-to-event are assumed to be conditionally independent given these latent classes. This conditional independence is a central assumption of the JLCM.

2.2 Class-specific marker trajectory

Given the latent class g , the vector of repeated measures of the longitudinal marker $Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{ij}), \dots, Y_i(t_{in_j}))$ is described at the different times of measurement t_{ij} ($j = 1, \dots, n_i$) by a standard linear mixed model¹⁸:

$$Y_i(t_{ij}) |_{c_i=g} = Z_i(t_{ij})^T u_{ig} + X_{li}(t_{ij})^T \beta_g + \epsilon_i(t_{ij}) \quad (2)$$

where the p -vector of class-specific random-effects $u_{ig} = u_i |_{c_i=g} \sim \mathcal{N}(\mu_g, B_g)$ or equivalently, the vector of random-effects $u_i \sim \sum_{g=1}^G \pi_{ig} \mathcal{N}(\mu_g, B_g)$ with π_{ig} defined in equation (1). The n_i -vector of measurement errors $\epsilon_i = (\epsilon_i(t_{i1}), \dots, \epsilon_i(t_{in_i}))^T \sim \mathcal{N}(0, \Sigma_i)$. The variance-covariance matrix B_g can be common over classes or class-specific. However, when considered as class-specific, usually $B_g = \omega_g^2 B$ with B unstructured and $\omega_G = 1$ to limit the number of parameters and identifiability concerns. The variance-covariance matrix Σ_i is usually restricted to the diagonal matrix $\sigma^2 I_{n_i}$ for homoscedastic independent errors but ϵ_i can also include a correlation process such as a Brownian motion or an auto-regressive process. The p -vector of time-dependent covariates $Z_i(t_{ij})$, that may include any function of time, is associated with the p -vector of random-effects u_{ig} . The q -vector of possibly time-dependent covariates $X_{li}(t_{ij})$ is associated with the possibly class-specific q -vector of fixed parameters β_g . No overlap between $Z_i(t_{ij})$ and $X_{li}(t_{ij})$ is assumed for identifiability.

2.3 Class-specific risk of event

Let T_i^* denote the time-to-event of interest, C_i the censoring time, $T_i = \min(T_i^*, C_i)$ and $E_i = 1_{T_i^* \leq C_i}$. Given the latent class g , the risk of event can be described using any survival model. For simplicity we consider here a proportional hazard model:

$$\lambda_i(t \mid c_i = g; \zeta_g, \delta_g) = \lambda_{0g}(t; \zeta_g) e^{X_{ei}(t)^T \delta_g} \quad (3)$$

where $X_{ei}(t)$ is the r -vector of (possibly time-dependent) covariates associated with the r -vector of parameters δ_g . The class-specific baseline hazard is $\lambda_{0g}(t; \zeta_g)$. Either a baseline hazard stratified on the latent class structure or baseline hazards proportional in each latent class ($\lambda_{0g}(t) = \lambda_0(t) e^{\zeta_g}$ with $\zeta_G = 0$) can be considered. To remain in the likelihood framework, only parametric hazard functions ($\lambda_{0g}(t)$ or $\lambda_0(t)$) are considered here, such as Weibull, piecewise constant or M-splines.¹²

2.4 Maximum likelihood estimation

For a fixed number of latent classes G , the log-likelihood $L(\theta_G)$ of the observed data can be decomposed using the conditional independence assumption so that:

$$L(\theta_G) = \sum_{i=1}^N L_i = \sum_{i=1}^N \log \left(\sum_{g=1}^G \pi_{ig} f(Y_i \mid c_i = g; \theta_G) \lambda_i(T_i \mid c_i = g; \theta_G)^{E_i} S_i(T_i \mid c_i = g; \theta_G) \right) \quad (4)$$

where θ_G is the entire vector of parameters for a JLCM with G classes; the class-membership probability π_{ig} is defined in (1); the instantaneous risk $\lambda_i(T_i \mid c_i = g; \theta_G)$ is defined in (3) and $S_i(T_i \mid c_i = g; \theta_G)$ is the corresponding class-specific survival function. The density $f(Y_i \mid c_i = g; \theta_G)$ of the longitudinal marker in class g is multivariate normal with mean $Z_i \mu_g + X_{li} \beta_g$ and covariance matrix $Z_i B_g Z_i^T + \Sigma_i$; Z_i and X_{li} being respectively the $n_i \times p$ and $n_i \times q$ matrices of j th row vectors $Z_i(t_{ij})^T$ and $X_{li}(t_{ij})^T$.

The JLCM estimation is implemented in the `Jointlcm` function of the `lcm` R package (<http://cran.r-project.org/web/packages/lcm>). The log-likelihood (4) is maximised using a Marquardt algorithm¹⁹ with stringent convergence criteria. In addition to parameter stability and log-likelihood stability, convergence is reached only when $d^T H^{-1} d < \epsilon$, where d is the gradient vector and H the Hessian matrix (by default $\epsilon = 10^{-4}$). The inverse of the Hessian matrix also provides estimates of the parameter variances.

Several difficulties arise in mixture model estimation. First, a permutation of the latent classes parameters in θ_G gives the same likelihood. Although this phenomenon, called ‘label switching’, may pose problems in Bayesian estimation,²⁰ it is not a concern for maximum likelihood estimation.²¹ Second, the likelihood in mixture problems may have multiple local maxima, so that it is highly recommended to run the algorithm starting from several sets of initial values to ensure convergence to the global maximum.^{22,23} Third, in some contexts, a lack of information in the data may result in difficulties fitting a latent class model. This is not the case with joint latent class model in which the latent class structure is based on a large amount of information with both continuous repeated data and a time-to-event. Finally, the likelihood maximisation is performed for a fixed number of latent classes, and the optimal number of latent classes is most often determined using the Bayesian Information Criterion (BIC) which is the preferred criterion in mixture models²⁴: $BIC(G) = -2L(\theta_G) + n_\theta \log(N)$ with n_θ the number of estimated parameters. Other criteria are discussed in Han et al.²⁵

We note that identifiability of finite mixture models was extensively discussed in Redner and Homer²¹ and that the estimation procedure for JLCM was specifically validated in a simulation study.¹²

2.5 Differences between joint latent class and shared random-effect models

Before comparing the two types of models, we first give a brief description of the SREM following closely the formulation in Wulfsohn and Tsiatis¹ and Rizopoulos.²⁶ In a SREM, the repeated measures of the longitudinal marker $Y_i(t_{ij})$ at time t_{ij} (for $j = 1, \dots, n_i$) are described by a standard linear mixed model¹⁸:

$$Y_i(t_{ij}) = Z_i(t_{ij})^T u_i + X_{it}(t_{ij})^T \beta + \epsilon_i(t_{ij}) \quad (5)$$

where $Z_i(t_{ij})$, $\epsilon_i(t_{ij})$ and $X_{it}(t_{ij})$ are defined above, and $u_i \sim \mathcal{N}(\mu, B)$ is the p -vector of random-effects. We assume a proportional hazard model for the risk of the event:

$$\lambda_i(t | u_i; \zeta, \delta, \eta) = \lambda_0(t; \zeta) e^{X_{ei}(t)^T \delta + f(u_i, \beta, Z_i(t), X_{it}(t))^T \eta} \quad (6)$$

where $\lambda_0(t; \zeta)$ and $X_{ei}(t)$ are defined as above and $f(u_i, \beta, Z_i(t), X_{it}(t))$ represents a univariate or multivariate function of the subject-specific random-effects, such as the subject-specific current mean marker level or/and slope, as implemented in JM R package.²⁶

The JLCM and the SREM have several differences. First, though both methods account for variability of the longitudinal profiles through the random-effects, the JLCM further accounts for heterogeneity of the population: through the latent classes, it assumes a heterogeneous population of subjects with each population having a different average profile of the marker and different risk of the event. In contrast, the SREM assumes a homogeneous population with a single average trajectory of the marker equation (5) and a continuous relationship between the marker and the risk of the event in equation (6).

