



# Modèles conjoints pour l'analyse longitudinale de la qualité de vie relative à la santé en présence de sorties d'étude informatives en oncologie

Benjamin Cuer

## ► To cite this version:

Benjamin Cuer. Modèles conjoints pour l'analyse longitudinale de la qualité de vie relative à la santé en présence de sorties d'étude informatives en oncologie. Biologie cellulaire. Université Montpellier, 2020. Français. NNT : 2020MONT080 . tel-03341261

**HAL Id: tel-03341261**

**<https://theses.hal.science/tel-03341261>**

Submitted on 10 Sep 2021

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

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

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biostatistique

École doctorale I2S – Information, Structures, Systèmes

Unité de recherche de l'Institut régional du Cancer Montpellier

## Modèles conjoints pour l'analyse longitudinale de la qualité de vie relative à la santé en présence de sorties d'étude informatives en oncologie

Présentée par Benjamin Cuer

Le 10 décembre 2020

Sous la direction de Caroline Mollevi  
et co-encadrée par Célia Touraine

Devant le jury composé de

Dr Caroline Mollevi, PhD

Dr Célia Touraine, PhD

Dr Carine Bellera, PhD

Pr Xavier Paoletti, PU-PH

Pr Nicolas Molinari, PU-PH

Pr Mathieu Ribatet, PU

Université de Montpellier – Institut du Cancer Montpellier

Institut du Cancer Montpellier

Université de Bordeaux – Institut Bergonié

Université Paris Saclay / Versailles St Quentin – Institut Curie

Université de Montpellier – CHU Montpellier

École Centrale de Nantes – Université de Nantes

Directrice de thèse

Co-encadrante de thèse

Rapporteure

Rapporteur

Président

Examineur



UNIVERSITÉ  
DE MONTPELLIER

Cette thèse a été préparée au sein des équipes de « Recherche intégrée pour une médecine personnalisée en oncologie digestive » (IRCM) et de Biométrie (ICM). Elle a été financée par une bourse de la Région Occitanie.





# Remerciements

---

Mes tous premiers remerciements s'adressent à ma directrice de thèse Caroline Mollevi et à Célia Touraine qui a co-encadré ce travail. Ces quelques lignes sont bien insuffisantes au regard de tout ce que vous m'avez apporté pendant ces trois années. Un travail de thèse est souvent long et difficile, mais le duo complémentaire que vous avez formé m'a permis d'évoluer dans un cadre de travail plus que confortable. Caroline, je te remercie grandement pour ton pragmatisme et ton suivi constant. Célia, je te remercie vivement pour ton investissement infailible, mais aussi pour ta rigueur et ton accompagnement. Vos mots sont souvent similaires, et même pendant les moments difficiles, vous avez toujours su trouver les bons pour continuer d'avancer. Merci pour votre confiance, votre soutien, et aussi pour vos très nombreuses relectures sans lesquelles ce travail n'aurait jamais pu être réalisé.

Je suis très heureux d'avoir la possibilité d'exprimer ici ma gratitude envers l'ensemble des membres du jury pour leur participation à mon jury de thèse et l'évaluation de ce travail.

Je remercie Nicolas Molinari d'avoir accepté de présider ce jury. Je n'oublie pas vos enseignements des biostatistiques appliqués à la santé, qui comptent parmi les premiers que j'ai reçus. Votre présence est un honneur, soyez-en remercié.

Je tiens à remercier très chaleureusement Carine Belera et Xavier Paoletti pour l'intérêt que vous avez porté à mes travaux. Je suis très honoré que vous ayez accepté de juger mon travail, en tant que rapporteurs de cette thèse. Vos larges connaissances sur l'utilisation de méthodes statistiques en oncologie m'apporteront certainement beaucoup. Veuillez trouver ici l'expression de ma reconnaissance.

---

Mes remerciements s'adressent également à Mathieu Ribatet pour avoir accepté sans hésitation et avec un enthousiasme évident de participer à ce jury. Ses conseils et son soutien inconditionnel tout au long de mes années de formation m'ont énormément appris. Qu'il en soit ici chaleureusement remercié.

Aux membres de l'équipe de biométrie que j'ai croisés, qui sont encore là, ou partis vers d'autres aventures, merci de m'avoir accueilli parmi vous. Je voudrais également remercier Sophie Gourgou pour m'avoir permis de réaliser mes travaux de recherche dans son équipe et dans de si bonnes conditions. Mention spéciale au petit groupe d'habitues ponctuels (plus ou moins prêts d'ailleurs) avec qui j'ai pu déjeuner à 12h00. Ce moment précieux et convivial m'était indispensable et nous aura permis de tisser des liens solides au fil des années. Je remercie également mes trois colocataires de bureau : vous qui m'avez si chaleureusement accueilli et où je pense m'être si rapidement intégré parmi vous. Aujourd'hui, il m'est difficile de le quitter, bien que cette rupture ait déjà été initiée en raison d'un certain dénommé « confinement ». Je remercie aussi les différents stagiaires. Votre présence m'a apporté beaucoup, mais aussi réconforté et aidé dans l'avancée de mon travail. Je vous souhaite bonne chance dans vos parcours respectifs.

Enfin, ce travail n'aurait pas pu voir le jour sans la présence des gens qui m'entourent et qui me sont chers.

A mes amis qui ont partagé de près ou de loin, mais qui ont toujours su être présents et vous me l'avez témoigné par maintes reprises. Votre présence à mes côtés est une chance et je vous en remercie.

A la venue de tous nos êtres chers qui ont pointé le bout de leur nez et qui nous rappelle qu'il n'y a vraiment pas que le travail dans la vie.

A mes collègues et amis rencontrés à l'université. Certains ont pris la voie de l'enseignement et/ou de la recherche, d'autres non. Malgré nos différents parcours, je suis fière d'avoir grandi avec vous et vous souhaite le meilleur pour la suite. Bonne chance également pour les futurs docteurs !

Aux (anciens) sportifs que j'ai rencontrés sans qui mes semaines (et parfois même mes week-ends) auraient été moins drôles ! Merci de m'avoir accompagné depuis de si nombreuses années. Une chose est sûre, en ces lieux, on y fait parfois

---

des rencontres extraordinaires dont de liens solides et affectueux aujourd'hui en sont la démonstration.

Y a-t-il d'autres personnes à remercier ? Assurément que oui. Ne m'en voulez pas, mais il m'est impossible de remercier individuellement toutes les personnes qui ont su m'accompagner, m'écouter, m'aider. Naturellement, mes derniers remerciements s'adressent à ma famille et en particulier à mes parents que j'ai impliqués directement dans cette thèse. Merci à tous les deux pour votre soutien sans lequel le chemin aurait été beaucoup plus difficile. Je les remercie d'avoir toujours été présents à mes côtés, et de m'avoir permis d'évoluer dans les meilleures conditions possibles.





# Valorisations scientifiques

---

## Publications

### Articles

Cuer B, Mollevi C, Anota A, Charton E, Juzyna B, Conroy T, Touraine C. Handling informative dropout in longitudinal analysis of health-related quality of life: application of three approaches to data from the esophageal cancer clinical trial PRODIGE 5/ACCORD 17. *BMC Medical Research Methodology*. 2020 Sep 3;20(1):223.

Charton E, Cuer B, Cottone F, Efficace F, Touraine C, Hamidou Z, Fiteni F, Bonnetain F, Woronoff-Lemsi M.-C, Bascoul Mollevi C, Anota A. Time to deterioration in cancer randomized clinical trials for patient-reported outcomes data: a systematic review. *Quality of Life Research*. 2020 Apr 1;29(4):867–78.

Cuer B, Conroy T, Juzyna B, Gourgou S, Mollevi C, Touraine C. Joint modelling with competing risks of dropout for longitudinal analysis of health-related quality of life in cancer clinical trials. (en soumission)

### Chapitre de livre

Cuer B, Touraine C, Mollevi C. Chapitre : Données manquantes. Mathoulin-Pelissier S, Gourgou S, Paoletti X. *Méthodes biostatistiques appliquées à la recherche clinique en cancérologie (2e édition)*. Arcueil : John Libbey Eurotext, 2021 (à paraître).

---

## Communications orales en congrès

### Internationales

#### **International Society for Quality of Life Research (ISOQOL), 26 octobre 2018, Dublin**

Cuer B, Mollevi C, Anota A, Charton E, Bonnetain F, Conroy T, Juzyna B, Gourgou S, Touraine C. Handling informative drop-out in longitudinal analysis of health-related quality of life: Application on data from the oesophageal cancer clinical trial PRODIGE 5/ACCORD 17.

#### **International Society for Quality of Life Research (ISOQOL), 26 octobre 2018, Dublin**

Charton E, Efficace F, Cottone F, Touraine C, Cuer B, Hamidou Z, Fiteni F, Bonnetain F, Mollevi C, Anota A. Time to health-related quality of life score deterioration in oncology phase III clinical trials: a systematic review.

### Nationales

#### **Conférence Francophone d'Epidémiologie Clinique (EPICLIN) - 25èmes Journées des statisticiens des Centre de Lutte Contre le Cancer (CLCC), 30 mai 2018, Nice**

Cuer B, Anota A, Conroy T, Juzyna B, Gourgou S, Mollevi C, Touraine. Prise en compte des données manquantes potentiellement informatives dans l'analyse longitudinale de la qualité de vie relative à la santé sur les données de l'essai clinique PRODIGE 5/ACCORD 17.

#### **Conférence Francophone d'Epidémiologie Clinique (EPICLIN) - 26èmes Journées des statisticiens des Centre de Lutte Contre le Cancer (CLCC), 15 mai 2019, Toulouse**

Touraine C, Cuer B, Conroy T, Juzyna B, Gourgou S, Mollevi C. Analyse conjointe de la survie et de la qualité de vie relative à la santé chez des patients atteints de cancer du pancréas métastatique : Application aux données de l'essai clinique PRODIGE 4/ACCORD 11.

---

**Groupe de recherche (GDR) « Statistiques et Santé », 11 octobre 2019, Paris**

Cuer B, Mollevi C, Touraine C. Modèles conjoints avec risques concurrents : analyse de données longitudinales de qualité de vie en oncologie en présence de différents types de sorties d'étude.

**Conférence Francophone d'Epidémiologie Clinique (EPICLIN) - 27èmes Journées des statisticiens des Centre de Lutte Contre le Cancer (CLCC), 16 septembre 2020, Angers**

Cuer B, Mollevi C, Conroy T, Juzyna B, Gourgou S, Touraine. Modèle conjoint à risques concurrents pour l'analyse longitudinale de la qualité de vie en oncologie en présence de différents types de sortie d'étude PRODIGE 5/ACCORD 17.

## **Communications affichées en congrès**

### **Internationales**

**International Society for Clinical Biostatistics (ISCB), 14 juillet 2019, Louvain**

Cuer B, Mollevi C, Conroy T, Juzyna B, Gourgou S, Touraine C. Joint modelling with competing risks of dropout for HRQoL longitudinal analysis in cancer clinical trials.

**International Society for Clinical Biostatistics (ISCB), 14 juillet 2019, Louvain**

Touraine C, Cuer B, Conroy T, Juzyna B, Gourgou S, Mollevi C. Joint analysis of health-related quality of life and overall survival in cancer clinical trials.

### **Nationales**

**Conférence Francophone d'Epidémiologie Clinique (EPICLIN) - 26èmes Journées des statisticiens des Centre de Lutte Contre le Cancer (CLCC), 15 mai 2019, Toulouse**

---

Cuer B, Mollevi C, Conroy T, Juzyna B, Gourgou S, Touraine C. Modélisation conjointe de la trajectoire de qualité de vie relative à la santé et du temps jusqu'à « dropout » pour la prise en compte des données manquantes monotones potentiellement informatives : application aux données de l'essai clinique PRODIGE 5/ACCORD 17.

**Conférence Francophone d'Epidémiologie Clinique (EPICLIN) - 27èmes Journées des statisticiens des Centre de Lutte Contre le Cancer (CLCC), 15 septembre 2020, Angers**

Touraine C, Cuer B, Conroy T, Juzyna B, Gourgou S, Mollevi C. Modèle conjoint versus modèle linéaire mixte pour l'analyse de données de qualité de vie dont le recueil est interrompu par la survenue d'un événement clinique en oncologie.

## **Autres productions**

### **Communications orales**

**Journée de l'école doctorale I2S (DOCTISS 2018), 14 juin 2018, Montpellier**

Prise en compte des données manquantes potentiellement informatives dans l'analyse longitudinale de la qualité de vie relative à la santé sur les données de l'essai clinique PRODIGE 5/ACCORD 17.

**Invitation au séminaire de l'Institut Gustave Roussy (IGR), 18 décembre 2018, Villejuif**

Prise en compte des données manquantes potentiellement informatives dans l'analyse longitudinale de la qualité de vie relative à la santé.

## **Prix scientifique**

**2019** Prix du meilleur poster à la Conférence Francophone d'Epidémiologie Clinique (EPICLIN) - 26èmes Journées des statisticiens des Centre de Lutte Contre le Cancer (CLCC)

# Table des matières

---

<b>Remerciements</b>	<b>v</b>
<b>Valorisations scientifiques</b>	<b>ix</b>
<b>Liste des abbréviations</b>	<b>xvii</b>
<b>Liste des figures</b>	<b>xx</b>
<b>Liste des tableaux</b>	<b>xxii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Qualité de vie relative à la santé . . . . .	3
1.1.1 Définition . . . . .	3
1.1.2 Instruments de mesure . . . . .	4
1.2 Analyse longitudinale de la QdV dans les essais cliniques . . . . .	8
1.2.1 Modèles linéaires mixtes . . . . .	8
1.2.2 Problématique des données manquantes . . . . .	12
1.3 Essai clinique PRODIGE 5/ ACCORD 17 . . . . .	12
1.4 Objectifs et plan . . . . .	14

## Table des matières

---

<b>2</b>	<b>Données manquantes</b>	<b>17</b>
2.1	Introduction . . . . .	17
2.2	Chapitre de livre : <i>Données manquantes</i> . . . . .	18
2.3	Discussion et conclusion . . . . .	26
<b>3</b>	<b>Étude de trois approches pour tenir compte de sorties d'étude informa- tives</b>	<b>27</b>
3.1	Introduction . . . . .	27
3.2	Article : <i>Handling informative dropout in longitudinal analysis of health- related quality of life</i> . . . . .	29
3.3	Modèles conjoints . . . . .	43
3.3.1	Structure générale de la modélisation conjointe . . . . .	43
3.3.2	Modèle conjoint à effets partagés . . . . .	44
3.3.3	Interprétation du modèle . . . . .	45
3.4	Discussion et conclusion . . . . .	46
<b>4</b>	<b>Modèles conjoints différenciant plusieurs causes de sorties d'étude</b>	<b>51</b>
4.1	Introduction . . . . .	51
4.2	Article : <i>Joint modelling with competing risks of dropout for longitudinal analysis of health-related quality of life in cancer clinical trials</i> . . . . .	53
4.3	Discussion et conclusion . . . . .	80
<b>5</b>	<b>Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude</b>	<b>85</b>
5.1	Prise en compte des trajectoires non linéaires . . . . .	85

5.1.1	Introduction . . . . .	85
5.1.2	Modèles conjoints avec des splines . . . . .	87
5.1.3	Application sur les données de l'essai clinique PRODIGE 5 / ACCORD 17 . . . . .	89
5.1.4	Conclusion . . . . .	97
5.2	Prise en compte du score en tant que donnée ordinale . . . . .	99
5.2.1	Introduction . . . . .	99
5.2.2	Modèle conjoint avec sous-GLMM à ratios de continuation . .	102
5.2.3	Application sur les données de l'essai clinique PRODIGE 5 / AC- CORD 17 . . . . .	107
5.2.4	Conclusion . . . . .	109
5.3	Discussion . . . . .	111
<b>6</b>	<b>Conclusion et perspectives</b>	<b>113</b>
6.1	Contributions par rapport aux objectifs initiaux . . . . .	114
6.2	Perspectives du travail de thèse . . . . .	117
6.3	Conclusion générale . . . . .	119
<b>A</b>	<b>Questionnaire EORTC QLQ-C30</b>	<b>123</b>
<b>B</b>	<b>Distributions des scores de QdV</b>	<b>127</b>
<b>C</b>	<b>Code pour le modèle de Diggle et Kenward</b>	<b>129</b>
<b>D</b>	<b>Etude des résidus du sous-modèle pour le temps jusqu'à sortie d'étude</b>	<b>137</b>