Second, in the SREM, the same random-effects u_i influence both the correlation between repeated measures of the marker and the dependency between the marker and the time-to-event. In contrast, these two dependence structures are separated in the JLCM where the random-effects in (2) only account for the correlation between repeated measures while the latent classes account for the dependency between the marker and the event.

Third, in the SREM, the characteristics of the marker trajectory that influence the risk of event are chosen a priori through the function $f(u_i, \beta, Z_i(t), X_{it}(t))$ included in the survival model. Comparing models with different functions $f(u_i, \beta, Z_i(t), X_{it}(t))$ is of interest when evaluating specific assumptions regarding the dependency between the two processes but this constraint can turn out to be too limited when the focus is on finding the best prediction model for the event. In contrast, the JLCM makes less precise assumptions on the link between the marker trajectory and the time-to-event, and a stratification over the latent classes allows the baseline risk of the event to vary flexibly according to the marker when the number of classes becomes large. However, this is achieved at the cost of a potentially large increase of the number of parameters, and sometimes a more difficult interpretation of the latent classes and the parameters within each class. Moreover, due to the conditional independence assumption, the JLCM assumes that inside a latent class and conditional on covariates, the risk of event is independent of the marker level. All of this taken into consideration, the JLCM is designed to describe the observations without specific a priori assumptions, and thus is potentially well suited for the purposes of prediction.

Lastly, regarding the estimation process, the log-likelihood computation for SREM requires a numerical integration over the random-effect distribution.^{1,26} In contrast, this integration is replaced in the JLCM by a sum over the latent classes, which is considerably easier computationally. However, the JLCM estimation needs to be repeated several times to ensure convergence to the global maximum and to choose the number of latent classes.

3 Goodness-of-fit evaluation

There has been limited study of goodness-of-fit of joint models in the literature. For the shared random-effect model, a few papers explored residual techniques.^{26,27} There has been more research for the joint latent class model with different aspects explored including: longitudinal and survival predictions,^{12,13} posterior classification of the subjects^{2,14} and the conditional independence assumption.^{28,29} In the JLCM framework, goodness-of-fit methods are needed not only to validate a specific model but also to guide the selection of the number of latent classes. In general, the choice of the number of classes should be based on a number of considerations, i.e. not only the smallest information criterion, but also a good discrimination between classes, correct predictions, satisfactory conditional independence and meaningful latent classes.

3.1 Posterior classification

A specific aspect of the JLCM is that posterior classification can be obtained from the posterior estimates of the latent class membership probabilities:

$$\begin{aligned}\hat{\pi}_{ig}^{Y,T} &= P(c_i = g \mid Y_i, (T_i, E_i); \hat{\theta}_G) \\ &= \frac{\hat{\pi}_{ig} f(Y_i \mid c_i = g; \hat{\theta}_G) \lambda_i(T_i \mid c_i = g; \hat{\theta}_G)^{E_i} S_i(T_i \mid c_i = g; \hat{\theta}_G)}{\sum_{l=1}^G \hat{\pi}_{il} f(Y_i \mid c_i = l; \hat{\theta}_G) \lambda_i(T_i \mid c_i = l; \hat{\theta}_G)^{E_i} S_i(T_i \mid c_i = l; \hat{\theta}_G)}\end{aligned}\quad (7)$$

where $\hat{\pi}_{ig}$ is the latent-class probability defined in equation (1) and computed at the parameter estimates $\hat{\theta}_G$; and $\lambda_i(T_i \mid c_i = g; \theta_G)$, $S_i(T_i \mid c_i = g; \theta_G)$ and $f(Y_i \mid c_i = g; \theta_G)$ are defined as in Section 2.4.

From these probabilities, each subject is classified in the class for which he has the highest posterior probability of belonging $\tilde{c}_i^{Y,T} = \text{argmax}_g(\hat{\pi}_{ig}^{Y,T})$. The more discriminatory the posterior classification is, the better the model. Discrimination of the latent classes can be considered in several ways: (1) the proportion of subjects with their maximal posterior latent class membership probability above a certain level, e.g. above 0.8 or above 0.9,¹⁴ (2) the measure of entropy $1 - \frac{\text{En}}{N \ln G}$ (where $\text{En} = -\sum_{i=1}^N \sum_{g=1}^G \hat{\pi}_{ig}^{Y,T} \log(\hat{\pi}_{ig}^{Y,T})$) that indicates a clear classification when close to one,³⁰ or (3) the posterior classification table¹² that provides the mean of the posterior probabilities for subjects classified in each class. For the latter, a very discriminatory classification would have diagonal terms close to 1 and non-diagonal terms close to 0.

3.2 Fitted values and comparison with observed values

As in any mixed model, marginal and subject-specific predictions from a JLCM (respectively $Y^{(M)}$ and $Y^{(SS)}$) can be computed and compared to observed data. The difference is that predictions are class-specific: for subject i , occasion j and class g , $\hat{Y}_{ijg}^{(M)} = Z_i(t_{ij})^T \hat{\mu}_g + X_{li}(t_{ij})^T \hat{\beta}_g$ and $\hat{Y}_{ijg}^{(SS)} = Z_i(t_{ij})^T (\hat{\mu}_g + \hat{u}_{ig}) + X_{li}(t_{ij})^T \hat{\beta}_g$ with $\hat{u}_{ig} = \hat{B}_g Z_i^T \hat{V}_{ig}^{-1} (Y_i - X_{li} \hat{\beta}_g - Z_i \hat{\mu}_g)$, the empirical Bayes estimates of the class-specific random-effects. From these class-specific individual predictions,

either individual predictions averaged over classes $\hat{Y}_{ij}^{(\cdot)} = \sum_{g=1}^G \hat{\pi}_{ig} \hat{Y}_{ijg}^{(\cdot)}$ or class-specific predictions averaged over individuals $\hat{Y}_g(t)^{(\cdot)} = \sum_{i=1}^{N(t)} \hat{\pi}_{ig} \hat{Y}_{ijg}^{(\cdot)}$ can be computed, $N(t)$ representing the number of subjects with measurements at t (in practice, time may be discretised in intervals for this computation). Depending on the type of predictions (marginal or subject-specific), $\hat{\pi}_{ig}$ can be computed either as marginal class-membership probabilities from equation (1) or as conditional probabilities from equation (7). Both averaged predictions $\hat{Y}_{ij}^{(\cdot)}$ and $\hat{Y}_g(t)^{(\cdot)}$ are useful to evaluate the fit of the observed data, $\hat{Y}_g(t)^{(\cdot)}$ being compared with the weighted class-specific mean of observed values $\bar{Y}_g(t) = \sum_{i=1}^{N(t)} \hat{\pi}_{ig} Y_i(t)$.

Marginal and subject-specific residuals $\hat{R}_{ij}^{(\cdot)} = Y_{ij} - \hat{Y}_{ij}^{(\cdot)}$ can be also computed from individual predictions. However, as in any joint model, these residuals may suffer from non-random dropout caused by the events so that standard residual analyses do not apply. Instead, the multiple-imputation technique proposed for the SREM³¹ may be used.

The fit of the survival part of the JLCM can also be evaluated by comparing, for example, the class-specific weighted individual survival functions $\hat{S}_g(t) = \sum_{i=1}^N \hat{\pi}_{ig} S_i(t | c_i = g; \hat{\theta}_G)$, where $S_i(t | c_i = g; \hat{\theta}_G)$ is derived from the survival model (3), with the corresponding class-specific weighted Kaplan–Meier estimates.

3.3 Conditional independence assumption

The JLCM assumes that the latent class structure captures the entire dependency between the longitudinal marker and the time-to-event. Several approaches were proposed to evaluate this assumption: analysis based on the posterior classification,²⁸ analysis of the residuals conditional on the event,¹² and a score test.²⁹ In the score test, the alternative hypothesis \mathcal{H}_1 is defined by a residual dependence between the time-to-event and the marker through the random-effects in addition to the dependence through the latent classes. Thus, the model under \mathcal{H}_1 is a JLCM with shared random-effects in which equation (3) is replaced by:

$$\lambda_i(t | c_i = g, X_{ei}; \zeta_g, \delta_g) = \lambda_{0g}(t; \zeta_g) e^{X_{ei}(t)\delta_g + u_{ig}\eta} \quad (8)$$

where η is a p -vector associating the p random-effects u_{ig} of the longitudinal model (2) with the time-to-event.

From this, evaluation of the conditional independence assumption consists in testing $\mathcal{H}_0: \eta = 0$ (vs $\mathcal{H}_1: \eta \neq 0$), and the corresponding score test statistic is:

$$U = \sum_{i=1}^N \sum_{g=1}^G \hat{\pi}_{ig}^{Y,T} (E_i - \Lambda_{ig}(T_i)) \hat{u}_{ig} \quad (9)$$

where $\Lambda_{ig}(T_i)$ is the class-specific cumulative hazard.

The score test statistic is an estimate of the covariance between the martingale residuals from the survival model and the class-specific empirical Bayes estimates of the random-effect weighted by the posterior class-membership probability. Under the null hypothesis, $U^T \text{Var}(U)^{-1} U$ follows a chi-square distribution with p degrees of freedom.²⁹ This approach was found to be much more powerful than the other methods to detect any departure from the conditional independence assumption in simulations.

4 Dynamic prediction

In recent years, there has been a growing interest in predictive/prognostic tools derived from joint models.^{2,5,7} Indeed, joint models have the ability to produce predictive tools that can be dynamically updated according to the observed trajectory of the marker, and thus offer a powerful aid for clinical decision making in patient monitoring. Such dynamic predictive tools were proposed both in SREM^{5,7} and in JLCM² contexts but their evaluation is challenging. Prognostic tools evaluation is already complex in standard survival analysis due to censoring, so that only a few methods have been proposed to extend predictive accuracy assessment to joint models.³²⁻³⁴ We introduce in this section the dynamic predictive tools computation and provide a review of measures to assess their predictive accuracy in JLCM and SREM frameworks.

4.1 Individual dynamic predictions and confidence bands

The dynamic predictive tool derived from a joint model consists of the predicted probability of an event in a window $[s, s + t]$ given covariates and marker measurements collected until time s . In the following, s is called the time at prediction ($s \geq 0$), and t is called the horizon ($t \geq 0$). For any subject i , let $Y_i^{(s)}$ denote the vector of marker repeated measures until time s , X_i all the other covariates, T_i the time of event, and θ the parameter vector of the joint model. In a JLCM, the predicted probabilities of events are given by:

$$\begin{aligned} P(T_i \leq s + t \mid T_i \geq s, Y_i^{(s)}, X_i; \theta) &= \sum_{g=1}^G P(T_i \leq s + t \mid T_i \geq s, c_i = g, X_i; \theta) \times P(c_i = g \mid T_i \geq s, Y_i^{(s)}, X_i; \theta) \\ &= \frac{\sum_{g=1}^G \pi_{ig} f(Y_i^{(s)} \mid c_i = g, X_i; \theta) (S_i(s \mid c_i = g, X_i; \theta) - S_i(s + t \mid c_i = g, X_i; \theta))}{\sum_{g=1}^G \pi_{ig} f(Y_i^{(s)} \mid c_i = g, X_i; \theta) S_i(s \mid c_i = g, X_i; \theta)} \end{aligned} \quad (10)$$

Predicted probabilities of events in a SREM are similarly obtained by replacing the sum over the latent classes by an integral over the random effects distribution:

$$\begin{aligned} P(T_i \leq s + t \mid T_i \geq s, Y_i^{(s)}, X_i; \theta) &= \int_{u_i} P(T_i \leq s + t \mid T_i \geq s, u_i, X_i; \theta) \times f(u_i \mid T_i \geq s, Y_i^{(s)}, X_i; \theta) du_i \\ &= \frac{\int_{u_i} f(Y_i^{(s)} \mid u_i, X_i; \theta) (S_i(s \mid u_i, X_i; \theta) - S_i(s + t \mid u_i, X_i; \theta)) f(u_i) du_i}{\int_{u_i} f(Y_i^{(s)} \mid u_i, X_i; \theta) S_i(s \mid u_i, X_i; \theta) f(u_i) du_i} \end{aligned} \quad (11)$$

An estimate of $P(T_i \leq s + t \mid T_i \geq s, Y_i^{(s)}, X_i; \theta)$ can be obtained by replacing θ by $\hat{\theta}$. However, to obtain 95% credibility bands, its posterior distribution must be approximated by a Monte Carlo method,^{2,7} using the 2.5% and 97.5% percentiles (and possibly the median for the point estimate) of the distribution of the probabilities computed from equation (10) with a large number D of parameter vectors $(\theta_d)_{d=1, \dots, D}$ drawn from the asymptotic distribution $\mathcal{N}(\hat{\theta}, \hat{V}(\hat{\theta}))$.

4.2 Predictive accuracy measures for dynamic predictions

Two types of predictive accuracy measures were proposed for assessing dynamic predictive tools: errors of prediction^{2,32,33} and more recently, a measure derived from the theory of information, the expected prognostic observed cross-entropy.³⁴

4.2.1 Quadratic error of prediction

Let $\hat{S}(s + t|s) = 1 - P(T_i \leq s + t | T_i \geq s, Y_i^{(s)}, X_i; \hat{\theta})$ denote the predicted event-free probability, and $Y(s + t)$ the survival status at time $s + t$. The quadratic error of prediction for rule \hat{S} , also known as half the expected Brier Score (BS), is $E[(Y(s + t) - \hat{S}(s + t|s))^2]$. In the dynamic prediction context, two difficulties arise in the estimation of this quantity. First, the survival status $Y(s + t)$ may be censored, and second the error of prediction is a two-dimensional surface that needs to be summarised. To take into account censoring, two estimators were proposed. The first one (called data-based BS) consists in weighting the observations according to their probability of being observed³³:

$$\hat{BS}_1(s, t) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left[\frac{I(T_i > s + t)}{\hat{G}(s + t)/\hat{G}(s)} (1 - \hat{S}(s + t|s))^2 + \frac{E_i I(T_i \leq s + t)}{\hat{G}(T_i)/\hat{G}(s)} (0 - \hat{S}(s + t|s))^2 \right] \tag{12}$$

where N_s is the number of subjects still at risk at time s and $\hat{G}(u)$ is the survival function of the censoring distribution at time u estimated using either a Kaplan–Meier estimate³³ or a regression model.³⁵