## **Table des matières**

---

D.1	Résidus de Cox-Snell . . . . .	138
D.2	Résidus martingales . . . . .	139
E	Résultats de simulations pour le modèle conjoint à ratio de continuation	141
F	Modèle cumulatif pour données ordinales	143
	Références bibliographiques	147

# Liste des abbréviations

---

<b>ACCORD</b>	Actions Concertées dans les Cancers Colorectaux et Digestifs
<b>AIC</b>	Akaike Information Criterion (critère d'information d'Akaike)
<b>EORTC</b>	European Organisation for Research and Treatment of Cancer
<b>FA</b>	Fatigue (fatigue)
<b>FDA</b>	Food and Drug Administration
<b>GLMM</b>	Generalized Linear Mixed Model (modèle linéaire mixte généralisé)
<b>HR</b>	Hazard Ratio
<b>HRQoL</b>	Health-Related Quality of Life (qualité de vie relative à la santé)
<b>INCa</b>	Institut National du Cancer
<b>LMM</b>	Linear Mixed Model (modèle linéaire mixte)
<b>MAR</b>	Missing At Random (données manquantes aléatoires)
<b>MCAR</b>	Missing Completely At Random (données manquantes complètement aléatoires)
<b>MCMC</b>	Méthode de Monte-Carlo par chaînes de Markov
<b>MNAR</b>	Missing Not At Random (données manquantes non aléatoires)
<b>OR</b>	Odds Ratio (rapport des cotes)
<b>OS</b>	Overall Survival (survie globale)
<b>PA</b>	Pain (douleur)
<b>PF</b>	Physical Functioning (fonction physique)
<b>PMM</b>	Pattern-Mixture Model (modèle de mélange de profils)

## Table des matières

---

<b>PRO</b>	Patients-Reported Outcome (donnée reportée par les patients)
<b>PRODIGE</b>	Partenariat de Recherche en Oncologie Digestive
<b>QdV</b>	Qualité de vie
<b>QL</b>	Global health status (santé globale des patients)
<b>QLQ-C30</b>	Quality of Life Questionnaire Core 30
<b>REML</b>	Restricted Maximum Likelihood (maximum de vraisemblance restreint)
<b>SISAQOL</b>	Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data
<b>SM</b>	Selection Model (modèle de sélection)
<b>SPM</b>	Shared-Parameter Model (modèle à effets partagés)

# Liste des figures

---

3.1	Représentation graphique d'un modèle conjoint constitué de deux processus (pour décrire les données répétées et la survenue d'un évènement) partageant une structure d'association commune décrite au travers des effets aléatoires. . . . .	44
5.1	Trajectoires individuelles (en gris), moyennes observées des scores et trajectoires prédites par le modèle conjoint linéaire dans le bras contrôle (symbole triangulaire vers le haut et courbe pleine) et le bras expérimental (symbole triangulaire vers le bas et courbe en pointillé). . . . .	87
5.2	Trajectoires prédites des scores de QdV modélisés par deux modèles conjoints, JM linéaire (en rouge) et JM splines (en bleu). . . . .	92
5.3	Résidus marginaux obtenus pour M = 20 imputations avec JM linéaire pour les 4 dimensions analysées. . . . .	96
5.4	Résidus marginaux obtenus pour M = 20 imputations avec JM splines pour les 4 dimensions analysées. . . . .	96
5.5	Q-Q plot des résidus standardisés du modèle conjoint avec splines pour vérifier l'adéquation à l'hypothèse de normalité. . . . .	97
5.6	Histogrammes des résidus standardisés du modèle conjoint avec splines pour vérifier l'hypothèse de normalité des résidus. . . . .	97

## Liste des figures

---

5.7	Estimation des <i>logits</i> de ratios de continuation pour l'analyse des scores de la dimension PA. . . . .	109
B.1	Histogrammes des scores de QdV à l'inclusion pour les données de l'essai clinique PRODIGE 5/ACCORD 17. . . . .	127
D.1	Résidus de Cox-Snell du modèle conjoint linéaire. En noir est tracée la fonction de survie de Kaplan-Meier des résidus avec l'intervalle de confiance à 95% et en gris la fonction de survie théorique distribuée selon une loi exponentielle de paramètre 1. . . . .	138
D.2	Résidus de Cox-Snell du modèle conjoint à splines. En noir est tracée la fonction de survie de Kaplan-Meier des résidus avec l'intervalle de confiance à 95% et en gris la fonction de survie théorique distribuée selon une loi exponentielle de paramètre 1. . . . .	138
D.3	Résidus martingales du modèle conjoint linéaire avec la courbe de lissage en rouge estimée par une régression non-paramétrique locale. . . . .	139
D.4	Résidus martingales du modèle conjoint à splines avec la courbe de lissage en rouge estimée par une régression non-paramétrique locale. . . . .	139
F.1	Représentation de la variable latente à partir du modèle cumulatif pour le score PA avec 7 catégories de réponse où les traits horizontaux correspondent aux paramètres de seuils $\lambda_k$ ( $k = 1, \dots, 6$ ) qui partitionnent l'échelle continue (à gauche) en 7 catégories ordonnées (à droite). . . . .	146

# Liste des tableaux

---

1.1	Caractéristiques du questionnaire EORTC QLQ-C30 (version 3.0). . .	5
1.2	Caractéristiques du questionnaire EORTC QLQ-OES18. . . . .	6
4.1	Attribution des causes de sorties d'étude (cause 1 : non liée à un évènement clinique, cause 2 : liée à un évènement clinique) en fonction de l'apparition de l'évènement clinique pendant l'évaluation de la QdV (O : observé, X : manquant). . . . .	80
5.1	Effet du bras de traitement sur l'évolution du score de QdV dans un modèle conjoint (JM linéaire) ou avec des splines (JM splines) donné par le test du rapport de vraisemblance (LR). . . . .	93
5.2	Estimations des paramètres pour les deux modèles conjoints avec des trajectoires de score de QdV linéaire (JM linéaire) ou non linéaire (JM splines). . . . .	94
5.3	Exemple de la réorganisation des observations pour transformer la variable ordinaire $Y$ en une nouvelle variable binaire $Y^*$ lorsque $K = 4$ .	106
5.4	Estimations des paramètres du modèle conjoint avec des trajectoires de score de QdV linéaire (JM linéaire, déjà présenté) et du modèle conjoint pour données ordinales avec sous-modèle à ratios de continuation (JM ordinal). . . . .	108

## Liste des tableaux

---

E.1	Résultats de simulations pour l'estimation du modèle conjoint à ratio de continuation en donnant la valeur moyenne estimée et l'écart-type des estimations pour chaque paramètre simulé. . . . .	142
F.1	Estimation du modèle cumulatif pour la modélisation des données de QdV de la dimension PA. . . . .	145







# Chapitre 1

## Introduction

---

La cancérologie est la spécialité médicale qui s'intéresse aux cancers, en particulier à leur diagnostic et à leur traitement. Le cancer est une maladie grave qui peut mettre en danger la survie de l'individu. D'après l'Institut National du Cancer (INCa), les cancers représentent en France métropolitaine la première cause de décès chez l'homme et la deuxième chez la femme ([INCa, 2011](#)), avec 382 000 nouveaux cas de cancer pour l'année 2018. Face à cet enjeu de santé publique, de nombreuses disciplines scientifiques se mobilisent pour améliorer les connaissances scientifiques sur cette maladie. Le cancer est généralement traité par chirurgie, radiothérapie, chimiothérapie ou hormonothérapie. Cependant, ces différents traitements, en luttant contre la maladie, peuvent provoquer des effets secondaires indésirables chez les patients, par exemple la chute de cheveux, des nausées et vomissements associés à la prise d'une chimiothérapie. Les traitements visent certes à prolonger la survie des patients malades, mais leur durée de vie (quantité de vie) est parfois augmentée au détriment de leur qualité de vie.

Les essais cliniques, ou essais thérapeutiques, sont des études scientifiques obligatoires et systématiques mises en place dans le cadre de l'évaluation de nouveaux traitements ou de nouvelles modalités de soin. Par exemple, les essais cliniques sont utilisés pour tester un nouveau médicament, une nouvelle méthode de diagnostic ou un nouveau protocole de traitement. Pour garantir la sécurité des malades et la rigueur scientifique, les essais cliniques se déroulent en plusieurs étapes

## Chapitre 1. Introduction

---

(phases) d'expérimentation avec d'abord une phase préclinique chez l'animal, suivie de 3 phases de développement.

Les essais de phase 1 constituent la première administration chez l'Homme. L'objectif principal est d'évaluer la tolérance du nouveau traitement en étudiant l'ensemble des effets indésirables attribué à son administration, sur un faible nombre de patients malades (en cancérologie entre 10 et 40 patients, à un stade avancé ou en situation d'échec thérapeutique). À l'issue de cette première phase d'expérimentation, il est ainsi possible de déterminer la dose maximale tolérée et la dose recommandée pour un futur essai de phase 2. Les essais de phase 2 ont pour objectifs d'explorer l'efficacité thérapeutique (objectif principal) et d'affiner le profil de tolérance (objectif secondaire) du nouveau traitement sur un nombre de patients supérieur à celui de la phase 1 (en général en incluant moins de 100 patients). Enfin, les essais de phase 3 permettent de montrer l'efficacité d'un nouveau traitement (traitement expérimental) en la comparant à la stratégie de référence (traitement standard). Cette dernière phase de développement nécessite l'inclusion d'un grand nombre de patients malades (plusieurs centaines ou milliers) pour tester le nouveau traitement en situation réelle : le traitement expérimental et le traitement standard sont alors comparés en matière de bénéfice clinique pour le patient. Pour supprimer les risques de biais liés à la comparaison des bras de traitement, l'attribution d'un patient dans un bras de traitement est faite au hasard, selon une méthode de randomisation et l'on parle alors d'essais cliniques « randomisés », qui sont reconnus et utilisés pour avoir un haut niveau de preuve scientifique. En général, le critère de jugement principal de référence est la survie globale (OS pour Overall Survival) qui est définie à partir du délai depuis l'inclusion jusqu'au décès des patients, quelle qu'en soient les causes. Cette définition, clairement définie et universelle, facilite la comparaison entre les études.

Les progrès de la médecine et les innovations ont permis de rendre les traitements de plus en plus efficaces en augmentant la survie des patients malades. Aujourd'hui, en cancérologie, il est de plus en plus difficile de trouver de nouveaux traitements qui montrent un gain significatif sur l'OS. Pour certaines localisations comme le cancer du sein, il serait compliqué de mener une étude avec un nombre de patients et/ou une durée d'étude raisonnable pour répondre à la question de l'efficacité d'un nouveau traitement. Pour pallier ces difficultés, des critères

de substitution intermédiaires centrés sur la tumeur et évaluables plus précocement ont été proposés, comme la survie sans maladie/progression ([Sargent et al., 2005](#)). Dans ce contexte, la qualité de vie est utilisée comme un critère de jugement pour s'assurer du bénéfice clinique pour le patient.

La qualité de vie relative à la santé (QdV ou en anglais HRQoL pour Health Related Quality of Life) est une mesure de l'état de santé du patient directement reportée par le patient lui-même et entre ainsi dans le champ des « Patients-Reported Outcomes » (PROs), dont l'utilisation représente un intérêt grandissant auprès de la Food and Drug Administration (FDA) ([Administration and others, 2019](#)). En général, la QdV est utilisée comme critère de jugement secondaire dans les essais, mais tend à s'imposer de plus en plus comme critère ou co-critère principal en gériatrie ([Pallis et al., 2011](#)) et/ou en soins de support, mais aussi dans certaines localisations spécifiques. Par exemple, il est actuellement recommandé d'utiliser la QdV en tant que co-critère principal avec un critère centré sur la tumeur dans le cancer colorectal métastatique en première ligne de traitement ([Bonnetain et al., 2017](#)). De même, il est également recommandé dans le cancer épithélial de l'ovaire, d'utiliser la QdV comme critère principal ou co-critère principal avec la survie sans progression pour des patientes résistantes aux sels de platine ([Joly et al., 2017](#)). La QdV apparaît donc comme un critère majeur utilisé dans les essais cliniques en cancérologie.

## 1.1 Qualité de vie relative à la santé

### 1.1.1 Définition

La constitution de l'Organisation Mondiale de la Santé (OMS), adoptée en 1948, déclare que la santé est « un état complet de bien-être physique, mental et social, et ne consiste pas seulement en une absence de maladie ou d'invalidité » ([Organization et al., 1948](#)). Plus tard en 1993, l'OMS définit la qualité de vie comme « la perception qu'a un individu de sa place dans l'existence, dans le contexte de la culture et du système de valeurs dans lequel il vit, en relation avec ses objectifs, ses attentes, ses normes et ses inquiétudes. Il s'agit donc d'un large champ concep-

## Chapitre 1. Introduction

---

tuel, englobant de manière complexe la santé physique de la personne, son état psychologique, son niveau d'indépendance, ses relations sociales, ses croyances personnelles et sa relation avec les spécificités de son environnement » (Group, 1993). Dans le contexte médical, la QdV est utilisée comme une mesure subjective de l'état de santé perçu par le patient pour rendre compte de l'impact de la maladie et du traitement sur son bien-être. Cette notion est cependant complexe et difficile à appréhender.

D'abord, la QdV est notamment influencée par la condition physique et l'environnement social, financier et émotionnel ce qui en fait un concept très général de nature multidimensionnelle. Ces différentes dimensions sont ainsi associées aux conséquences directes du traitement, mais aussi aux conséquences indirectes induites par la maladie, comme des difficultés financières. La QdV est également un concept subjectif. Chaque individu dispose de sa propre perception et définition de la QdV (par exemple, la perception de la douleur) ce qui se traduit par une forte variabilité inter-individuelle. La QdV est enfin un concept dynamique qui peut évoluer dans le temps, en se dégradant ou s'améliorant en fonction de l'évolution de la maladie, l'impact du traitement ainsi que les attentes et espérances du patient au cours de la maladie. Notons que l'appréciation de la QdV ne se fait pas forcément selon les mêmes critères au cours du temps ce qui se traduit par une variabilité intra-individuelle d'une mesure à l'autre (*response shift*).

### 1.1.2 Instruments de mesure

La QdV n'étant pas directement mesurable, des instruments de mesure standardisés intégrant les difficultés conceptuelles liées à la QdV (multidimensionnelle, subjective et dynamique) ont été proposés pour retranscrire au mieux la perception du patient quant à l'impact de la maladie et du traitement sur son bien-être. En cancérologie, les questionnaires sont spécifiques à la maladie et les patients sont interrogés sur la fréquence et la gravité de leurs symptômes et de leurs capacités. Les questionnaires sont administrés aux patients sous la forme d'auto-questionnaires, qui sont des questionnaires remplis par les patients eux-mêmes. Ces instruments génériques de mesure sont validés et respectent des propriétés psychométriques garantissant la facilité de compréhension et une faible variabilité

## 1.1. Qualité de vie relative à la santé

**Tableau 1.1 – Caractéristiques du questionnaire EORTC QLQ-C30 (version 3.0).**

	Légende	Nombre d'item(s)	Numéro item(s)	Nombre de modalités par item
QdV/santé globale des patients				
QdV/santé globale des patients	QL	2	29, 30	7
Echelles fonctionnelles				
Fonction physique	PF	5	1 - 5	4
Fonction personnelle	RF	2	6 - 7	4
Fonction émotionnelle	EF	4	21 - 24	4
Fonction cognitive	CF	2	20, 25	4
Fonction sociale	SF	2	26, 27	4
Echelles/items symptomatiques				
Fatigue	FA	3	10, 12, 18	4
Nausées et vomissements	NV	2	14, 15	4
Douleur	PA	2	9, 19	4
Dyspnée	DY	1	8	4
Insomnie	SL	1	11	4
Perte d'appétit	AP	1	13	4
Constipation	CO	1	16	4
Diarrhées	DI	1	17	4
Difficultés financières	FI	1	28	4

d'interprétation. Les principaux questionnaires dans le domaine de la cancérologie sont ceux de l'EORTC (European Organisation of Research and Treatment of Cancer) et du groupe FACT (Functional Assessment of Cancer Therapy). Nous nous concentrerons uniquement sur les questionnaires de l'EORTC, qui sont les plus utilisés en Europe alors que l'utilisation des questionnaires FACT est plus répandue aux Etats-Unis.