The second estimator (called model-based BS) consists in predicting the contribution to the error of prediction of censored observations directly using the joint model^{2,32}:

$$\begin{aligned} \hat{BS}_2(s, t) = \frac{1}{N_s} \sum_{i=1}^{N_s} & \left\{ I(T_i > s + t) (1 - \hat{S}(s + t|s))^2 + E_i I(T_i \leq s + t) (0 - \hat{S}(s + t|s))^2 \right. \\ & \left. + (1 - E_i) I(T_i \leq s + t) \left[(1 - \hat{S}(s + t|s))^2 \frac{\hat{S}(s + t|s)}{\hat{S}(T_i|s)} + (0 - \hat{S}(s + t|s))^2 \left(1 - \frac{\hat{S}(s + t|s)}{\hat{S}(T_i|s)} \right) \right] \right\} \end{aligned} \tag{13}$$

While the second estimator may be biased with misspecified models, the first one may lack efficiency³⁶ and may require modelling the probability of being observed.³⁵ We provide both estimators to validate the dynamic predictive tools.

Summary measures of these 2D estimators may be useful in practice. For a given time at prediction s , we summarise the error of prediction over a $[0, \tau]$ -window of horizons using the weighted average³²:

$$\hat{IBS}_h(s) = \frac{\sum_{k=1}^{n_\tau^{(s)}} d_k^{(s)} (\hat{G}(s)/\hat{G}(t_k)) \hat{BS}_h(s, t_k - s)}{\sum_{k=1}^{n_\tau^{(s)}} d_k^{(s)} (\hat{G}(s)/\hat{G}(t_k))} \quad \text{for } h = 1, 2 \tag{14}$$

where $n_\tau^{(s)}$ is the number of different times of events in the window $(s, s + \tau)$, $d_k^{(s)}$ is the number of events at time t_k among subjects at risk at time s and $\hat{G}(u)$ is the Kaplan–Meier estimate of the censoring distribution at time u .

4.2.2 Expected prognostic observed cross-entropy

Let $f_{T|Y^{(s)}, T^* \geq s}$ denote the conditional density of the right-censored time of event $T = \min(T^*, C)$ derived from the joint model. The expected prognostic observed cross-entropy (EPOCE) is defined as $E(-\ln f_{T|Y^{(s)}, T^* \geq s} | T^* \geq s)$. The EPOCE can be estimated by leave-one-out

cross-validation.³⁴ For a fixed time at prediction s , an approximate cross-validated estimator is $CVPOL_a$ defined as:

$$CVPOL_a(s) = -\frac{1}{N_s} \sum_{i=1}^{N_s} F_i(\hat{\theta}, s) + N \text{Trace}(H^{-1}K_s) \quad (15)$$

where N_s is the number of subjects still at risk in s , H is the Hessian matrix of the joint log-likelihood, $K_s = \frac{1}{N_s(N-1)} \sum_{i=1}^N I(T_i \geq s) \hat{v}_i(s) \hat{d}_i^T$ with $\hat{v}_i(s)$ and \hat{d}_i the gradients of the individual contributions respectively to the conditional log-likelihood in s using only $Y_i^{(s)}$ and the joint log-likelihood computed in $\hat{\theta}$ using the total vector of repeated measures Y_i . Finally, F_i is the individual contribution to the conditional log-likelihood defined for $i = 1, \dots, N_s$ in respectively a JLCM and a SREM as:

$$F_i(\theta_G, s) = \ln \left(\frac{\sum_{g=1}^G \pi_{ig} f(Y_i^{(s)} | c_i = g; \theta_G) \lambda_i(T_i | c_i = g; \theta_G)^{E_i} S_i(T_i | c_i = g; \theta_G)}{\sum_{g=1}^G \pi_{ig} f(Y_i^{(s)} | c_i = g; \theta_G) S_i(s | c_i = g; \theta_G)} \right) \quad (16)$$

$$\text{and } F_i(\theta, s) = \ln \left(\frac{\int_{u_i} f(Y_i^{(s)} | u_i; \theta) \lambda_i(T_i | u_i; \theta)^{E_i} S_i(T_i | u_i; \theta) f(u_i) du_i}{\int_{u_i} f(Y_i^{(s)} | u_i; \theta) S_i(s | u_i; \theta) f(u_i) du_i} \right) \quad (17)$$

Advantages of the EPOCE over previously described measures of predictive accuracy are multiple. First, EPOCE can be estimated either on the data used for estimating the joint model thanks to the approximate cross-validation correction ($\text{Trace}(H^{-1}K_s)$) with $CVPOL_a$, or on external data like other errors of prediction using MPOL which equals $CVPOL_a$ without this correction. A correction of over-optimism using the cross-validation technique was also proposed for errors of prediction³⁷ but as no approximate formula was given, it remains too computationally demanding for joint models evaluation. Second, no assumption is made in the $CVPOL_a$ regarding the window of horizons $[0, \tau]$ evaluated nor the type of summary measure over the horizons. Third, no assumption is made regarding the censoring distribution in contrast with the estimators in equations (12) and (13). Finally, a 95% tracking interval of the difference in EPOCE between two joint models can be computed, which enables a better evaluation of whether the difference in predictive accuracy between two models is of importance (see details in Commenges et al.³⁴). Finally, since the models are fitted using the likelihood, it seems more natural to evaluate them using a method based on the log of the density, rather than a quadratic loss as in the BS criteria.

5 Application to prostate cancer

We illustrate the JLCM on prostate cancer data. The objective was to propose a dynamic prognostic tool to detect clinical recurrence of prostate cancer based on repeated measures of Prostate Specific Antigen (PSA) after external radiation beam therapy (EBRT), PSA being a well-known biomarker of prostate cancer progression routinely collected after treatment.

5.1 University of Michigan hospital cohort

The joint models were estimated using the data from the University of Michigan hospital cohort.⁵ All subjects with localised prostate cancer of stage T1 to T4, node and metastasis negative, who underwent EBRT and did not initiate any androgen deprivation therapy during the follow-up were

included. Cases were required to have at least 1 year follow-up without clinical recurrence and at least two PSA measurements before the end of follow-up. For the purposes of this paper clinical recurrence was defined as any of the following: distant metastases, nodal recurrence or any palpable or biopsy-detected local recurrence 3 years or later after radiation; any local recurrence within 3 years of EBRT if the last PSA value was >2 ng/mL; death from prostate cancer. All PSA measures collected after EBRT and until the end of follow-up (minimum time to clinical recurrence or lost to follow-up) were analysed. Three prognostic factors were considered: initial level of PSA at diagnosis (iPSA) as continuous in the log scale, T-stage category (Stage 1–2 vs. 3–4), and Gleason score (7, 8–10 vs. 2–6), an indicator grading prostate cancers. In brief, the sample included 459 subjects with a median follow-up of 5.16 (interquartile range (IQR)=2.68,7.69) years. During the follow-up, 74 patients (16.1%) had a clinical recurrence with a median of 2.77 (IQR=1.87,4.41) years after EBRT. The mean iPSA on the logarithm scale was 2.18 (SE=0.90), 41 (8.9%) patients had a T-stage of 3 or 4 and, respectively, 173 (37.7%) and 34 (7.4%) patients had a Gleason score of 7 and above 7.