### Questionnaire EORTC QLQ-C30

Le questionnaire EORTC QLQ-C30 (Quality of Life Questionnaire Core 30) est un questionnaire standard spécifique au cancer et administré aux patients pour évaluer leur QdV ([Aronson et al., 1993](#)) (voir Annexe A). Dans sa version la plus récente (version 3.0), il est disponible dans une centaine de langues différentes. Ce questionnaire est notamment utilisé dans des essais cliniques de phase 2 ou 3 et peut être administré quel que soit le type de cancer. Le QLQ-C30 est composé de 30 questions (items) qui permettent d'évaluer 15 dimensions de la QdV, réparties en 5 échelles fonctionnelles, 9 échelles symptomatiques et 1 échelle pour l'état de QdV/santé globale des patients (voir Tableau 1.1).

Les dimensions fonctionnelles regroupent les échelles de fonctions : physique (4 items), personnelle (2 items), émotionnelle (4 items), cognitive (2 items) et so-

## Chapitre 1. Introduction

**Tableau 1.2 – Caractéristiques du questionnaire EORTC QLQ-OES18.**

	Légende	Nombre d'item(s)	Numéro item(s)	Nombre de modalités par items
Echelles/items symptomatiques				
Dysphagie	OESDYS	3	1 - 3	4
Faim	OESEAT	4	6 - 9	4
Reflux	OESRFX	2	14, 15	4
Douleur	OESPA	3	16 - 18	4
Difficulté à avaler de la salive	OESSV	1	4	4
Étouffement par la déglutition	OESCH	1	5	4
Sécheresse de la bouche	OESDM	1	10	4
Problèmes de goût	OESTA	1	11	4
Difficulté à tousser	OESCO	1	12	4
Difficulté à parler	OESSP	1	13	4

ciale (2 items). Les dimensions symptomatiques regroupent les échelles liées : à la fatigue (3 items), aux nausées et vomissements (2 items), à la douleur (2 items), à la dyspnée (1 item), à la perte d'appétit (1 item), à l'insomnie (1 item), à la constipation (1 item), aux diarrhées (1 item) et aux difficultés financières (1 item). Les items 1 à 28 sont construits selon une échelle à quatre modalités de réponse (1 « pas du tout » ; 2 « un peu » ; 3 « assez » ; 4 « beaucoup ») alors que les deux derniers items, évaluant l'état de QdV/santé globale, sont eux construits sur une échelle à 7 modalités de réponse (de 1 « très mauvais » à 7 « excellent »).

### Modules spécifiques

Des modules supplémentaires spécifiques à la localisation cancéreuse sont également disponibles. Ils sont le plus souvent donnés en complément du QLQ-C30, comme les modules additionnels QLQ-OES18 (associé au cancer de l'œsophage) ou QLQ-BR23 (associé au cancer du sein). À titre d'exemple, le questionnaire QLQ-OES18 évalue 10 échelles symptomatiques supplémentaires liées au cancer de l'œsophage et aux effets secondaires du traitement ; il évalue plus précisément la dysphagie, la déglutition, des symptômes gastro-intestinaux et abdominaux et des difficultés alimentaires ([Blazeby et al., 2003](#)) (voir Tableau 1.2).

### Calcul des scores

Pour chaque dimension décrite dans les questionnaires (QLQ-C30 ou modules spécifiques additionnels), il est possible de calculer un score standardisé de 0 à

100 à partir des réponses aux items, en suivant les recommandations de l'EORTC (procédure de *scoring*) (Fayers et al., 2001). Les scores ainsi calculés ont l'avantage d'être facilement comparables (d'une dimension à l'autre ou d'un questionnaire à l'autre) et plus facilement interprétables en vue d'une analyse statistique.

Soit  $J$  le nombre d'items à  $M$  modalités de réponse et  $I_j \in 1, \dots, M$  la réponse à l'item  $j$  ( $j = 1, \dots, J$ ). Pour les échelles symptomatiques et l'état de QdV/santé globale des patients le score  $S$  est égal à :

$$S = \frac{\left(\frac{1}{J} \sum_{j=1}^J I_j\right) - 1}{M - 1} \times 100.$$

Pour les échelles fonctionnelles le score  $S$  est égal à :

$$S = \left[ 1 - \frac{\left(\frac{1}{J} \sum_{j=1}^J I_j\right) - 1}{M - 1} \right] \times 100.$$

Pour les dimensions fonctionnelles ou la QdV/santé globale des patients, un score élevé correspond à un haut niveau de capacités fonctionnelles (physique, sociale, ...) ou un haut niveau de QdV/santé globale et donc reflète un bon niveau de qualité de vie. À l'inverse, pour les dimensions symptomatiques, un score élevé correspond à de forts symptômes (fatigue, douleur, ...) et donc reflète un mauvais niveau de qualité de vie. L'échelle de scores est donc inversée entre les dimensions fonctionnelles ou la QdV/santé globale et les dimensions symptomatiques.

En cas d'absence de réponse à certains items, le manuel pour calculer les scores de QdV recommande de calculer un score pour une dimension uniquement si au moins la moitié des items de la dimension concernée est renseignée. Ainsi, le score associé à la dimension se calcule tel que présenté précédemment, mais uniquement sur la base des items renseignés (en ignorant les items manquants). Sinon, le score associé à la dimension est considéré comme manquant.

Par exemple, prenons le cas de l'échelle fonctionnelle du QLQ-C30 à  $J = 4$  items et  $M = 4$  modalités où les deux premiers items sont renseignés ( $I_1$  et  $I_2$ )



alors que les deux derniers sont manquants ( $I_3$  et  $I_4$ ). Dans ce cas, la moitié des items pour cette échelle fonctionnelle étant renseignés, le score  $S$  est égal à :

$$S = \left[ 1 - \frac{\left(\frac{1}{2}(I_1 + I_2)\right) - 1}{4 - 1} \right] \times 100.$$

## 1.2 Analyse longitudinale de la QdV dans les essais cliniques

Dans les essais cliniques, les questionnaires de QdV sont généralement collectés à l'inclusion puis tout au long de la prise en charge du patient (pendant la période de traitement et au cours du suivi). L'administration des questionnaires à différents temps génère une répétition de mesures pour un même patient qui d'un point de vue statistique se traite comme des mesures répétées et corrélées entre elles. Pour en tenir compte, il est indispensable d'utiliser des méthodes d'analyse longitudinale adaptées à ce type de données, telles que les modèles linéaires mixtes (Fairclough, 2010; Brown and Prescott, 2014).

### 1.2.1 Modèles linéaires mixtes

Les modèles linéaires à effets aléatoires (modèles linéaires mixtes) introduits par (Laird and Ware, 1982) sont aujourd'hui très utilisés pour l'étude d'une variable quantitative gaussienne au cours du temps. Ils reposent sur deux niveaux d'analyse différents : un niveau global, commun à l'ensemble de la population décrit par des effets fixes et un niveau individuel, décrit par des effets aléatoires. L'introduction d'effets aléatoires permet de tenir compte de la corrélation intra-individuelle des données répétées et de l'hétérogénéité des patients. Le terme « linéaire » signifie ici « linéaire en les paramètres » dont l'une des caractéristiques est que l'espérance et la variance du modèle s'expriment comme des fonctions linéaires des paramètres. Ainsi, il est possible de décrire des trajectoires de diverses formes sans se limiter forcément à la modélisation de trajectoires linéaires dans le temps, ce qui en fait des modèles très flexibles et très utilisés.

## 1.2. Analyse longitudinale de la QdV dans les essais cliniques

Dans une population de  $n$  individus, soit le vecteur  $Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{in_i}))$  avec  $Y_i(t_{ij})$  la variable aléatoire d'une quantité d'intérêt de l'individu  $i$  mesurée au temps  $t_{ij}$  pour  $j = 1, \dots, n_i$  où  $n_i$  est le nombre de mesures répétées de l'individu  $i$ . Le modèle linéaire mixte s'écrit comme suit :

$$Y_i = X_i\beta + Z_ib_i + \varepsilon_i, \quad (1.1)$$

où  $X_i$  est une matrice de dimension  $n_i \times p$  de variables explicatives (incluant notamment le temps) associée au vecteur d'effets fixes  $\beta$ ,  $Z_i$  est une sous-matrice de  $X_i$  de dimension  $n_i \times q$  (avec  $q \leq p$ , le nombre d'effets aléatoires) associée au vecteur d'effets aléatoires  $b_i$  et  $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})$  est le vecteur des erreurs de mesure. Les effets aléatoires sont supposés être distribués selon une loi normale d'espérance zéro et de matrice de variance-covariance  $D : b_i \sim \mathcal{N}(0, D)$ . Le vecteur des erreurs de mesure  $\varepsilon_i$  est supposé indépendant des effets aléatoires  $b_i$  et distribué selon une loi normale d'espérance zéro et de matrice de variance-covariance  $\Sigma_i : \varepsilon_i \sim \mathcal{N}(0, \Sigma_i)$ . Les erreurs de mesures sont généralement supposées mutuellement indépendantes et de même variance  $\sigma^2$  et la matrice de variance-covariance des erreurs de mesures se simplifie par  $\Sigma_i = \sigma^2 I_{n_i}$ . Dans certaines situations, il peut être utile de les supposer corrélées entre elles ou bien ajouter aux erreurs indépendantes une structure entre les mesures répétées d'un même individu  $i$ , par exemple en utilisant un processus autorégressif gaussien (Diggle, 2009).

La partie fixe  $X_i\beta$  représente l'évolution moyenne de la quantité d'intérêt au cours du temps et la partie aléatoire  $Z_ib_i$  les déviations individuelles par rapport au terme moyen. Le terme  $X_i\beta + Z_ib_i$  représente la valeur de la quantité d'intérêt au cours du temps, mais débarrassée de l'erreur de mesure, que l'on nomme souvent la « vraie » valeur non observée. En utilisant l'Équation 1.1, on peut montrer que la distribution marginale du vecteur  $Y_i$  est une loi normale multivariée de taille  $n_i$  :

$$Y_i \sim N(X_i\beta, V_i = Z_i D Z_i^\top + \Sigma_i) \quad (1.2)$$

de moyenne  $X_i\beta$  et de matrice de variance-covariance  $V_i = Z_i D Z_i^\top + \Sigma_i$ . L'approche classique pour estimer les paramètres du modèle est basée sur le maximum de vraisemblance (ML) ou bien le maximum de vraisemblance restreint (REML), dont les détails techniques d'optimisation et les algorithmes utilisés sont connus

## Chapitre 1. Introduction

---

et clairement décrits dans la littérature (Verbeke and Molenberghs, 2009; Pinheiro et al., 2000).

Dans le contexte d'analyse de la qualité de vie dans les essais cliniques, les modèles linéaires mixtes sont fréquemment utilisés pour l'analyse longitudinale des scores de QdV (Ferrandina et al., 2014; Pan et al., 2012; Schaake et al., 2013). Désignons par  $Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{in_i}))$  les scores de QdV du patient  $i$  aux temps de mesure  $t_{ij}$ ,  $j = 1, \dots, n_i$ .

Dans le cas de deux groupes randomisés, la modélisation classique pour comparer l'évolution de la QdV entre deux bras de traitement consiste à considérer un modèle linéaire mixte à ordonnée à l'origine (*intercept*) et pente aléatoires. De façon générale, le score de QdV de l'individu  $i$  au temps  $t_{ij}$  peut s'écrire alors :

$$Y_i(t_{ij}) = \beta_0 + \beta_1 \text{arm}_i + \beta_2 t_{ij} + \beta_3 (t_{ij} \times \text{arm}_i) + b_{0i} + b_{1i} t_{ij} + \varepsilon_i(t_{ij})$$

$$\text{avec } b_i = (b_{0i}, b_{1i})^\top \sim N\left(0, D = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right) \text{ et } \varepsilon_i \sim N(0, \Sigma_i = \sigma^2 I_{n_i})$$

où :

- $\beta_0$  est l'ordonnée à l'origine du modèle dans le bras de référence ( $\text{arm}_i = 0$ ) ;
- $\beta_0 + \beta_1$  est l'ordonnée à l'origine du modèle dans le bras expérimental ( $\text{arm}_i = 1$ ) ;
- $\beta_2$  est la pente dans le bras de référence ( $\text{arm}_i = 0$ ) ;
- $\beta_3$  est la différence de pente entre les deux bras et  $\beta_2 + \beta_3$  correspond à la pente dans le bras expérimental ( $\text{arm}_i = 1$ ) ;
- $b_i = (b_{0i}, b_{1i})^\top$  est le vecteur des effets aléatoires avec  $b_{0i}$  et  $b_{1i}$  les effets aléatoires liés à l'ordonnée à l'origine et à la pente.

Cette modélisation permet de pouvoir facilement comparer les évolutions moyennes entre les deux bras de traitement grâce aux effets fixes et l'introduction d'effets aléatoires permet de modéliser la variabilité entre les individus. En particulier, l'effet aléatoire  $b_{0i}$  permet de considérer que les individus n'ont pas les

## 1.2. Analyse longitudinale de la QdV dans les essais cliniques

mêmes valeurs de scores à  $t = 0$  (à l'inclusion) et l'effet aléatoire  $b_{1i}$  permet de considérer que le score n'a pas la même évolution entre les individus.

En cancérologie, il est fréquent de rencontrer un modèle très similaire à celui présenté au modèle (1.2), en supprimant l'effet du bras de traitement à l'inclusion (effet  $\beta_1$ ). Avant de commencer leur traitement, les patients recrutés dans l'essai clinique forment, du fait de la randomisation, deux groupes assez homogènes tant sur leur maladie (type de cancer, stades d'évolution) que sur leurs symptômes. Par conséquent, l'ajout de cet effet du bras présenterait peu d'intérêt dans la modélisation, si ce n'est de confirmer que les deux groupes de traitement sont comparables à l'inclusion. C'est la raison pour laquelle nous nous intéresserons plutôt au modèle qui omet cet effet du bras. Dans le contexte d'essais cliniques randomisés en cancérologie, le score de QdV de l'individu  $i$  au temps  $t_{ij}$  peut finalement se simplifier par :

$$Y_i(t_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 (t_{ij} \times \text{arm}_i) + b_{0i} + b_{1i} t_{ij} + \varepsilon_i(t_{ij}) \quad (1.3)$$

avec  $b_i = (b_{0i}, b_{1i})^\top \sim N\left(0, D = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right)$  et  $\varepsilon_i \sim N(0, \Sigma_i = \sigma^2 I_{n_i})$

où :

- $\beta_0$  est l'ordonnée à l'origine du modèle ;
- $\beta_1$  est la pente dans le bras de référence ( $\text{arm}_i = 0$ ) ;
- $\beta_2$  est la différence de pente entre les deux bras et  $\beta_1 + \beta_2$  correspond à la pente dans le bras expérimental ( $\text{arm}_i = 1$ ) ;
- $b_i = (b_{0i}, b_{1i})^\top$  est le vecteur des effets aléatoires avec  $b_{0i}$  et  $b_{1i}$  les effets aléatoires liés à l'ordonnée à l'origine et à la pente.

Dans la suite de ce manuscrit, c'est généralement à ce modèle (1.3) que nous ferons référence dès que nous évoquerons l'analyse des scores de QdV par des modèles linéaires mixtes.

### 1.2.2 Problématique des données manquantes

La collecte des données de QdV dans les essais cliniques est devenue de plus en plus courante, mais la présence de données manquantes est inéluctable et les données de QdV sont la plupart du temps incomplètes (par opposition aux données complètes, en l'absence de données manquantes). En effet, des questionnaires de QdV peuvent être incomplets (trop d'items mal ou non renseignés), voire manquants à certaines visites (non complétés, non rendus ou non transmis au patient) ce qui ne permet pas le calcul du score de QdV selon les recommandations en vigueur.

Actuellement, les données manquantes de QdV sont décrites en détaillant les caractéristiques à l'inclusion des patients ayant des données manquantes, en reportant la compliance (nombre de questionnaires reçus par rapport au nombre de questionnaires attendus) pendant toute la durée de l'essai ([Calvert et al., 2013](#)), mais les données manquantes sont généralement ignorées dans la modélisation. Les problématiques liées à la présence de données manquantes est un point soulevé et décrit dans la littérature, néanmoins trop peu considéré dans les analyses des données de QdV issues des essais cliniques. Pourtant, les patients qui sont à des stades très avancés de la maladie sont souvent les moins susceptibles de remplir les questionnaires transmis, en raison d'évènements tels que les toxicités du traitement ou la progression de la maladie. C'est pourquoi il est tout à fait plausible que l'absence de questionnaire soit liée au niveau de la QdV du patient au moment de l'évaluation ([Bernhard et al., 1998](#)) et qu'il faille ainsi utiliser des méthodes adaptées à cette situation.

## 1.3 Essai clinique PRODIGE 5/ACCORD 17

Tout au long de ce manuscrit, nous exploiterons les données de l'essai clinique PRODIGE (Partenariat de Recherche en Oncologie DIGEstive) 5/ACCORD (Actions Concertées dans les Cancers Colorectaux et Digestifs) 17 afin d'illustrer les différentes méthodes proposées. L'essai clinique PRODIGE 5/ACCORD 17 est un essai multi-centrique randomisé de phases 2/3 pour des patients atteints d'un can-

cer de l'œsophage localement avancé. Cet essai clinique a été initié pour comparer deux chimiothérapies administrées en concomitance avec la radiothérapie, à savoir une chimiothérapie à base d'oxaliplatine, acide folinique et 5-fluorouracile cisplatine (FOLFOX) et une chimiothérapie de fluorouracile et cisplatine (5FU-CISPLATIN). L'essai a inclus au total 267 patients dont 134 patients dans le bras expérimental (FOLFOX) et 133 dans le bras contrôle (5FU-CISPLATIN) ([Conroy et al., 2014](#)).

L'objectif principal de l'essai de phase 3 consistait en la comparaison de la survie sans progression (un critère centré sur la tumeur) entre les deux groupes : radiochimiothérapie avec du FOLFOX (traitement expérimental) versus radiochimiothérapie avec du 5FU-CISPLATIN (traitement standard). Le traitement expérimental n'a pas montré de bénéfice sur la survie sans progression ; cependant, l'administration du FOLFOX en ambulatoire est plus facile pour les patients et moins de décès liés aux toxicités ont été relevés par rapport au 5FU-CISPLATIN. Pour ces raisons, le FOLFOX a donc été proposé comme le nouveau traitement standard pour les patients atteints d'un cancer de l'œsophage localement avancé.

La QdV est l'un des critères secondaires de cet essai. Elle a été évaluée par le questionnaire EORTC QLQ-C30 (version 3.0) ainsi que par le module complémentaire EORTC QLQ-OES18, spécifique aux cancers de l'œsophage. Ce sont les patients qui remplissaient eux-mêmes les questionnaires à l'hôpital, au cours du traitement et de la prise en charge. Il y avait 8 évaluations, planifiées à l'inclusion, au dernier jour de la radiothérapie, au premier jour de la dernière chimiothérapie (semaine 10 ou 11), à la semaine 15 et enfin pendant la période de suivi à 6, 12, 24 et 36 mois. Selon le protocole de l'essai, la collecte des mesures de QdV a ainsi été poursuivie au-delà de la progression de la maladie pendant un délai de 36 mois après l'inclusion. Parmi les 30 dimensions du QLQ-C30, 4 échelles étaient identifiées comme particulièrement pertinentes : la QdV/santé globale des patients (QL), l'échelle fonctionnelle de capacité physique (PF) ainsi que les échelles symptomatiques de fatigue (FA) et de douleur (PA). Les distributions des scores pour ces dimensions à l'inclusion sont données en Annexe B. L'analyse longitudinale des données de QdV a montré une détérioration de la QdV pendant la période de traitement, mais aussi que les patients retrouvaient leur niveau initial de QdV pendant leur suivi ([Bascoul-Mollevis et al., 2017](#)). Dans ce même article, les analyses n'ont

pas montré de différence d'évolution de QdV entre les deux traitements (pas d'effet du bras de traitement). Comme il a été judicieusement discuté dans l'article, il y a une forte proportion de données manquantes ce qui pourrait compromettre les résultats obtenus. Une attention particulière a été portée sur la description des données manquantes : la proportion des données manquantes est équilibrée entre les deux bras de traitement et la comparaison des patients ayant des données manquantes par rapport à ceux qui n'en n'ont pas, n'a montré aucune différence sur leurs caractéristiques initiales. Néanmoins, de nombreux questionnaires de QdV sont manquants avec une faible compliance atteignant seulement 41 % après 6 mois de suivi. La forte proportion de données manquantes dans cet essai, incluant des patients à un stade avancé de cancer de l'œsophage, pose la question de la robustesse des résultats obtenus. Par conséquent, cette thèse portera sur l'analyse de la QdV en présence de données manquantes et sera illustrée sur les données de cet essai clinique.

### 1.4 Objectifs et plan

Mon travail de thèse s'est concentré sur les méthodes statistiques pour l'analyse longitudinale de la QdV dans les essais cliniques en cancérologie. Plus particulièrement, la problématique majeure de ce travail concerne une situation très souvent rencontrée dans l'étude de la qualité de vie dans les essais cliniques, à savoir, la présence de données manquantes dans les suivis longitudinaux. En effet, les données manquantes, lorsqu'elles sont associées avec la variable d'intérêt, nécessitent d'utiliser des méthodes d'analyse statistique adaptées, pour ne pas obtenir des résultats incorrects et biaisés (données manquantes informatives). Nous avons choisi de considérer seulement le cas des données manquantes de type monotone (sorties d'étude) dont la prise en compte est un véritable enjeu au moment de l'analyse statistique. L'objectif principal de mon travail de thèse a été d'étudier des méthodes statistiques adéquates dans cette situation en expliquant et en justifiant l'intérêt de leur utilisation afin de proposer des recommandations au moment de l'analyse pour prendre en compte les sorties d'étude informatives. Bien que ce travail naisse directement des problématiques rencontrées dans l'analyse des données longitudinales de QdV, il concerne plus généralement n'importe quelle application où la

variable d'intérêt est une PRO et où les sorties d'étude auraient un possible lien avec la variable d'intérêt étudiée.

Le début de ce manuscrit commence par un état de l'art sur les données manquantes (Chapitre 2). Il introduit principalement les trois mécanismes de données manquantes, les outils statistiques pour les détecter ainsi que les méthodes à utiliser en présence de sorties d'étude informatives dont la prise en compte est un véritable enjeu au moment de l'analyse statistique. Ce travail préliminaire a fait l'objet d'un chapitre rédigé de livre dans lequel trois approches sont présentées pour tenir compte des sorties d'étude informatives dans les analyses longitudinales : les modèles de sélection, les modèles de mélange de profils et les modèles à effets partagés.

La suite de ce travail, décrit dans le Chapitre 3, s'intéresse à l'application de ces trois approches pour l'analyse des données de QdV. L'objectif est de les comparer au modèle linéaire mixte, qui est la modélisation classiquement utilisée et qui ignore la présence de données manquantes. Les avantages et les inconvénients de chacune de ces approches sont explicités d'un point de vue méthodologique puis illustrés sur les données de l'essai clinique PRODIGE 5/ACCORD 17 pour l'analyse des données de QdV. Ce travail a fait l'objet d'un premier article intitulé « *Handling informative dropout in longitudinal analysis of health-related quality of life: application of three approaches to data from the esophageal cancer clinical trial PRODIGE 5/ACCORD 17* », publié dans la revue *BMC Medical Research Methodology*. Dans cet article, nous avons souligné l'intérêt et la pertinence des modèles conjoints à effets partagés. La suite de ce manuscrit sera plus particulièrement consacrée à leur étude.

Les modèles conjoints (à effets partagés) nécessitent la formulation de deux sous-modèles. Le principe est de modéliser la variable longitudinale (score de QdV) par un modèle linéaire mixte, de modéliser le processus à l'origine des données manquantes (temps jusqu'à sortie d'étude) par un modèle de survie, et d'associer les deux modèles par une structure latente commune capable de capturer l'association entre les deux processus. Le Chapitre 4 est dédié à l'étude du sous-modèle qui décrit le processus des données manquantes, en permettant de distinguer différentes causes de sortie d'étude grâce à un modèle à risques concurrents



## Chapitre 1. Introduction

---

pour le temps jusqu'à sortie d'étude. Un second article intitulé « *Joint modelling with competing risks of dropout for longitudinal analysis of health-related quality of life in cancer clinical trials* » et prêt à être soumis, propose une étude de simulations afin d'étudier la capacité des modèles conjoints à distinguer les sorties d'étude informatives des sorties d'étude non informatives et à donner des estimations valides pour les trajectoires de QdV. Ceci est particulièrement utile en vue d'une application sur les données réelles, lorsque les sorties d'étude sont supposées être associées différemment à la QdV. L'utilisation d'un tel modèle conjoint en considérant deux causes possibles de sorties d'étude a également été appliqué aux données de l'essai clinique PRODIGE 5/ACCORD 17.

Après avoir utilisé un sous-modèle plus sophistiqué pour le temps jusqu'à sortie d'étude, nous avons envisagé de nous intéresser au sous-modèle décrivant la variable longitudinale. Afin de ne pas se restreindre à des trajectoires uniquement linéaires, une modélisation plus flexible basée sur des splines est utilisée dans le Chapitre 5 pour les trajectoires de scores de QdV dans l'essai clinique PRODIGE 5/ACCORD 17. Les données de scores ont jusqu'alors toujours été considérées comme des variables continues et gaussiennes. Afin de s'affranchir de cette hypothèse, la dernière partie de ce chapitre est consacrée à la modélisation des scores de QdV en tant que données catégorielles présentant une notion d'ordre (données ordinales). Ceci est particulièrement utile pour certaines dimensions du QLQ-C30 qui ne présentent que quelques modalités de réponses et pour lesquelles l'hypothèse de normalité est difficilement vérifiée.

Enfin, nous terminerons ce travail par un chapitre synthétisant les principales conclusions pour l'analyse des données de QdV en présence de données manquantes, ainsi que les perspectives envisagées de ce travail.

# Chapitre 2

## Données manquantes

---

### 2.1 Introduction

Dans presque toutes les disciplines, la présence de données manquantes est un problème courant et difficile à résoudre et peut compliquer l'analyse des données longitudinales. Le terme de « données manquantes » est alors utilisé pour indiquer que certaines mesures prévues (par exemple par un protocole) n'ont pas été observées. Un cadre théorique rigoureux a été développé pour modéliser la survenue des données manquantes, en particulier, une terminologie, des notations et des mécanismes pour les décrire ([Little, 1988](#); [Rubin, 1976](#)). Depuis les années 1990, le traitement statistique des données manquantes a connu de nombreux développements, notamment autour des modèles utilisés dans le contexte de données longitudinales, là où les données manquantes et en particulier les sorties d'étude sont le plus fréquemment rencontrées ([Molenberghs, 2014](#); [Fitzmaurice et al., 2009](#); [Molenberghs and Kenward, 2007](#); [Verbeke and Molenberghs, 2009](#)). Dans ce chapitre, nous avons souhaité faire un état de l'art de la littérature pour traiter le problème des données manquantes. Ce travail a fait l'objet d'un chapitre de livre (Chapitre *Données manquantes*<sup>1</sup>) qui se concentre sur la présence de données manquantes

---

1. Mathoulin-Pelissier S, Gourgou S, Paoletti X. *Méthodes biostatistiques appliquées à la recherche clinique en cancérologie (2e édition)*. Arcueil : John Libbey Eurotext, 2021, (à paraître)

## Chapitre 2. Données manquantes

---

dans les essais cliniques, en s'appuyant sur l'exemple du recueil des données de QdV. Nous avons décrit brièvement les trois mécanismes associés aux données manquantes, leurs hypothèses et leurs implications sur les stratégies d'analyse. Nous avons souligné les difficultés liées aux données manquantes puis présenté les trois méthodes décrites dans la littérature pour tenir compte de la présence de sorties d'étude informatives pour l'analyse longitudinale de la QdV.

### 2.2 Chapitre de livre : *Données manquantes*

# Données manquantes

Benjamin Cuer, Célia Touraine, Caroline Mollevi

## 1. Introduction

Dans les essais cliniques, il est fréquent de collecter des données de manière répétée pour étudier une variable d'intérêt au cours du temps. Une situation souvent rencontrée et inévitable dans les suivis longitudinaux est l'apparition de données manquantes. Les données disponibles sont donc dites incomplètes. Il existe des recommandations pour tenir compte des données manquantes et limiter leur occurrence dans les essais cliniques [1], [2]. Une fois l'essai clinique initié, Little et al. [2] recommandent de contrôler le taux de données manquantes au cours du temps par bras de traitement et d'en collecter si possible la raison. Des moyens supplémentaires peuvent ainsi être mis en œuvre pour réduire en temps réel l'occurrence de certaines données manquantes tout en s'assurant d'une répartition homogène dans chaque bras de traitement. Selon la raison pour laquelle une donnée est manquante (ex. : erreur d'organisation indépendante du patient, refus du patient ou patient trop malade), il peut être nécessaire d'en tenir compte pour ne pas obtenir de résultats incorrects, à la fois dans la conception et l'analyse statistique de l'étude [3]. Avant l'analyse, il est recommandé d'évaluer la quantité et le type de données manquantes, de façon à utiliser des méthodes adaptées. La stratégie d'analyse générale doit par ailleurs être spécifiée en amont dans le protocole.

Dans ce chapitre, nous décrivons les stratégies d'analyse à adopter en présence de données manquantes de la variable d'intérêt selon leur type en détaillant le cas des données manquantes potentiellement informatives. Nous illustrerons nos propos en considérant l'exemple du recueil de données de qualité de vie relative à la santé (QdV) dans les essais cliniques où la présence de données manquante est un véritable enjeu. La QdV y est évaluée de manière répétée au cours de la prise en charge du patient à l'aide d'auto-questionnaires standardisés. Des scores de QdV sont ensuite calculés selon les recommandations associées au questionnaire utilisé et analysés en tant que variables quantitatives.

## 2. Différents types de données manquantes

Avant de discuter du choix des méthodes d'analyse statistique à considérer en présence de données manquantes, il est important de bien définir la terminologie utilisée pour les caractériser.

D'abord, il est possible de distinguer deux types de données manquantes selon leur répartition : les données manquantes peuvent être manquantes de façon intermittente ou monotone. Les données manquantes sont dites monotones lorsqu'elles sont manquantes à partir d'un certain temps et jusqu'à la fin de l'étude. C'est le cas lorsqu'un patient sort prématurément de l'étude : les questionnaires de QdV ne sont donc plus disponibles à partir de ce moment-là. L'autre situation est celle des données manquantes dites intermittentes correspondant à des questionnaires ou items manquants par intermittence. En pratique, c'est souvent une combinaison de ces deux types de données manquantes qui est rencontrée.

Ensuite, il est également important de définir le type de mécanisme sous-jacent à cette non-réponse. Selon Rubin [4], il est possible d'en distinguer trois : les données peuvent être manquantes complètement au hasard (MCAR pour *Missing Completely At Random*), manquantes au hasard (MAR pour *Missing At Random*) ou manquantes non au hasard (MNAR pour *Missing Not At Random*). En présence de données manquantes de la variable d'intérêt, les modèles pour son analyse statistique doivent être choisis en fonction du type de mécanisme suspecté.