5.2 Joint latent class model estimation

The trajectory of PSA on the logarithm scale ($\ln(\text{PSA} + 0.1)$) was described in a three-component parametric linear mixed model² with baseline (post-treatment level of PSA), short-term drop of PSA approximated by $f_1(t) = (1 + t)^{1.5} - 1$, and linear long-term trend, each of the three components being class- and subject-specific with class-specific correlated random-effects. The baseline hazard functions were class-specific Weibull functions starting at 1 year. As the aim was to propose a dynamic prognostic tool, we chose to include the three covariates in all parts of the JLCM, that is in the latent class membership probability in equation (1), in the interaction with the three components of the trajectory in equation (2) and in the survival model in equation (3) with common effects over classes in equations (2) and (3). For comparison, exactly the same model structure was adopted for the SREM with the same three trajectory components, a Weibull hazard function with a 1-year delay time in the survival model and the three covariates included in equation (6), and in equation (5) with interaction with each of the three components. In addition, either the current true PSA level alone, the current slope of PSA alone, or both the current PSA level and slope was considered in the survival model.³ JLCM models with from 1 to 5 latent classes and SREM were estimated and summarised in Table 1. We note that the 1-class JLCM represents the model assuming independence between PSA and time-to-recurrence and as such, is both a special case of the JLCM ($G = 1$) and of the SREM ($\eta = 0$).

The JLCM with the best BIC included three latent classes but the conditional independence assumption was rejected for this model so that the model with four latent classes for which the CI assumption was not rejected ($p = 0.0523$) was preferred. In the SREM, inclusion of the PSA functions as covariates improved the goodness-of-fit with a maximal difference in log-likelihood between the SREM with both current PSA level and slope as covariates and the model assuming independence of 81. In contrast, the difference in log-likelihood between the JLCM with only two latent classes and the model assuming independence was more than 260. This greater difference illustrates the flexibility of the JLCM to model the biomarker trajectory, the time-to-event and their dependence. However, this is accomplished in this example with a large increase of the number of parameters: 61 for the 4-class JLCM vs. 30 for the SREM, which might suggest some over-fitting. We note however that the 4-latent class JLCM with a reduced adjustment for covariates (i.e. covariates only in the survival model rather than in the survival model, the longitudinal model

Table 1. Summary of joint model (JLCM and SREM) fits: log-likelihood (\mathcal{L}), number of parameters (p), Bayesian information criterion (BIC), Score Test Statistic (ST) and ST p -value, and latent class proportion (in %). For JLCM, $G = g$ indicates g latent classes, for SREM, $Y(t)$ and $\delta Y(t)$, respectively, refer to the current PSA level and the current PSA slope included as covariates in the survival model.

Joint model	\mathcal{L}	p	BIC	ST (p -value)	Latent class proportion (%)
JLCM					
$G = 1$	-2711.55	28	5594.71	134.44 (<0.001)	100
$G = 2$	-2446.69	39	5132.42	33.22 (<0.001)	(88.67, 11.33)
$G = 3$	-2376.54	50	5059.54	14.06 (0.0028)	(85.62, 9.15, 5.23)
$G = 4$	-2347.24	61	5068.35	7.72 (0.0523)	(85.19, 8.93, 4.14, 1.74)
$G = 5$	-2330.03	72	5101.35	10.80 (0.0129)	(84.75, 9.59, 2.18, 1.74, 1.74)
SREM					
$Y(t)$	-2650.90	29	5479.55	–	–
$\delta Y(t)$	-2636.26	29	5450.26	–	–
$Y(t), \delta Y(t)$	-2630.60	30	5445.07	–	–

and the class-membership model) gave also a substantially improved fit: $\mathcal{L} = -2502.4$ (BIC = 5231.5) with only 37 parameters.

Class-specific predicted trajectories and survival functions, displayed in Figure 1(A) and (B) show a large latent class (class 1) representing 85.2% of the subjects with a very low long-term increase of PSA and a very small risk of recurrence over years. The three other latent classes 2 to 4 (representing respectively 8.9%, 4.1% and 1.7% of the subjects) correspond to different profiles of PSA trajectory associated with risks of recurrence from moderate to intense. Weighted subject-specific predicted trajectories and weighted subject-specific event-free probabilities displayed in Figure 1(C) and (D) demonstrate a very good fit of both the longitudinal and the time-to-event data.

The four latent classes of the JLCM provided very good discrimination with an entropy measure of 0.94 very close to 1 and proportions of maximal posterior probabilities above 0.8 of respectively 97.7%, 87.8%, 89.5% and 100% in classes 1 to 4. Finally, mean maximal posterior probabilities of subjects classified in each of the four latent classes were very close to 1 with, respectively, 0.98, 0.92, 0.96 and 0.96 for classes 1 to 4.

5.3 Evaluation of dynamic predictions

An example of the dynamic predictive tool derived from the 4-class compared to the one derived from the 1-class JLCM that does not account for PSA repeated measures (i.e. a standard survival model) is given in Figure 2 for a subject who recurred 3.8 years after the end of EBRT. While a low risk of recurrence is predicted by the standard survival model during the whole follow-up, the risk of recurrence predicted by the 4-class JLCM increases with the increasing PSA post-treatment trajectory.

Predictive accuracy can be evaluated on the same dataset as used for estimation using the cross-validated estimate of EPOCE. Graph 3 displays (A) this estimate, the $CVPOL_a$, computed for a time at prediction from 1 to 6 years after EBRT from the different JLCM and SREM, as well as (B) the differences in $CVPOL_a$ between the 4-class JLCM and, respectively, the best SREM and the 2- and 3-class JLCM with 95% tracking intervals. JLCM with at least three latent classes have a better predictive accuracy than the 1-class prognostic model not accounting for PSA repeated measures. This illustrates the relevance of updating the risk of recurrence using dynamic prognostic tools