### 2.1 Données manquantes de type MCAR

#### Définition

Les données manquantes de la variable d'intérêt sont dites MCAR lorsque la probabilité qu'elles soient manquantes est indépendante des autres valeurs de la variable, observées ou non observées. Par exemple, le fait que l'équipe médicale oublie de transmettre le questionnaire de QdV au patient n'est *a priori* pas lié à sa QdV et est donc indépendant des données de score, observées ou non. Little [5] a également introduit le concept de données manquantes dépendantes de variables explicatives. Les données manquantes sont définies comme telles si, conditionnellement à des variables explicatives observées à l'inclusion, elles peuvent

être considérées MCAR. Supposer que les données manquantes sont MCAR ou dépendent seulement de variables explicatives est une hypothèse très forte et en général peu crédible dans les essais cliniques.

### **Méthodes d'analyse**

Sous l'hypothèse MCAR, même les méthodes qui n'utilisent pas toute les données disponibles mais seulement les cas complets (les données des patients ayant des données à tous les temps d'évaluation) de type MANOVA peuvent être appliquées sans conséquence particulière si ce n'est une perte de puissance statistique [6]. En effet, la sous-population utilisée est représentative de la population totale. De la même manière, il est possible de réaliser des analyses transversales de la variable d'intérêt, c'est-à-dire des analyses à un temps d'évaluation donné.

### **Identifier les variables explicatives du mécanisme de données manquantes**

Afin d'explorer le mécanisme de données manquantes, il faut d'abord définir la variable de données manquantes, *i.e.* de l'absence de réponse à la variable d'intérêt. Pour une variable d'intérêt de QdV, ce serait une variable indicatrice de non-réponse au questionnaire à une visite spécifique ou à plusieurs visites. Les variables associées à la probabilité d'être manquant peuvent ensuite être identifiées avec des calculs de corrélation via le tau de Kendall ou à l'aide de régressions logistiques. Dans le cas de données manquantes de structure monotone, les variables expliquant le mécanisme de sortie d'étude pourraient être identifiées en appliquant un modèle de Cox à risques proportionnels du temps jusqu'à sortie d'étude, la variable de données manquantes étant l'indicatrice de sortie d'étude [6].

En présence de données manquantes, il est opportun de réaliser une analyse préliminaire pour identifier les variables expliquant le mécanisme de données manquantes, celle-ci devant se focaliser en priorité sur les caractéristiques initiales des patients qui sont potentiellement liées à la variable d'intérêt.

## **2.2 Données manquantes de type MAR**

### **Définition**

Les données manquantes de la variable d'intérêt sont dites MAR lorsque, conditionnellement aux variables explicatives, la probabilité qu'elles soient manquantes dépend des autres valeurs observées de la variable. Par exemple, les patients dont le score de QdV observé à l'inclusion est faible pourraient être les plus susceptibles de ne pas remplir les questionnaires au cours du suivi.

### **Etudier la validité de l'hypothèse MCAR versus MAR**

La dépendance du mécanisme de données manquantes avec les données observées de la variable d'intérêt peut être explorée en examinant leur association avec la variable de données manquantes [6], [7]. Il est possible d'utiliser des approches graphiques qui consistent à tracer l'évolution moyenne de la variable d'intérêt par groupe, en regroupant les patients ayant un profil de données manquantes similaire [8]. De telles représentations ont été utilisées notamment dans le cas de sorties d'étude où les patients étaient regroupés en fonction de la dernière visite à laquelle ils avaient complété leur questionnaire de QdV [6], une évolution différente des scores moyens de QdV observés dans chacun des groupes suggérant que le mécanisme de données manquantes dépend des données de score observées et donc invalidant l'hypothèse de données manquantes MCAR.

Il est également possible de tester formellement l'hypothèse MCAR contre MAR. Little [9] a proposé un test basé sur l'idée que, sous l'hypothèse MCAR, il y a égalité des moyennes des données observées de la variable d'intérêt entre les groupes définis par le profil de données manquantes. Pour les données manquantes de structure monotone, Listing et Schlittgen ont proposé un test paramétrique [10] et un test non paramétrique [11] basés sur la comparaison des valeurs de la variable d'intérêt à la dernière observation avant sortie d'étude avec celles des patients qui ne sont pas sortis d'étude au même temps d'observation.

Les approches décrites précédemment peuvent également être utilisées afin d'étudier la validité de l'hypothèse MCAR contre MAR. Il est possible d'étudier la corrélation entre la variable de données manquantes et une variable des données observées telle que, par exemple, le dernier score de QdV avant sortie d'étude ou le score de QdV à l'inclusion. Pour étudier la validité de l'hypothèse MCAR contre MAR conditionnellement à des variables explicatives, il est possible d'utiliser les approches de régression logistique ou de modèle de Cox ajustés sur les caractéristiques initiales des patients.

### Méthodes d'analyse

Lorsque les données manquantes sont MAR, les méthodes utilisant toute l'information observée comme les modèles linéaires mixtes (basés sur la vraisemblance ou la vraisemblance restreinte) peuvent être appliquées (voir chapitre « Critères de qualité de vie relative à la santé »). En revanche, ignorer une partie des données observées de la variable d'intérêt ou omettre des variables expliquant le mécanisme des données manquantes conduirait à des estimations biaisées.

## 2.3 Données manquantes de type MNAR

### Définition

Les données manquantes de la variable d'intérêt sont dites MNAR lorsque, conditionnellement aux variables explicatives, la probabilité qu'elles soient manquantes dépend des autres valeurs observées mais aussi non observées de la variable. C'est le cas par exemple lorsque des patients ne remplissent pas leur questionnaire à cause d'une détérioration de leur état de santé, comme l'augmentation de symptômes liés au traitement. Souvent, les données manquantes de type MNAR sont dites *informatives*.

### Etudier la validité de l'hypothèse MAR versus MNAR

Il est impossible de tester formellement l'hypothèse MAR contre MNAR sur la base des données observées puisque, par définition, des données manquantes MNAR dépendent des données non observées de la variable d'intérêt. Il est en revanche possible de mettre en avant des arguments suggérant que l'hypothèse MNAR est la plus plausible, en particulier lorsque des variables fortement associées à la variable d'intérêt, telles que la présence d'effets secondaires du traitement ou la progression de la maladie, sont également associées à la variable de données manquantes.

### Méthodes d'analyse

Sous l'hypothèse MNAR, *i.e.* dans le cas de données manquantes informatives, la modélisation seule de la variable d'intérêt ne suffit pas, il faut modéliser simultanément la variable d'intérêt et le mécanisme à l'origine de ses données manquantes. Des modèles spécifiques, basés sur la distribution conjointe des deux variables, ont ainsi été développés et peuvent être classés en trois catégories : les modèles de sélection, les modèles de mélange de profils et les modèles conjoints. Ces modèles sont généralement utilisés dans la littérature pour prendre en compte les données manquantes informatives de structure monotone (sorties d'étude) en supposant que les données manquantes intermittentes sont de type MCAR ou MAR. Dans le paragraphe suivant, où ces modèles sont décrits en détail, nous nous placerons également dans ce cas. Notons que les données manquantes de type MCAR et MAR sont souvent regroupées sous le terme de données manquantes « ignorables » par opposition aux données manquantes informatives. Cela ne signifie pas que l'analyse peut être réalisée en ignorant les données manquantes, comme nous l'avons vu, mais qu'il est possible d'obtenir des estimations non biaisées sans avoir besoin de recourir à une modélisation du mécanisme à l'origine des données manquantes.

## 3. Modèles en présence de données manquantes informatives

Dans ce paragraphe, nous détaillons les trois approches qui peuvent être utilisées lorsque l'on suspecte que le mécanisme à l'origine des données manquantes de la variable d'intérêt  $Y$  est informatif, c'est-à-dire associé aux valeurs non observées de  $Y$ . En supposant que les données manquantes informatives sont de structure monotone, notons  $D$  la variable de sortie d'étude associée. Les trois approches décrites ci-après sont basées sur la distribution conjointe de la variable d'intérêt  $Y$  et de la variable de sortie d'étude  $D$ , conditionnellement aux variables explicatives observées à l'inclusion  $X$  [12].

### 3.1 Modèle de sélection (*Selection model*)

Supposons que la variable  $D$  prend la valeur  $k$  lorsque la dernière observation avant sortie d'étude de la variable  $Y$  a eu lieu à la  $k^{\text{ème}}$  évaluation et  $K$ , le nombre total d'évaluations, lorsqu'il n'y a pas de sortie d'étude. Les modèles de sélection sont basés sur la décomposition suivante de la distribution conjointe  $f(Y, D | X)$  :

$$f(Y, D | X) = f(Y|X)f(D|Y, X)$$

où  $f(Y|X)$  est la distribution marginale de la variable d'intérêt et  $f(D|Y, X)$  est la distribution conditionnelle de la variable de sortie d'étude sachant les données observées et non observées de  $Y$ .

La modélisation généralement adoptée pour  $f(D|Y, X)$  est une régression logistique proposée par Diggle et Kenward [13] :

$$\text{logit} \left[ P \left( D_i = j \mid D_i \geq j, Y_i(t_j), Y_i(t_{j+1}) \right) \right] = \psi_0 + \psi_1 Y_i(t_j) + \psi_2 Y_i(t_{j+1})$$

où la probabilité que le patient  $i$  soit sorti d'étude à la visite  $j + 1$  sachant qu'il était toujours présent dans l'étude à la visite précédente  $j$  dépend de sa QdV (observée) à la visite précédente  $Y_i(t_j)$  et de sa QdV (non observée) à la visite de sortie d'étude  $Y_i(t_{j+1})$ . Ce modèle est combiné avec le modèle pour la distribution marginale de la variable d'intérêt  $f(Y|X)$ , habituellement un modèle linéaire mixte, et les paramètres associés aux deux sous-modèles sont simultanément estimés. Notons que les paramètres de la régression logistique nous renseignent directement sur le type de mécanisme à l'origine des données manquantes : tester  $\psi_2 = 0$  revient à tester l'hypothèse MAR contre MNAR et tester  $\psi_1 = 0$  revient à tester l'hypothèse MCAR contre MAR. Néanmoins, la validité du modèle repose sur l'hypothèse de normalité des données complètes (observées et non observées) de  $Y$  qui est non vérifiable sur la base des données disponibles. Or, il a été montré que les résultats des modèles de sélection  $y$  sont particulièrement sensibles [14].

### 3.2 Modèle de mélange de profils (*Pattern mixture model*)

Supposons que la variable de sortie d'étude  $D$  correspond à un profil (*pattern*) de données manquantes, c'est-à-dire qu'il regroupe les patients ayant des profils similaires de sortie d'étude. Par exemple, au cours d'un essai où la QdV est évaluée à l'inclusion, puis 3 fois durant le traitement, puis 2 fois durant le suivi post-traitement, la variable  $D$  peut prendre 3 modalités selon que les patients sont sortis d'étude durant le traitement, après le traitement ou ne sont pas sortis d'étude ;  $D$  pourrait aussi prendre autant de modalités que de possibilités de sortie d'étude ( $3+2+1=6$ ). Les modèles de mélange de profils sont basés sur la décomposition suivante de la distribution conjointe  $f(Y, D | X)$  :

$$f(Y, D | X) = f(Y|D, X)f(D|X)$$

où  $f(Y|D, X)$  est la distribution conditionnelle de la variable d'intérêt sachant le profil et  $f(D|X)$  est la distribution marginale de la variable d'appartenance au profil. Cette dernière est supposée multinomiale, c'est-à-dire que la probabilité d'appartenance à un profil est estimée par sa fréquence. La variable d'intérêt est ici modélisée conditionnellement aux profils de sortie d'étude. Il s'agit donc d'appliquer le même modèle qui serait utilisé pour analyser  $Y$  sous l'hypothèse de données manquantes MAR, tel un modèle linéaire mixte, mais dans chacun des sous-échantillons de patients définis par leur profil de données manquantes. Néanmoins, on peut vouloir des estimations relatives à la distribution marginale et non conditionnelle de  $Y$  ; avec le modèle de mélange de profils contrairement au modèle de sélection nous n'y avons pas directement accès. Dans notre exemple illustratif, les estimations conditionnelles permettent de savoir si les trajectoires de QdV sont différentes entre bras de traitement chez les patients sortis d'étude durant le traitement, sortis d'étude après traitement et non sortis d'étude. Cependant, on souhaite généralement comparer l'évolution de la QdV entre bras de traitement de façon globale. Une estimation relative à la distribution marginale de  $Y$ ,  $\hat{\beta}$ , peut être obtenue en calculant la somme des estimations conditionnelles au profil  $l$ ,  $\hat{\beta}^l$ , pondérée par la probabilité d'appartenance au profil,  $\pi_l$  :

$$\hat{\beta} = \sum_{l=1}^L \hat{\beta}^l \pi_l$$

où  $L$  est le nombre de profils. Remarquons qu'en appliquant une telle formule on réalise implicitement une hypothèse sur l'évolution de la variable  $Y$  pour chacun des profils : la QdV des patients sortis d'étude durant le traitement est extrapolée jusqu'à la fin du traitement et après le traitement et la QdV des patients sortis d'étude après le traitement est extrapolée jusqu'à la fin de la période de suivi. Tout comme l'hypothèse de normalité des modèles de sélection, cette hypothèse est non vérifiable sur la base des données disponibles. Elle a cependant l'avantage d'être simple et explicite.

### 3.3 Modèle conjoint

Dans les modèles conjoints, la variable  $D$  correspond au délai jusqu'à la sortie d'étude. Le principe des modèles conjoints, aussi appelés modèles à paramètres partagés, est que les distributions de  $Y$  et  $D$  ont un certain nombre de paramètres communs, dont les effets aléatoires  $b$ , et que conditionnellement aux effets aléatoires les variables  $Y$  et  $D$  sont supposées indépendantes [15]–[20]. La distribution conjointe s'écrit alors :

$$f(Y, D | X) = \int f(Y | b, X) f(D | b, X) f(b | X) db$$

où  $f(b | X)$  est la distribution des effets aléatoires  $b$  (en général supposée gaussienne).

Les modèles conjoints sont constitués de deux sous-modèles : un sous-modèle pour la variable d'intérêt  $Y$  qui correspond au même modèle qui serait utilisé sous l'hypothèse de données manquantes MAR et un sous-modèle de survie (en général un modèle à risques proportionnels de Cox) pour le délai jusqu'à sortie d'étude. L'association entre les deux variables est prise en compte dans le sous-modèle de survie en incluant une fonction des paramètres du modèle de  $Y$ , correspondant le plus souvent au prédicteur linéaire, en tant que variable dépendant du temps dans le modèle de survie. La fonction du risque instantané de sortie d'étude s'écrit alors :

$$\lambda_i(t | W) = \lambda_0(t) \exp\{\gamma^t W + \alpha Y_i^*(t)\}$$

où  $\lambda_0$  est la fonction de risque de base,  $\gamma$  le vecteur des effets des variables explicatives  $W$  et  $\alpha$  le paramètre d'association entre le risque de sortie d'étude et la vraie valeur de  $Y$  au temps  $t$  (débarassée de l'erreur résiduelle),  $Y^*(t)$ . Notons qu'un paramètre  $\alpha$  qui ne serait pas significativement non nul irait en faveur d'une hypothèse MAR contre MNAR. Lorsque la variable d'intérêt est un score de QdV, une association négative (respectivement, positive) signifie qu'une augmentation de  $x$  points de la valeur du score diminue (respectivement, augmente) le risque de sortie d'étude de  $HR = \exp(x\alpha)$ .

Il est également possible de considérer d'autres structures d'association : au lieu de la valeur courante  $Y^*$ , on peut par exemple supposer que c'est la pente de  $Y^*$  ou les effets aléatoires seuls qui sont associés au risque de sortie d'étude.

#### 4. Conclusion

Malgré des efforts croissants pour limiter leur occurrence et en récolter la cause, les données manquantes d'une variable d'intérêt mesurée tout au long d'un essai clinique représentent un véritable enjeu au moment de l'analyse statistique. En présence de données manquantes, il est indispensable de se questionner sur le type de mécanisme qui est à leur origine car les stratégies d'analyse en dépendent. Dans le cas de données manquantes supposées non informatives, les méthodes basées sur la vraisemblance donnent des résultats valides à condition d'utiliser toutes les données observées de la variable d'intérêt et d'avoir identifié les variables en lien avec le mécanisme de données manquantes. En présence de données manquantes informatives, il est nécessaire de modéliser conjointement à la variable d'intérêt le mécanisme de données manquantes. Les modèles de sélection et les modèles de mélange de profils ont d'abord été proposés. Ils reposent cependant sur des hypothèses fortes et non vérifiables sur la base des données disponibles. Par ailleurs, ils ne sont pas implémentés dans les logiciels statistiques standards contrairement aux modèles conjoints. Ces derniers connaissent un intérêt grandissant depuis quelques décennies, et qui ne se limite pas au contexte dans lequel nous les avons décrits. Ils reposent évidemment sur des hypothèses de modélisation relatives au mécanisme de données manquantes mais de nombreuses extensions ont déjà été proposées permettant d'en considérer différentes formes. Notons que l'utilisation d'un des 3 types de modèles peut donner des résultats en faveur ou en défaveur de l'hypothèse de données manquantes informatives. Dans le deuxième cas, cela ne veut pas nécessairement dire qu'elles ne sont pas informatives mais que le modèle utilisé n'est peut-être pas correctement spécifié. De façon générale, pour renforcer la fiabilité des résultats et des conclusions qui en découlent, il est recommandé de tester différents modèles en faisant varier les hypothèses de modélisation.



### Modèles conjoints dans les logiciels statistiques

Les modèles conjoints sont implémentés dans les logiciels statistiques standards tel que R, SAS et Stata. Dans ce paragraphe, nous détaillons les principaux packages et modules qui peuvent être utilisés.

- R

La fonction `jointModel()` du package **JM** [18], [21] permet l'utilisation d'un modèle conjoint. Les deux principaux arguments sont : un objet issu d'un modèle linéaire mixte (obtenu à partir de la fonction `lme()` du package **nlme**) et un objet issu d'un modèle de survie (obtenu à partir de la fonction `coxph()` ou `survreg()` du package **survival**). De plus, **JM** propose de choisir parmi différentes structures d'association (valeur courante, pente, ...), différentes formes pour le risque de base du modèle de survie (continu par morceaux, Weibull, B-splines, ...) ainsi que différents arguments d'optimisation (nombre de points et type de quadrature pour l'approximation des intégrales, ...).

Il est également possible d'utiliser le package **JMbayes** [22] qui est comparable à **JM** mais est fondé sur une approche bayésienne (basée sur les algorithmes MCMC). Ce package est le plus récent des deux et est celui qui est toujours en cours de maintenance et de développement. Il offre la possibilité d'utiliser des modèles plus flexibles, notamment en considérant une ou plusieurs variables longitudinales, quantitatives ou qualitatives, via un modèle linéaire mixte généralisé.

- SAS

La macro `%JM()` [23] a été développée sous SAS dans la continuité du package **JM** sous R. L'argument `LongiType` permet de considérer des modèles paramétriques ou bien des modèles plus flexibles (splines naturelles, B-splines) pour la variable longitudinale qui peut être de différents types (gaussienne, binaire, binomiale ou de Poisson) via un modèle linéaire mixte généralisé. L'argument `EventModel` permet un large choix pour la fonction de risque de base du modèle de survie : distributions paramétriques (exponentielle, Weibull, Gompertz, ...) mais aussi des formes plus flexibles (splines naturelles ou de Royston-Parmar). De plus, cette macro offre différentes options pour la structure d'association avec l'argument `SharedParam`.

- Stata

La commande **merlin** [24] est le résultat de nombreux développements et peut à présent être utilisée avec la version 16 du logiciel Stata. Une variété de modèles peut être considérée en utilisant différentes distributions pour la variable longitudinale quantitative ou qualitative (gaussienne, binaire, binomiale ou de Poisson) et pour la fonction du risque de base du modèle de survie (Weibull, Gompertz, splines) avec l'option `family()`. Des options permettent de choisir également parmi différentes structures d'association (valeur courante, dérivées d'ordre 1 ou 2, ...).

## Références

- [1] N. R. Council, *The prevention and treatment of missing data in clinical trials*. Washington, DC: The National Academies Press, 2010.
- [2] R. J. Little *et al.*, 'The prevention and treatment of missing data in clinical trials', *New England Journal of Medicine*, vol. 367, no. 14, pp. 1355–1360, 2012.
- [3] S. Halabi and S. Michiels, *Textbook of clinical trials in oncology: A statistical perspective*. CRC Press, 2019.
- [4] D. B. Rubin, 'Inference and Missing Data', *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976, doi: 10.2307/2335739.
- [5] Roderick J. A. Little, 'Modeling the Drop-Out Mechanism in Repeated-Measures Studies', *Journal of the American Statistical Association*, vol. 90, no. 431, pp. 1112–1121, 1995, doi: 10.2307/2291350.
- [6] D. L. Fairclough, *Design and analysis of quality of life studies in clinical trials*. Chapman and Hall/CRC, 2010.
- [7] D. Curran, M. Bacchi, S. Schmitz, G. Molenberghs, and R. Sylvester, 'Identifying the types of missingness in quality of life data from clinical trials', *Statistics in medicine*, vol. 17, no. 5–7, pp. 739–756, 1998.
- [8] J. W. Hogan, J. Roy, and C. Korkontzelou, 'Handling drop-out in longitudinal studies', *Statistics in medicine*, vol. 23, no. 9, pp. 1455–1497, 2004.
- [9] R. J. A. Little, 'A Test of Missing Completely at Random for Multivariate Data with Missing Values', *Journal of the American Statistical Association*, vol. 83, no. 404, pp. 1198–1202, Dec. 1988, doi: 10.1080/01621459.1988.10478722.
- [10] J. Listing and R. Schlittgen, 'Tests if dropouts are missed at random', *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, vol. 40, no. 8, pp. 929–935, 1998.
- [11] J. Listing and R. Schlittgen, 'A nonparametric test for random dropouts', *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, vol. 45, no. 1, pp. 113–127, 2003.
- [12] J. G. Ibrahim and G. Molenberghs, 'Missing data methods in longitudinal studies: a review', *Test (Madr)*, vol. 18, no. 1, pp. 1–43, May 2009, doi: 10.1007/s11749-009-0138-x.
- [13] P. Diggle and M. G. Kenward, 'Informative Drop-Out in Longitudinal Data Analysis', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 43, no. 1, pp. 49–93, 1994, doi: 10.2307/2986113.
- [14] M. G. Kenward, 'Selection models for repeated measurements with non-random dropout: an illustration of sensitivity', *Statistics in medicine*, vol. 17, no. 23, pp. 2723–2732, 1998.
- [15] M. C. Wu and R. J. Carroll, 'Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process', *Biometrics*, pp. 175–188, 1988.
- [16] M. C. Wu and K. Bailey, 'Analysing changes in the presence of informative right censoring caused by death and withdrawal', *Statistics in Medicine*, vol. 7, no. 1–2, pp. 337–346, 1988.
- [17] D. Follmann and M. Wu, 'An approximate generalized linear model with random effects for informative missing data', *Biometrics*, pp. 151–168, 1995.
- [18] D. Rizopoulos, *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press, 2012.
- [19] R. Henderson, 'Joint modelling of longitudinal measurements and event time data', *Biostatistics*, vol. 1, no. 4, pp. 465–480, Dec. 2000, doi: 10.1093/biostatistics/1.4.465.
- [20] 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*, vol. 28, no. 16, pp. 2796–2801, Jun. 2010, doi: 10.1200/JCO.2009.25.0654.
- [21] D. Rizopoulos, 'JM: An R package for the joint modelling of longitudinal and time-to-event data', *Journal of Statistical Software, Articles*, vol. 35, no. 9, pp. 1–33, 2010, doi: 10.18637/jss.v035.i09.
- [22] D. Rizopoulos, 'The R Package JMBayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC', *Journal of Statistical Software; Vol 1, Issue 7 (2016)*, Aug. 2016, [Online]. Available: <https://www.jstatsoft.org/v072/i07>.
- [23] A. Garcia-Hernandez and D. Rizopoulos, '%JM: A SAS macro to fit jointly generalized mixed models for longitudinal data and time-to-event responses', *Journal of Statistical Software, Articles*, vol. 84, no. 12, pp. 1–29, 2018, doi: 10.18637/jss.v084.i12.
- [24] M. J. Crowther, 'merlin—A unified modeling framework for data analysis and methods development in Stata', *The Stata Journal*, vol. 20, no. 4, pp. 763–784, 2020.

### 2.3 Discussion et conclusion

La présence de données manquantes est un véritable enjeu pour les analyses dans les essais cliniques : un comité a même été créé à la demande de la FDA pour établir un rapport sur les moyens à engager pour réduire les données manquantes ainsi que les méthodes statistiques à privilégier pour en tenir compte dans l'analyse ([Council, 2010](#)).

Dans ce chapitre, nous avons introduit les différents concepts utilisés pour traiter les données manquantes, en utilisant l'exemple des données de QdV dans les essais cliniques. Différentes stratégies et méthodes d'analyse basées sur la vraisemblance (comme les modèles linéaires mixtes) ont été présentées lorsque le mécanisme de données manquantes est supposé MCAR ou MAR (non informatif). Lorsque le mécanisme de données manquantes est supposé MNAR (informatif), trois approches ont été proposées : le modèle de sélection, le modèle de mélange de profils et le modèle conjoint à effets partagés. Ces approches peuvent venir en complément des modèles linéaires mixtes, mais sont pourtant très peu utilisées dans le contexte des essais cliniques.

Il est actuellement recommandé de reporter le nombre de questionnaires reçus par rapport au nombre de questionnaires attendus (compliance au questionnaire), de vérifier si la répartition des données manquantes dans les bras de traitement est homogène et d'identifier les variables expliquant une différence observée ([Calvert et al., 2013](#)). Cependant, il n'existe pas de recommandations précises sur les approches à utiliser en cas de suspicion de données manquantes informatives, si ce n'est de bien référencer et justifier la méthode utilisée. Dans le chapitre suivant, nous étudierons les trois approches déjà décrites (modèle de sélection, modèle de mélange de profils et modèle conjoints à effets partagés) sur les données de QdV de l'essai clinique PRODIGE 5/ACCORD 17 en présence de données manquantes suspectées informatives.

## Étude de trois approches pour tenir compte de sorties d'étude informatives

---

### 3.1 Introduction

Les données de QdV de l'essai clinique PRODIGE 5/ACCORD 17 présentent une forte proportion de données manquantes. Les questionnaires de QdV sont souvent incomplets, voire manquants à certaines visites. Ce constat est d'autant plus vrai que le suivi des patients est long. En particulier, les données manquantes monotones (sorties d'étude) représentent une grande majorité des données manquantes et interrompent le suivi initialement prévu de la QdV (par exemple en raison du décès du patient). Cet essai clinique inclut des patients à un stade avancé de leur maladie (cancer de l'œsophage métastatique) et potentiellement dans un état faible de santé ou affaiblis par la prise du traitement. Dans cette situation, il est légitime d'imaginer qu'il existerait potentiellement un lien entre le fait de ne plus répondre au questionnaire de QdV et l'état de santé du patient lui-même (obtenu et mesuré par la QdV). Dans le chapitre précédent, nous avons vu comment tenir compte du mécanisme de données manquantes. Alors qu'il est peu probable

### Chapitre 3. Étude de trois approches pour tenir compte de sorties d'étude informatives

---

que les données manquantes soient MCAR, il est souvent plausible qu'elles soient MAR, mais il n'est jamais certain qu'elles ne soient pas MNAR.

Nous avons aussi vu qu'il existe différentes approches pour tenir compte des sorties d'étude suspectées informatives (sous l'hypothèse MNAR) : d'abord les modèles de sélection (SMs) et les modèles de mélange de profils (PMMs), puis plus récemment, les modèles conjoints à effets partagés (SPMs). Ces trois types de modèle proposent de tenir compte des sorties d'étude de différentes façons en supposant : une dépendance aux données non observées (SM), l'existence d'une hétérogénéité de la population (PMM) ou que les effets aléatoires sont capables de capturer la dépendance entre le processus de données répétées et le processus de survenue des sorties d'étude (SPM). Selon [Fielding et al. \(2016\)](#), il y a un trop grand écart entre le développement des méthodes statistiques pour traiter les données manquantes et leur utilisation dans des applications comme les essais cliniques. En effet, ces modèles reposent sur des hypothèses qu'il est parfois difficile de vérifier et d'appréhender, comme l'hypothèse gaussienne de la distribution complète des données (SM), l'extrapolation au-delà des sorties d'étude (PMM) ou l'hypothèse gaussienne de la distribution des effets aléatoires (SPM), ce qui peut complexifier le choix de la méthode à utiliser. De plus, il est parfois difficile de mettre en pratique ces modèles par manque de solution logicielle, ce qui pourrait aussi expliquer leur sous-utilisation dans les essais cliniques.

Nous proposons de présenter chacune des trois méthodes permettant de tenir compte des sorties d'étude informatives lors de l'analyse de la QdV, en dressant d'abord leurs avantages et leurs inconvénients d'un point de vue théorique. Cela nous permet d'expliquer clairement les hypothèses des modèles et leurs implications. Nous illustrons ensuite ces trois méthodes sur les données de QdV de l'essai clinique PRODIGE 5/ACCORD 17, puis nous les étudions afin de mettre en évidence leurs différences et similitudes entre elles et par rapport au modèle linéaire mixte, qui est le modèle de référence ignorant le caractère informatif des données manquantes (hypothèse MAR). Ce travail a fait l'objet d'un premier article intitulé « *Handling informative dropout in longitudinal analysis of health-related quality of life: Application of three approaches to data from the esophageal cancer clinical trial PRODIGE 5/ACCORD 17* » et publié dans la revue *BMC Medical Research Methodology* ([Cuer et al., 2020](#)).

### 3.2. Article : *Handling informative dropout in longitudinal analysis of health-related quality of life*

---

Au-delà des descriptions faites des modèles dans l'article, nous décrirons ensuite plus en détail les SPMs qui sont des modèles plus récents (et ceux utilisés dans la suite de cette thèse) afin de mieux comprendre comment ils peuvent être utilisés dans le contexte des sorties d'étude informatives. Enfin, ce chapitre se terminera sur une discussion des trois modèles étudiés, en complément de celle faite dans l'article présenté.

### 3.2 Article : *Handling informative dropout in longitudinal analysis of health-related quality of life: application of three approaches to data from the esophageal cancer clinical trial PRODIGE 5/ACCORD*

17

RESEARCH ARTICLE

Open Access

# Handling informative dropout in longitudinal analysis of health-related quality of life: application of three approaches to data from the esophageal cancer clinical trial PRODIGE 5/ACCORD 17



B. Cuer<sup>1,2,3\*</sup> , C. Mollevi<sup>1,2,3</sup>, A. Anota<sup>2,4,5</sup>, E. Charton<sup>2,4,5</sup>, B. Juzyna<sup>6</sup>, T. Conroy<sup>7,8</sup> and C. Touraine<sup>1,2</sup>

## Abstract

**Background:** Health-related quality of life (HRQoL) has become a major endpoint to assess the clinical benefit of new therapeutic strategies in oncology clinical trials. Typically, HRQoL outcomes are analyzed using linear mixed models (LMMs). However, longitudinal analysis of HRQoL in the presence of missing data remains complex and unstandardized. Our objective was to compare the modeling alternatives that account for informative dropout.

**Methods:** We investigated three alternative methods—the selection model (SM), pattern-mixture model (PMM), and shared-parameters model (SPM)—in relation to the LMM. We first compared them on the basis of methodological arguments highlighting their advantages and drawbacks. Then, we applied them to data from a randomized clinical trial that included 267 patients with advanced esophageal cancer for the analysis of four HRQoL dimensions evaluated using the European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 questionnaire.