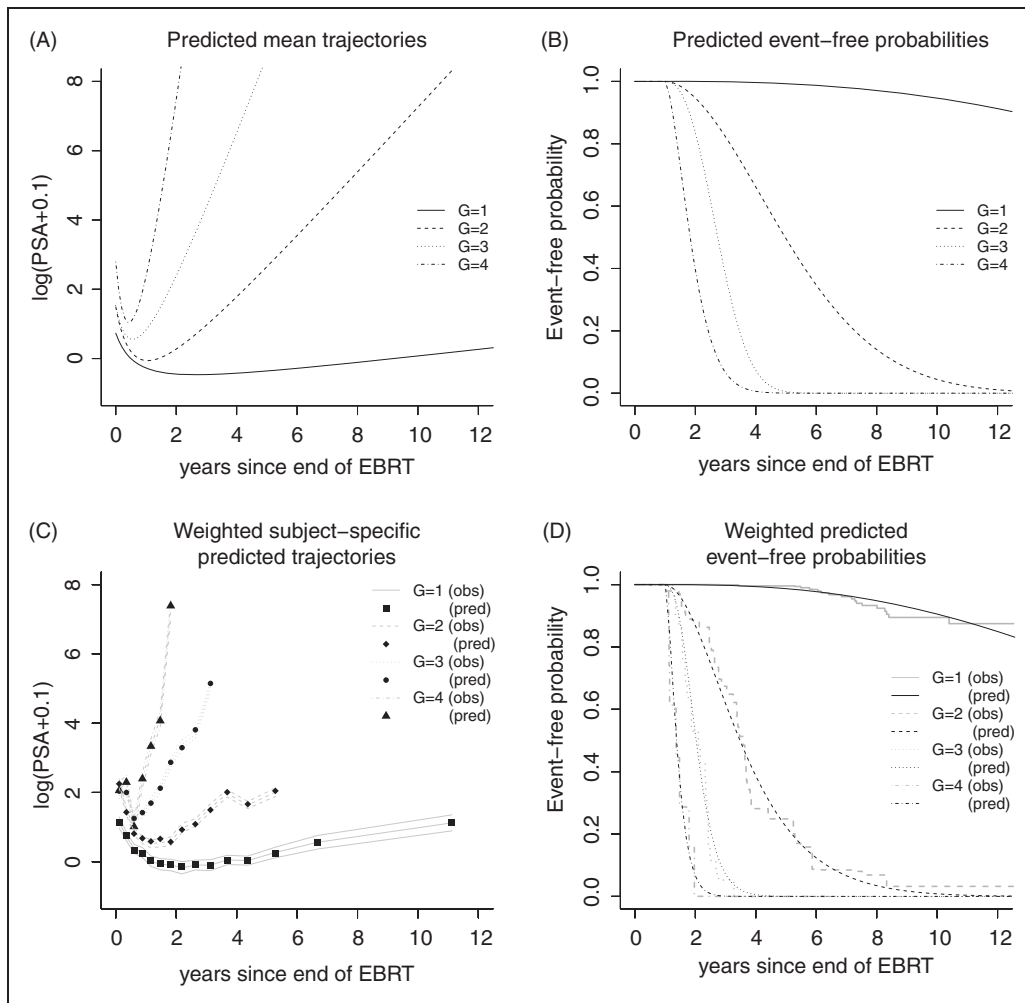


Figure 1. (A) Class-specific predicted mean trajectories and (B) class-specific event-free probabilities from the 4-class JLCM for a subject with Tstage <3, Gleason <7 and iPSA = 2 ng/mL. (C) Weighted subject-specific predicted trajectories (pred) and weighted observed trajectories (obs) from the 4-class JLCM. (D) Weighted predicted event-free probabilities (pred) and weighted Kaplan–Meier estimates (obs) from the 4-class JLCM.

rather than standard prognostic tools based only on baseline information. Moreover, according to the 95% tracking interval, the 4-class JLCM seems better than the simpler 2-class JLCM at least in the four first years after EBRT where detecting subjects at high risk of recurrence is particularly important. Similarly, the 4-class JLCM also appears better than the SREM (including the PSA current level and slope as covariate) in these four first years after EBRT according to the 95% tracking interval (Figure 3).

In prognostic model development, validation on external data is recommended. We used a second independent dataset (VANC) to compute predictive accuracy measures based on the joint models estimated on UM data. VANC data came from the British Columbia Cancer Agency (Vancouver, Canada).³⁸ Inclusion criteria were the same as in the UM sample. In brief, 719 patients including 126

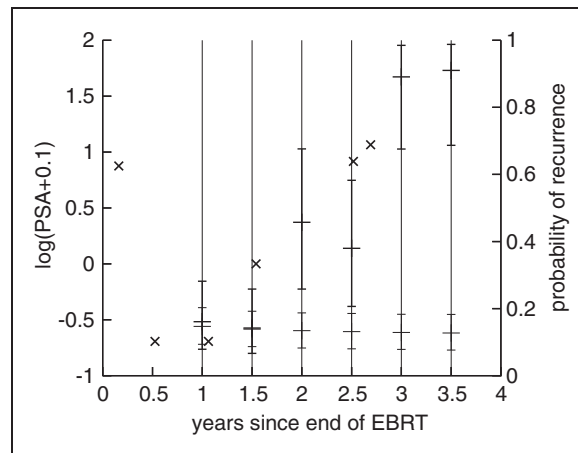


Figure 2. Example of individual predicted probability of clinical recurrence within 3 years updated every 6 months from the 4-class JLCM (bold plain line) and the 1-class JLCM (plain line) for a subject who recurred 3.8 years after the end of EBRT and had T-stage = 2 Gleason = 7, iPSA = 9.7 ng/mL and repeated PSA measurements denoted by \times .

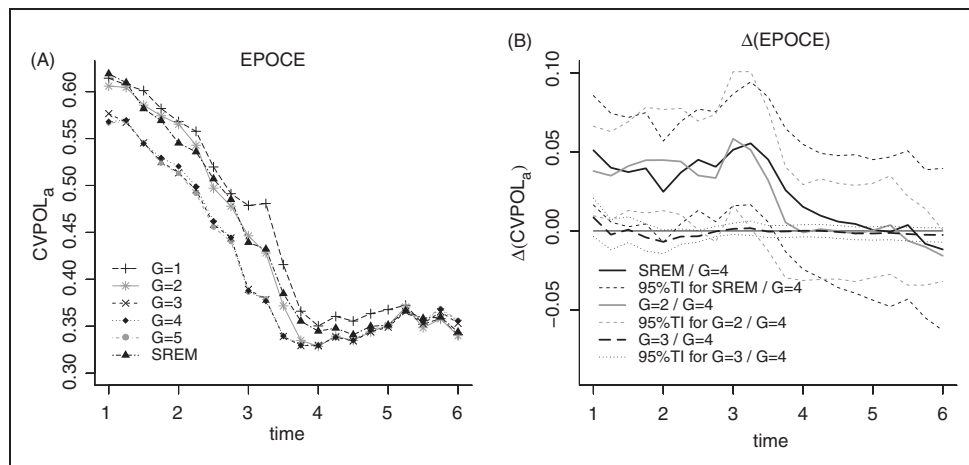


Figure 3. (A) $CVPOL_a$, cross-validated estimate of EPOCE, computed from the SREM (with current PSA level and current PSA slope included as covariates in the survival model), and from JLCM models with $G = 1$ to $G = 5$ latent classes for times at prediction from 1 to 6 years after EBRT, (B) Differences in EPOCE with 95% tracking interval (TI) on UM dataset ($N = 459$).

(17.5%) who had a clinical recurrence were followed-up for a median of 6.48 (IQR=4.28,8.22) years. The mean iPSA on the logarithm scale was 1.97 (SE=0.74), 105 (14.6%) patients had a T-stage of 3 or 4 and, respectively, 145 (20.2%) and 38 (5.3%) patients had a Gleason score of 7 and above 7.