**Results:** We highlighted differences in terms of outputs, interpretation, and underlying modeling assumptions; this methodological comparison could guide the choice of method according to the context. In the application, none of the four models detected a significant difference between the two treatment arms. The estimated effect of time on HRQoL varied according to the method: for all analyzed dimensions, the PMM estimated an effect that contrasted with those estimated by the SM and SPM; the LMM estimated effects were confirmed by the SM (on two of four HRQoL dimensions) and SPM (on three of four HRQoL dimensions).

**Conclusions:** The PMM, SM, or SPM should be used to confirm or invalidate the results of LMM analysis when informative dropout is suspected. Of these three alternative methods, the SPM appears to be the most interesting from both theoretical and practical viewpoints.

(Continued on next page)

\* Correspondence: [benjamin.cuer@icm.unicancer.fr](mailto:benjamin.cuer@icm.unicancer.fr)

<sup>1</sup>Biometrics Unit, Montpellier Cancer Institute (ICM), University of Montpellier, 208, avenue des Apothicaires, 34298 Montpellier, France

<sup>2</sup>French National Platform Quality of Life and Cancer, Montpellier, France  
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Trial registration:** This study is registered with [ClinicalTrials.gov](https://clinicaltrials.gov), number [NCT00861094](https://clinicaltrials.gov/ct2/show/study/NCT00861094).

**Keywords:** Pattern-mixture model, Selection model, Shared-parameters model, Joint modeling, Health-related quality of life, Informative dropout, Cancer clinical trial

## Background

Health-related quality of life (HRQoL) is often a secondary endpoint in cancer clinical trials. It is also increasingly being used as a primary or co-primary endpoint [1]. HRQoL is assessed at different time points throughout the care process (at baseline, during treatment, and during follow-up) by self-administered questionnaires composed of items assessing different HRQoL dimensions. The HRQoL outcome to be analyzed consists of longitudinal dimension-specific score data. However, the rate of completed questionnaires generally decreases over time and, in addition, some items may be missing among available questionnaires. This leads to missing data that are said to be monotone if the score is not available from a certain time point until the end of the study, and intermittent otherwise. The nature of the missing data mechanism depends on how the missingness is related to the HRQoL outcome.

Missing data are classified as missing completely at random (MCAR) if missingness is independent of the (observed or unobserved) HRQoL outcome or depends only on observed characteristics, as missing at random (MAR) if missingness additionally depends on the observed HRQoL outcome, and as missing not at random (MNAR) if missingness is dependent of the unobserved HRQoL outcome [2, 3]. The terms informative or non-ignorable are also used to refer to MNAR data. In the presence of incomplete longitudinal outcome data, the strategy of analysis should be adapted to the nature of the missing data mechanism in order to avoid biased or inaccurate results. In most studies, the missing data mechanism is not characterized, so methods used to analyze longitudinal HRQoL data in randomized clinical trials [4] are potentially inadequate.

Linear mixed models (LMMs) are powerful and flexible models for the analysis of repeated measures of a continuous outcome. This class of model is classically used to compare changes in HRQoL over time between experimental and control arms in cancer clinical trials [5, 6]. However, the occurrence of intermittent or monotone missing data could compromise the longitudinal analysis of HRQoL data, leading to a loss of statistical power at best, and, at worse, biased estimates; for instance, in palliative or advanced disease situations, where missing data could be related to the health status of patients too ill to complete their HRQoL questionnaires [7, 8]. Likelihood-based methods that use all the observed

information (as in LMMs) are valid when the missing data are MAR [9]. However, in the presence of informative missing data (i.e., MNAR), the two processes that are the longitudinal HRQoL outcome and the missing data mechanism have to be jointly modeled to prevent a biased estimation [10, 11].

Since the end of the 1980s, different models have been proposed for the joint distribution of the longitudinal outcome and the missingness process. More attention has been devoted to monotone missing data, corresponding to dropout, which is more likely to be informative and generally easier to handle. Pattern-mixture models (PMMs) and selection models (SMs) are based on the two possible decompositions of the joint distribution [12, 13]. In recent years, the joint models or shared-parameter models (SPMs), where the association between the two processes is captured by shared parameters, have received much attention [14, 15]. In clinical trials, SPMs are mostly used to jointly analyze a longitudinal outcome and overall survival. They can also be used to take into account and study the relationship between a longitudinal HRQoL outcome and time-to-dropout [16].

There are relatively few publications that compare these three approaches from a perspective of their practical application to clinical trial data [17–19]. This is needed to further our understanding of their use and interpretation; the insufficient knowledge about these models could explain why they are rarely used in clinical trials.

The objectives of this paper were to compare the PMM, the SM, and the SPM with each other and then to compare these models with the LMM, for the analysis of an HRQoL outcome in the presence of informative dropout. First, we compare the models from a methodological point of view, highlighting the advantages and drawbacks of each one. Then, we illustrate and interrogate them in the longitudinal analysis of four HRQoL dimensions in patients with advanced esophageal cancer from the PRODIGE 5/ACCORD 17 clinical trial.

## Methods

We highlighted the differences between the PMM, SM, and SPM in handling informative dropout when analyzing a longitudinal HRQoL outcome and interpreted their results in relation to those from the LMM. For this purpose, we first made a methodological comparison of the four models by highlighted their differences in terms of



underlying modelling assumptions and interpretation. The advantages and drawbacks of each of model are then illustrated through an analysis of data from the PRODIGE 5/ACCORD 17 clinical trial (NCT00861094).

### Illustrative clinical trial

#### Study design

In the PRODIGE 5/ACCORD 17 clinical trial, 267 patients with advanced esophageal cancer were randomly assigned to either an experimental arm ( $N = 134$ ) receiving a FOLFOX (fluorouracil plus leucovorin and oxaliplatin) regimen or a control arm ( $N = 133$ ) receiving a fluorouracil and cisplatin regimen as part of chemoradiotherapy treatment. The primary endpoint was progression-free survival and one of the secondary endpoints was HRQoL. The statistical analysis of the primary endpoint revealed no significant difference between the two treatment arms. More details concerning inclusion and exclusion criteria, study design, protocol treatment, HRQoL assessment, and compliance have been previously published [20, 21].

#### HRQoL assessment

HRQoL was prospectively assessed using the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30, version 3.0) [22] at baseline, during treatment (months 1.25 and 3), at month 4, and after treatment during follow-up (at months 6, 12, 24, and 36). This self-administered questionnaire contains 30 items evaluating five functional scales, nine symptomatic scales/items, and one global health status/HRQoL scale. Standardized scores from 0 to 100 can be calculated for each scale according to the scoring procedure recommended by the EORTC [23]. A high score for the functional and global health status scales corresponds to good functional capacities and reflects a high level of HRQoL, whereas a high score for the symptom scales corresponds to a high level of symptoms and reflects a poor HRQoL. Four dimensions were pre-specified in the protocol as targeted dimensions: global health status/HRQoL (QL scale), physical functioning (PF scale), pain (PA scale), and fatigue (FA scale). In what follows, we will consider only these four dimensions (or scales).

#### Statistical analysis

All analyses were performed in the evaluable intent-to-treat population: a patient was considered as evaluable for a given scale when the score was available at least once during the study, whatever the corresponding measurement time. We used the four models described below in Eqs. (1), (3), (5) and (8) to analyze the longitudinal HRQoL score data conditionally to baseline covariates in the presence of potentially informative monotone missing data (dropout).

We first used the LMM that is valid under the MAR assumption. We then modeled the joint distribution of the longitudinal outcome and the dropout process using three models that are valid under the MNAR assumption: the SM and the PMM, which are based on the two existing and converse factorizations of the joint distribution, and the SPM, where the longitudinal outcome and the time-to-dropout are linked through a function of the random effects. In these three models, we used the LMM presented below as the sub-model for the HRQoL score.

#### Linear mixed model (LMM)

We modeled the HRQoL score trajectories by a random coefficients LMM. The HRQoL score for patient  $i$  at time  $t_j$  of the  $j$ -th planned visit was expressed as follows:

$$Y_i(t_j) = \beta_0 + \beta_1 t_j + \beta_2 \{arm_i \times t_j\} + b_{0i} + b_{1i} t_j + \varepsilon_i(t_j) \quad (1)$$

where  $arm_i$  is the arm indicator variable for patient  $i$  (0: control, 1: experimental),  $\beta_0$  is the intercept,  $\beta_1$  the slope in the control arm, and  $\beta_2$  the interaction effect corresponding to the difference between the slopes in the experimental and control arms. With this parametrization, the quantity  $\beta_1 + \beta_2$  represents the slope in the experimental arm. The random intercept  $b_{0i}$  and the random slope  $b_{1i}$  take into account the repeated measurements on the same patient and correspond to the individual deviations from the fixed intercept and slope, respectively. They are assumed to be normally distributed with a mean of 0 and a  $2 \times 2$  unconstrained covariance matrix to estimate. The error term denoted by  $\varepsilon_i(t_j)$  is also assumed to be normally distributed with a mean of 0 and a variance to estimate.

In what follows,  $Y_i$ ,  $X_i$ , and  $D_i$  denote respectively the vector of longitudinal HRQoL scores, the vector of covariates, and the dropout variable for patient  $i$ .

#### Selection model (SM)

The SM is based on the decomposition of the joint distribution into the marginal distribution of the HRQoL score and the conditional distribution of the dropout variable given the HRQoL score:

$$f(Y_i, D_i | X_i) = f(Y_i | X_i) \times f(D_i | Y_i, X_i) \quad (2)$$

where the dropout variable  $D_i$  corresponds to the visit at which the last available HRQoL assessment took place, i.e., before patient  $i$  dropout. In cases of no dropout,  $D_i = J$ , where  $J$  is the number of planned visits. We modeled the HRQoL score using the LMM in Eq. (1). We modeled the conditional probability of dropout at each visit  $j = 1, \dots, J$  by the logistic regression proposed by Diggle and Kenward [24]:

$$\begin{aligned} \text{logit}[P(D_i = j | D_i \geq j, Y_i(t_j), Y_i(t_{j+1}))] \\ = \psi_0 + \psi_1 Y_i(t_j) + \psi_2 Y_i(t_{j+1}) \end{aligned} \quad (3)$$

The dropout probability is allowed to depend on the last (observed) HRQoL score  $Y_i(t_j)$  and the current (unobserved) HRQoL score  $Y_i(t_{j+1})$ . A non-zero parameter  $\psi_1$  would be in favor of the MAR assumption and a non-zero parameter  $\psi_2$  in favor of the MNAR assumption (informative dropout). If only the  $\psi_0$  parameter is non-zero, the dropout can be considered to be independent of the HRQoL score (MCAR assumption).

#### Pattern-mixture model (PMM)

The PMM is based on the other possible decomposition of the joint distribution, that is, the decomposition into the marginal distribution of the dropout variable and the conditional distribution of the HRQoL score given the dropout variable:

$$f(Y_i, D_i | X_i) = f(D_i | X_i) \times f(Y_i | D_i, X_i) \quad (4)$$

where the dropout variable corresponds to the pattern of missing data:  $D_i = k$ ,  $k = 1, \dots, K$ , where  $K$  is the number of possible patterns. In the simplest case, the variable is defined as a dropout indicator ( $K=2$ ); in the most complex case, the variable is defined as the number of dropout possibilities:  $D_i = k$ ,  $k = 1, \dots, J$ , where  $J$  is the number of planned visits. In our application, we classified a patient as belonging to a certain pattern when she/he dropped out within a specific time interval covering one or several visits.

In the PMM, a multinomial distribution is assumed for the dropout probability, meaning that the probability of belonging to pattern  $k$  is simply estimated by the proportion  $\pi_k$  of patients belonging to pattern  $k$ .

We modeled the conditional HRQoL score trajectory using an LMM similar to the LMM in Eq. (1) in each pattern  $k$ :

$$\begin{aligned} Y_{ik}(t_j) = \beta_0^k + \beta_1^k t_j + \beta_2^k \{arm_{ik} \times t_j\} + b_{0ik} \\ + b_{1ik} t_j + \varepsilon_{ik}(t_j) \end{aligned} \quad (5)$$

Note that in the PMM approach, the fixed effects differ according to the dropout pattern. The following formula allows estimates to be obtained for the marginal distribution of the HRQoL score (irrespective of the pattern):

$$\beta_l = \sum_{k=1}^K \beta_l^k \pi_k, \quad l = 0, 1, 2 \quad (6)$$

It corresponds to a weighted sum of the pattern-specific parameters. Confidence intervals can then be calculated using the delta method [25].

#### Shared-parameter model (SPM)

The SPM captures the association between the time-to-dropout and the longitudinal HRQoL outcome through shared parameters that include the random effects  $b_i$ , so that the HRQoL score and the dropout variable are supposed to be conditionally independent given the random effects:

$$\begin{aligned} f(Y_i, D_i | X_i) &= \int f(Y_i, D_i | X_i) db_i \\ &= \int f(Y_i | b_i, X_i) \times f(D_i | b_i, X_i) \times f(b_i) db_i \end{aligned} \quad (7)$$

where the dropout variable  $D_i$  corresponds to a time-to-dropout variable. In our application, dropout is not related to an event occurring at any time but corresponds to non-response after a certain visit. Thus, we defined  $D_i$  as the delay between inclusion and the last visit in which HRQoL assessment occurred. We modeled the HRQoL score using the LMM in Eq. (1). We modeled the risk of dropout at time  $t_j$  using a Cox-type survival model.

In the SPM, the association between the HRQoL score and dropout is modeled by including a function of the variables and parameters from the model for  $Y_i$  as a time-dependent variable in the survival model. We used the current value parametrization, which means that the time-dependent variable corresponded to the true current HRQoL score value:  $Y_i^*(t_j) = Y_i(t_j) - \varepsilon_i(t_j)$ . More precisely, we used the following model for  $D_i$ :

$$\lambda_i(t_j | arm_i) = \lambda_0(t) \exp\{\gamma arm_i + \alpha Y_i^*(t_j)\} \quad (8)$$

where  $\lambda_0$  is the baseline hazard function,  $\gamma$  denotes the arm effect on the instantaneous risk of dropout, and  $\alpha$  is the parameter that quantifies the association between risk of dropout and true current HRQoL score.

#### Statistical software

We fitted the four models to the PRODIGE 5/ACCORD 17 data using the R software (code available on request). For LMM estimation, we used the restricted maximum likelihood method (REML) from the R package nlme [26]. The SM was not available in standard statistical software and required sophisticated programming: the Diggle and Kenward model involved marginalization over the unobserved outcomes and the computation of the likelihood required evaluation of integrals approximated by the Romberg numerical algorithm. We implemented a maximum likelihood function procedure based on a Newton-type algorithm. To apply the PMM required that we apply an LMM with indicator variables for the pattern. We then combined the PMM estimates following Eq. (6) to obtain marginal estimates and implemented a delta method to obtain their confidence

intervals. For the SPM, we used the R package JM [27] by assuming a piecewise-constant function for the baseline hazard  $\lambda_0$  with seven intervals for the baseline (six internal knots placed at months 1.25, 3, 4, 6, 12, and 24) and the pseudo-adaptive Gauss-Hermite method with nine quadrature points to approximate the integrals over the random effects.

## Results

### Methodological comparison

Table 1 compares the four approaches (LMM, SM, PMM, and SPM) from a methodological point of view.

In cases of non-informative dropout (MAR assumption), the likelihood-based LMM that uses all observed data provides valid results; in cases of informative dropout (MNAR assumption), the risk of dropout needs to be modeled using one of the three other approaches.

The SM explains the probability of dropout by a logistic regression; the PMM estimates the probability of belonging to a certain pattern of dropout with a multinomial distribution; the SPM uses a survival model for the time-to-dropout. The SM and PMM suppose that dropout occurs at the discrete assessment times of the HRQoL. By contrast, the SPM treats the time variable as continuous, making it possible to take into account the fact that the dropout could arise at any time during the study.