Figure 4 displays the estimate of EPOCE, the MPOL, and its differences, as well as two estimates of the Integrated Brier Score (IBS), computed using the data-based approach defined in equation (12) or the model-based approach defined in equation (13). The two IBS estimates gave roughly the

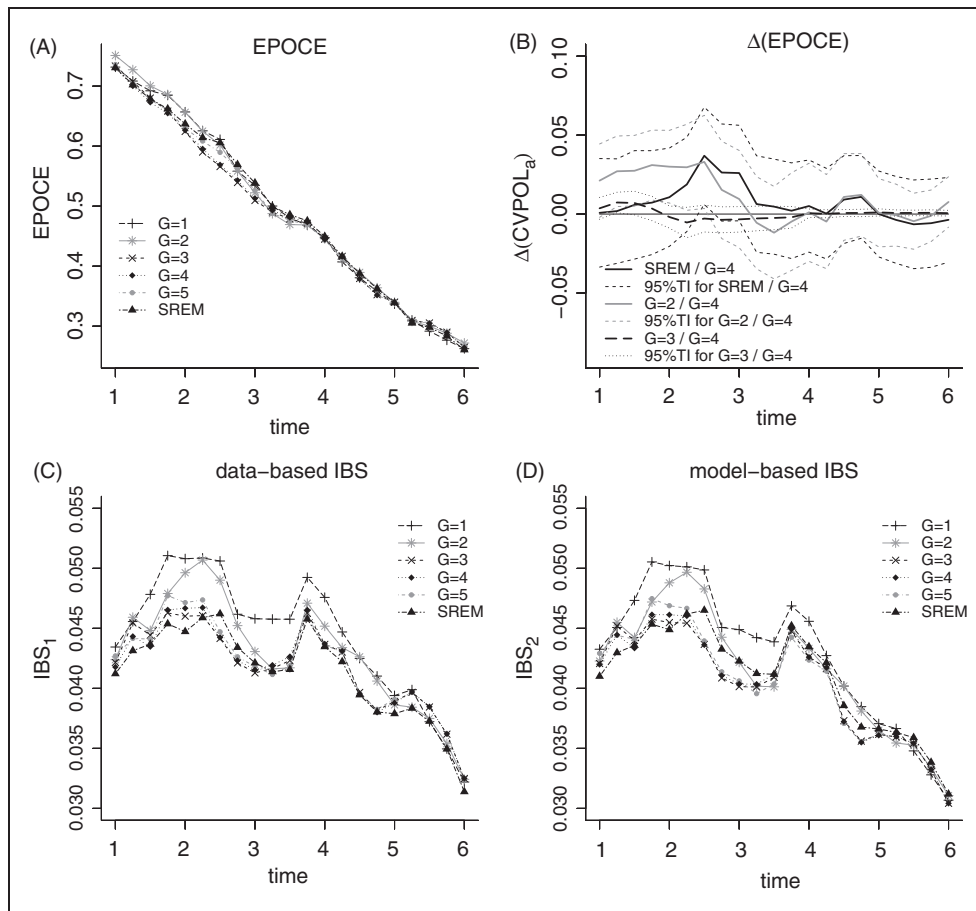


Figure 4. Predictive accuracy measures at times at prediction from 1 to 6 years after EBRT computed on VANC dataset ($N = 719$) from joint models estimated on UM dataset: (A) EPOCE estimate, (B) difference in EPOCE and 95% tracking interval (TI), (C) data-based estimate of Integrated Brier Score (IBS), and (D) model-based estimate of IBS. The censoring distribution for the IBS data-based estimate is modelled in a semi-parametric proportional hazard model with Gleason, T-stage, iPSA as covariates. SREM refers to the SREM with current PSA level and slope included as covariates in the survival model while $G = 1$ to $G = 5$ refer to JLCM with $G = 1$ to $G = 5$ latent classes.

same results, both indicating a reduced error of prediction during the entire follow-up for JLCM with at least three latent classes and the SREM. According to EPOCE, the SREM and the JLCM from two to five latent classes seemed to have relatively close predictive accuracies. But difference in EPOCE between the 4-class JLCM and either the 2-class JLCM or (to a lesser extent) the SREM remained substantial in the first years of prediction according to the 95% tracking intervals of EPOCE differences, since these intervals exclude zero.

6 Concluding remarks

Joint latent class models have received much less attention than shared-random-effect models in the joint modelling literature. Yet, although they may not be suitable to evaluate specific assumptions

regarding the link between a longitudinal marker and the time to an event, as in surrogate marker evaluation for example, they offer a flexible framework to model the joint distribution of the longitudinal marker and the time-to-event. The JLCM summarises their dependency by the assumed categorical structure explaining heterogeneous profiles of the marker and risks of the event, while the SREM summarises their dependency by specific a priori determined functions of the marker trajectory. This characteristic of the JLCM, which may restrict its usefulness to descriptive analyses, can turn out to be a great asset in prediction studies as it approximates any structure, even complex, of the correlated data without a priori assumption. As an example, in the illustration about prostate cancer progression after radiation therapy, the JLCM with four latent classes gave a substantial gain in goodness-of-fit compared to the SREM.

Evaluating predictive accuracy of prognostic tools derived from joint models is not straightforward because of the censoring process and of the dynamic nature of the prognostic tool. We reviewed two types of predictive accuracy measures, quadratic errors of prediction and expected prognostic observed cross-entropy, and showed how to compute them with a JLCM and with a SREM. The quadratic error of prediction (or Brier Score) is reasonably well known in the prediction literature, but its estimates rely on assumptions regarding the censoring distribution and the window of horizons in which predictive accuracy is evaluated. In contrast, the expected prognostic observed cross-entropy does not require any assumption regarding the censoring distribution, makes use of all the available information from the time at prediction, and provides a tracking interval for the difference of predictive accuracy that makes the comparison of models possible.

In this article we illustrated the methods using a dataset that was small relative to the complexity of the JLCM that included a large number of parameters. This was for illustrative purposes only. It emphasised the flexibility of this approach to include covariates in different ways, and model the dependency (stratification of the risks of events over classes for example). However, it produced large credibility bands for the individual predicted probabilities and large tracking intervals for the EPOCE differences. In practice, when developing dynamic predictive tools and applying such joint models, estimation on large datasets (possibly pooled datasets) is recommended to ensure precise estimates of the model parameters and less uncertainty in the individual predictions. Moreover, validation of the dynamic predictive tools on different external datasets (rather than a single one as in this example) with different characteristics would be recommended. Further validation of dynamic predictive tools from a joint model on external data, and comparison with other approaches such as survival models including the previous biomarker measures are described in Proust-Lima and Taylor.²

Finally, we note that all the models fitted in the illustration were estimated using available R packages: `lcmm` package for JLCM (`Jointlcmm` and `epoce` functions) and `JM` package for SREM. Other codes for the computation of predictive accuracy measures are available on request from the authors.

Funding

This work was supported by the French National Institute of Cancer INCa [grant PREDYC number 2010-059] and by the US National Cancer Institute [grant CA110518].

Acknowledgements

The authors thank Tom Pickles from the British Columbia Cancer Agency (Vancouver, British Columbia, Canada) for providing the VANC dataset.

References

1. Wulfsohn MS and Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**(1): 330–339.
2. Proust-Lima C and Taylor JMG. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of post-treatment PSA: a joint modelling approach. *Biostatistics* 2009; **10**: 535–549.
3. Yu M, Taylor JMG and Sandler HM. Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *J Am Stat Assoc* 2008; **103**: 178–187.
4. Prentice RL. Covariate measurement errors and parameter estimation in cox's failure time regression model. *Biometrika* 1982; **69**(2): 331–342.
5. Taylor JMG, Yu M and Sandler HM. Individualized predictions of disease progression following radiation therapy for prostate cancer. *J Clin Oncol* 2005; **23**(4): 816–825.
6. Henderson R, Diggle P and Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**(4): 465–480.
7. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; **67**(3): 819–829.
8. Huang X, Li G, Elashoff RM and Pan J. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Anal* 2011; **17**(1): 80–100.
9. Rizopoulos D and Ghosh P. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat Med* 2011; **30**(12): 1366–1380.
10. Vermunt JK and Magidson J. Latent class models for classification. *Comput Stat Data Anal* 2003; **41**(3-4): 531–537.
11. Lin H, Turnbull BW, McCulloch CE and Slate EH. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *J Am Stat Assoc* 2002; **97**: 53–65.
12. Proust-Lima C, Joly P and Jacqmin-Gadda H. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Comput Stat Data Anal* 2009; **53**(4): 1142–1154.
13. Garre FG, Zwiderman AH, Geskus RB and Sijpkens YWJ. A joint latent class changepoint model to improve the prediction of time to graft failure. *J Roy Stat Soc Ser A* 2008; **171**(1): 299–308.
14. Beunckens C, Molenberghs G, Verbeke G and Mallinckrodt C. A latent-class mixture model for incomplete longitudinal gaussian data. *Biometrics* 2008; **64**(1): 96–105.
15. Dantan E, Proust-Lima C, Letenneur L and Jacqmin-Gadda H. Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. *Int J Biostat* 2008; **4**. Article 14.
16. Zhang S, Müller P and Do KA. A Bayesian semi-parametric survival model with longitudinal markers. *Biometrics* 2010; **66**(2): 435–443.
17. Diggle PJ, Sousa I and Chetwynd AG. Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture. *Stat Med* 2008; **27**(16): 2981–2998.
18. Laird NM and Ware JH. Random effects models for longitudinal data. *Biometrics* 1982; **38**: 963–974.
19. Marquardt D. An algorithm for leastsquares estimation of nonlinear parameters. *SIAM J Appl Math* 1963; **11**: 431–441.
20. Stephens M. Dealing with label switching in mixture models. *J Roy Stat Soc Ser B* 2000; **62**(4): 795–809.
21. Redner RA and Walker FH. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev* 1984; **26**(2): 195–239.
22. Hipp JR and Bauer DJ. Local solutions in the estimation of growth mixture models. *Psychol Meth* 2006; **11**: 36–53.
23. Biernacki C, Celeux G and Govaert G. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Comput Stat Data Anal* 2003; **41**(3-4): 561–575.
24. Hawkins DS, Allen DM and Stromberg AJ. Determining the number of components in mixtures of linear models. *Comput Stat Data Anal* 2001; **38**(1): 15–48.
25. Han J, Slate EH and Pena EA. Parametric latent class joint model for a longitudinal biomarker and recurrent events. *Stat Med* 2007; **26**(29): 5285–5302.
26. Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. *J Stat Softw* 2010; **35**(9): 1–33.
27. Dobson A and Henderson R. Diagnostics for joint longitudinal and dropout time modeling. *Biometrics* 2003; **59**: 741–751.
28. Lin H, McCulloch CE and Rosenheck RA. Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* 2004; **60**(2): 295–305.
29. Jacqmin-Gadda H, Proust-Lima C, Taylor JMG and Commenges D. Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. *Biometrics* 2010; **66**(1): 11–19.
30. Muthén B, Brown CH, Masyn K, et al. General growth mixture modeling for randomized preventive interventions. *Biostatistics* 2002; **3**(4): 459–475.
31. Rizopoulos D, Verbeke G and Molenberghs G. Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* 2010; **66**(1): 20–29.
32. Henderson R, Diggle P and Dobson A. Identification and efficacy of longitudinal markers for survival. *Biostatistics* 2002; **3**(1): 33–50.
33. Schoop R, Schumacher M and Graf E. Measures of prediction error for survival data with longitudinal covariates. *Biometrical J* 2011; **53**(2): 275–293.
34. Commenges D, Liqueur B and Proust-Lima C. Choice of prognostic estimators in joint models by estimating differences of expected conditional Kullback-Leibler risks. *Biometrics* 2012; (in press).
35. Gerds TA and Schumacher M. Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical J* 2006; **48**(6): 1029–1040.
36. Rosthøj S and Keiding N. Explained variation and predictive accuracy in general parametric statistical models: the role of model misspecification. *Life-time Data Anal* 2004; **10**(4): 461–472.
37. Gerds TA and Schumacher M. Efron-type measures of prediction error for survival analysis. *Biometrics* 2007; **63**(4): 1283–1287.
38. Pickles T, Kim-Sing C, Morris W, et al. Evaluation of the Houston biochemical relapse definition in men treated with prolonged neoadjuvant and adjuvant androgen ablation and assessment of followup lead-time bias. *Int J Radiat Oncol Biol Phys* 2003; **57**: 1118.

Liste des publications et communications scientifiques

Publications

- Sène M., Taylor J.M.G., Dignam J.J., Jacqmin-Gadda H., Proust-Lima C. Personalized dynamic prediction of prostate cancer recurrence according to the initiation of a second treatment : development and validation. Soumis.
- Sène M., Bellera C.A., Proust-Lima C. Shared random-effect models for the joint analysis of longitudinal and time-to-event data : application to the prediction of prostate cancer recurrence. *Journal de la Société Française de Statistique*. Accepté.
- Proust-Lima C., Sène M., Taylor J.M.G., Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data : A review. *Statistical Methods in Medical Research*. Publié en ligne le 19 Avril 2012. doi : 10.1177/0962280212445839.

Communications orales

- Sène M., Proust-Lima C. Development and validation of dynamic tools to predict the recurrence of prostate cancer after a first treatment and its use to evaluate the benefit of immediate initiation of an hormonal therapy. 9^e Journées du canceropôle GSO, Limoges - France, 16-18 Octobre 2013.
- Sène M., Bellera C.A., Proust-Lima C. Joint modeling of longitudinal and time-to-event data with application to the prediction of prostate cancer recurrence. 28th International Workshop on Statistical Modelling, Palermo - Italy, 8-12 July 2013.
- Sène M., Jacqmin-Gadda H., Bellera C.A., Proust-Lima C. Joint modeling of longitudinal and time-to-event data with application to the prediction of prostate cancer

Liste des publications et communications scientifiques

- recurrence. 4th *IBS Channel Network Conference*, St Andrews - Scotland, 3-5th July 2013.
- Sène M., Bellera C.A., Proust-Lima C. Joint modeling of longitudinal and time-to-event data with application to the prediction of prostate cancer recurrence. 45^e *Journées de Statistique de la de la SFDS*, Toulouse - France, 27-31 Mai 2013.
 - Proust-Lima C., Sène M. Dynamic predictions from joint models for longitudinal biomarker trajectory and time to clinical event : development and validation. 33rd *Annual Conference of the International Society for Clinical Biostatistics*, Bergen - Norway, 19-23rd August 2012.
 - Sène M., Proust-Lima C. Joint modelling of longitudinal and time-to-event data : Dynamic predictions. 44^e *Journées de Statistique de la SFDS*, Bruxelles - Belgique, 21-25 Mai 2012.
 - Sène M., Proust-Lima C. Prédications dynamiques à partir de la modélisation conjointe de données longitudinales et de temps d'événement. *EPICLIN 6 / 19^{èmes} Journées de Statisticiens de CLCC*, Lyon - France, 9-11 Mai 2012.

Séminaires invités

- Sène M., Proust-Lima C. Modélisation conjointe de données longitudinales et de temps d'événements : Illustration avec des données de cancer de la prostate. *Le Laboratoire de Mathématiques de Besançon*, Besançon - France, 11 Mars 2013.
- Sène M., Proust-Lima C. Modélisation conjointe de données longitudinales et de temps d'événements : Illustration avec des données de cancer de la prostate. *Institut de cancérologie Gustave Roussy*, Villejuif - France, 14 Janvier 2013.