The fixed parameters  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$  characterizing the mean HRQoL score trajectories are directly estimated using the LMM, SM, and SPM, or obtained indirectly by extrapolation using the PMM. More precisely, the PMM estimates the HRQoL score trajectory parameters at the level of each pattern  $k$ ; afterwards, marginal estimates can be calculated as weighted averages using the proportion  $\pi_k$  of patients in each dropout pattern. Note that this calculation implicitly extrapolates the HRQoL score trajectories beyond the dropout. Thus, all models can be used to graphically represent the mean HRQoL score over time according to treatment arm, directly (LMM, SM, SPM) or indirectly (PMM). The PMM provides complementary graphs specific to the dropout pattern, which can be useful to understand and visualize how the risk of dropout is linked to the HRQoL. The SPM allows a graphical representation of the risk of dropout over time. The informative nature of the dropout can also be tested using additional parameters of the SM or SPM: the  $\psi_2$  coefficient in the logistic regression of the SM indicates how the probability of the HRQoL score to be missing at a certain time depends on the missing value at this time, while the  $\alpha$  coefficient in the Cox regression of the SPM indicates how the instantaneous risk of dropout at any time is associated with the current HRQoL score.

Nevertheless, the models used to study the evolution of HRQoL scores in the presence of informative dropout require additional assumptions that are untestable on the basis of the observed data. We have already mentioned extrapolating the HRQoL trajectories beyond the dropout in the PMM. The SM is based on the assumption of a normal distribution of the complete (i.e., observed and unobserved) HRQoL score variable. The SPM assumes independence between the longitudinal outcome and dropout process conditionally to the random effects.

The estimates of each model can be obtained using usual statistical software (including R, SAS, and Stata). Specific software has already been developed for LMM and SPM. However, applying the SM and the PMM requires a programming effort. In particular, applying the SM requires implementation and maximization of the likelihood function.

### Application on data from the PRODIGE 5/ACCORD 17 clinical trial

#### Monotone missing data in HRQoL outcomes

At each scheduled visit, there were missing HRQoL score data. From the 267 patients of the intent-to-treat population (experimental arm:  $N = 134$ ; control arm:  $N = 133$ ), the remaining evaluable patients, i.e., with at least one available HRQoL score, were 252 for scale QL (experimental arm:  $N = 130$ ; control arm:  $N = 122$ ), and 254 for scales PF, PA and FA (experimental arm:  $N = 131$ ; control arm:  $N = 123$ ). In fact, the proportion of available scores for scales QL, PF, PA, and FA decreased over time, mostly because of monotone missing data that can be attributed to dropouts (see Fig. 1). For example, for the QL scale, 16/130 patients (12%) in the experimental arm and 17/122 patients (14%) in the standard arm dropped out after the baseline visit (V0, baseline); at the last scheduled visit (V7, month 36), 125/130 patients (96%) in the experimental arm and 115/122 patients (94%) in the standard arm had dropped out (i.e., only 5/130 (4%) and 7/122 (6%) patients completed the questionnaire or the items associated with the QL scale until V7). The distribution of the dropouts seemed homogeneous in both treatment arms, regardless of the dimension. The compliance in completing the entire questionnaire was high at baseline (89 and 90% in experimental and standard regimen arms, respectively), then reduced during treatment and follow-up. Some missing items led to a lower compliance for dimension QL than for the others (for example, at baseline: 83% for QL vs. 89% for PF and 88% for PA and FA in the experimental regimen arm, and 86% for QL vs. 90% for PF, PA and FA in the standard regimen arm) (see Supplementary Figure 1).

**Table 1** Methodological comparison of the four models used for analysis of longitudinal HRQoL score data

	LMM	SM	PMM	SPM
<b>MODELING</b>				
<b>Validity of the model</b>	Under MAR assumption	Under MNAR assumption	Under MNAR assumption	Under MNAR assumption
<b>Model for the HRQoL outcome Y</b>	LMM	LMM	LMM by pattern	LMM
<b>Model for the dropout variable</b>	–	Logistic Dropout at specific time (discrete)	Multinomial Dropout at specific time (discrete)	Survival model Dropout at any time (continuous)
<b>Graphical outputs</b>	Mean HRQoL score over time according to treatment arm	Mean HRQoL score over time according to treatment arm	(Mean HRQoL score over time according to treatment arm) Mean HRQoL score over time according to treatment arm for each dropout pattern	Mean HRQoL score over time according to treatment arm Hazard function of dropout according to treatment arm
<b>ESTIMATIONS AND INTERPRETATION</b>				
<b>Main estimated parameters</b>	Fixed effects ( $\beta_0$ , $\beta_1$ , and $\beta_2$ )	Fixed effects ( $\beta_0$ , $\beta_1$ , and $\beta_2$ ) Logistic regression coefficients ( $\psi_0$ , $\psi_1$ , and $\psi_2$ )	(Fixed effects overall patterns ( $\beta_0$ , $\beta_1$ , and $\beta_2$ )) Fixed effects in each pattern $k$ ( $\beta_0^k$ , $\beta_1^k$ , and $\beta_2^k$ ) Proportion in each pattern ( $\pi_k$ )	Fixed effects ( $\beta_0$ , $\beta_1$ , and $\beta_2$ ) Association parameter ( $\alpha$ ) Effect of arm on instantaneous risk of dropout ( $\gamma$ )
<b>Interpretation</b>	Improvement/deterioration of the HRQoL	Improvement/deterioration of the HRQoL Testing MNAR assumption: a non-null $\psi_2$ when probability of dropout is associated with unobserved Y	(Improvement/deterioration of the HRQoL) Improvement/deterioration of the HRQoL in each dropout pattern	Improvement/deterioration of the HRQoL Risk of dropout over time Testing MNAR assumption: a non-null $\alpha$ when instantaneous risk of dropout is associated with current value of Y
<b>Underlying assumptions</b>	–	Normality of the complete (observed and unobserved) Y	Extrapolation of the conditional distribution of Y (given the dropout pattern) beyond the dropout to obtain estimations for the marginal distribution of Y	Conditional independence of Y and T given the random effects Normality assumption of the random effects distribution
<b>Key limitations</b>	Do not account for informative dropout	Dropout in discrete time Not directly available in classical statistical software	Dropout in discrete time Do not directly provide marginal estimates	Computationally challenging to approximate integrals over random effects
<b>Main software</b>	R (nlme) SAS (PROC MIXED) Stata (mixed)	S plus (OSWALD, pcמיד function but not currently available) Implemented with R in our application (sophisticated programming)	Implemented with R in our application (easy programming)	R (JM, JMBayes) SAS (%JM) Stata (stjm)

Legend: LMM Linear Mixed Model, SM Selection Model, PMM Pattern-Mixture Model, SPM Shared-Parameter Model, MCAR Missing Completely At Random; MAR Missing At Random, MNAR Missing Not At Random, HRQoL Health-Related Quality of Life

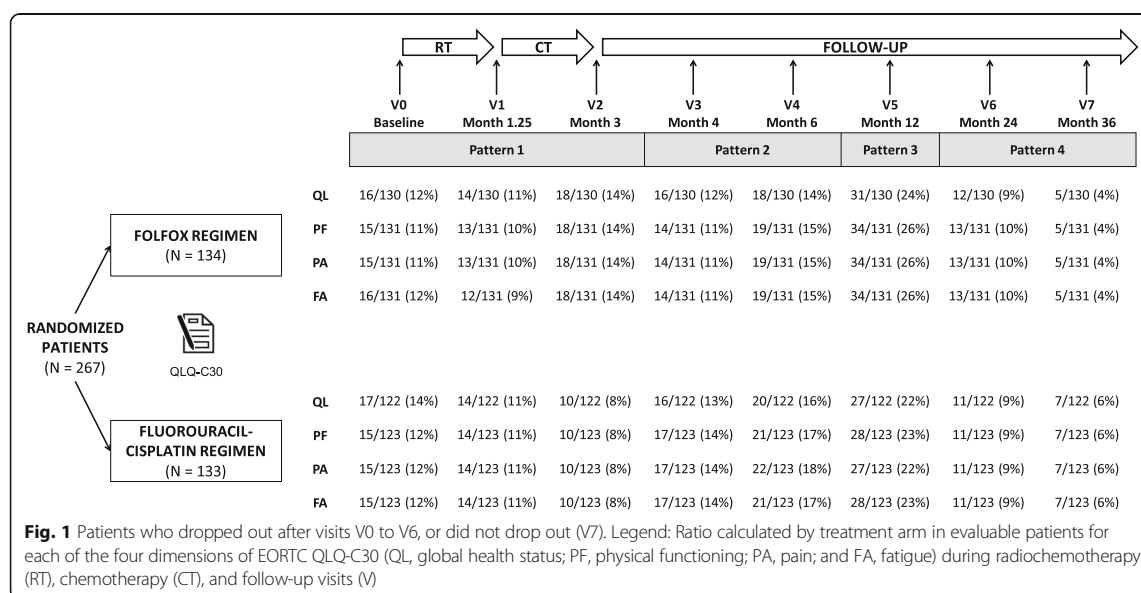
#### Definition of the patterns for the PMM approach

We defined four patterns of dropout with well balanced effectives and a reasonable number of patients by pattern as well as clinically pertinent (see Fig. 1).

The first pattern grouped the patients who dropped out before visit V3 (last HRQoL measurement at V0, V1, or V2), that is, during or just after the period of radio-chemotherapy and chemotherapy treatment. The patients who dropped out between V3 and V5 (last measurement at V3 or V4) formed the second pattern, and between V5 and V6 (last measurement at V5) the third pattern. The last pattern grouped the patients who

dropped out between V6 and V7 (last measurement at V6) and the patients who did not drop out. For the QL dimension for example, the 252 evaluable patients were distributed as follows: 89/252 ( $\pi_1 = 35\%$ ), 70/252 ( $\pi_2 = 28\%$ ), 58/252 ( $\pi_3 = 23\%$ ), and 35/252 ( $\pi_4 = 14\%$ ) in the four respective patterns (for the other dimensions, see Fig. 1).

The results of the longitudinal analysis of the QL, PF, PA, and FA scales of the EORTC QLQ-C30 using the four previously described approaches are summarized in Table 2 (estimates, 95% confidence intervals, and associated  $p$ -values of the Wald test) and graphically



represented in Fig. 2 (estimated slope  $\hat{\beta}_1$  and interaction  $\hat{\beta}_2$  parameters).

No significant treatment-by-time interaction effect  $\beta_2$  was exhibited by the LMM. This was also the case for the SM, PMM, and SPM that had taken into account the dropout. Thus, none of the models suggested a significant effect of the treatment on the score evolution of the QL, PF, PA, and FA scales. The interaction parameters for the LMM (QL:  $\hat{\beta}_2 = -0.130$ ; PF:  $\hat{\beta}_2 = -0.112$ ; PA:  $\hat{\beta}_2 = 0.276$ ; and FA:  $\hat{\beta}_2 = 0.275$ ) were very close to those for the SM and SPM for all four dimensions. The interaction estimates from the PMM differed greatly from those of the other methods (QL:  $\hat{\beta}_2 = 0.464$ ; PF:  $\hat{\beta}_2 = 0.300$ ; PA:  $\hat{\beta}_2 = -0.501$ ; and FA:  $\hat{\beta}_2 = -0.098$ ) but also showed a greater uncertainty (larger confidence intervals).

The LMM showed a significant time effect for three of the four dimensions. More precisely, this model showed an increase in scale QL ( $\hat{\beta}_1 = 0.513$ ,  $p < 0.001$ ) and a decrease in scales PA ( $\hat{\beta}_1 = -0.472$ ,  $p = 0.008$ ) and FA ( $\hat{\beta}_1 = -0.514$ ,  $p = 0.003$ ), reflecting a better level of HRQoL.

The SM confirmed or contradicted these results, depending on whether an association with the probability of dropout was detected or not. The SM and LMM estimated similar effects of time in the QL and in the PA scale where the dropout seemed to be ignorable (non-significant  $\hat{\psi}_2$ ). However, there were unclear results with optimization difficulties: for scale QL, a numerical issue when inverting the Hessian matrix made it impossible to estimate the standard errors of  $\hat{\beta}_1$ , and therefore its confi-

dence interval and associated  $p$ -value were not available; in view of the results for the PA scale, we could question whether or not the algorithm converged to a local minimum. When the SM detected an informative dropout (PF:  $\hat{\psi}_2 = 0.107$ ,  $p < 0.001$  and FA:  $\hat{\psi}_2 = -0.097$ ,  $p < 0.001$ ), the estimated effect of time was larger than that estimated by the LMM, with a substantial increase in PF (SM:  $\hat{\beta}_1 = 1.434$ ,  $p < 0.001$  vs. LMM:  $\hat{\beta}_1 = -0.164$ ,  $p = 0.266$ ) and decrease in FA (SM:  $\hat{\beta}_1 = -2.484$ ,  $p < 0.001$  vs. LMM:  $\hat{\beta}_1 = -0.514$ ,  $p = 0.003$ ). However, the values of  $\hat{\psi}_2$  were counterintuitive, suggesting that the probability of dropout increased with an unobserved score value that corresponded to a higher level of HRQoL.

The marginal effect of time derived from the PMM estimates was ambiguous for all dimensions. For scales QL and PA, the direction of the time effect (i.e., the sign of  $\hat{\beta}_1$ ) was reversed and no longer significant compared to the LMM. For the PF and FA scales, the HRQoL deterioration was aggravated compared to the LMM, with a significant increase in PF (PMM:  $\hat{\beta}_1 = -2.652$ ,  $p < 0.001$  vs. LMM:  $\hat{\beta}_1 = -0.164$ ,  $p = 0.266$ ) and FA (PMM:  $\hat{\beta}_1 = 3.157$ ,  $p < 0.001$  vs. LMM:  $\hat{\beta}_1 = -0.514$ ,  $p = 0.003$ ), corresponding exactly with the same dimensions for which the SM had detected informative dropout.

We observed that the estimated effect of time in the first pattern differed greatly from those in all other patterns (see also Fig. 3, which depicts the score trajectories by pattern).

The estimates in this pattern with a maximum of three repeated measures showed poor functional capacities

**Table 2** Results of the longitudinal analysis of the four dimensions of interest of the EORTC QLQ-C30

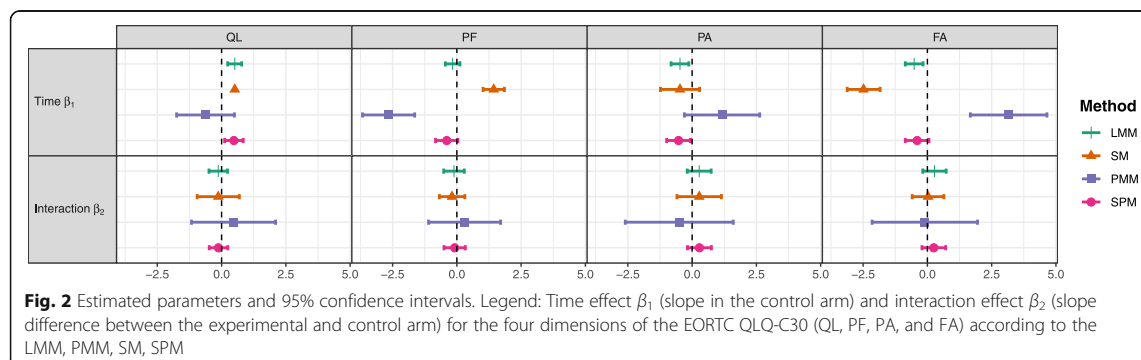
	LMM			SM			PMM			SPM		
	Estimate	CI 95%	p	Estimate	CI 95%	p	Estimate	CI 95%	p	Estimate	CI 95%	p
<b>QL (N = 252)</b>												
<b>Longitudinal outcome</b>												
$\beta_1$	0.513	[0.24; 0.78]	<0.001	$\beta_1$	0.513	NA	$\beta_1$	-0.628	[-1.75; 0.50]	0.275	$\beta_1$	0.477
$\beta_2$	-0.130	[-0.49; 0.23]	0.472	$\beta_2$	-0.130	[-0.95; 0.69]	$\beta_2$	0.464	[-1.17; 2.10]	0.578	$\beta_2$	-0.122
<b>Dropout</b>												
$\psi_0$				$\psi_0$	-0.808	[-1.31; -0.31]	$\beta_1^1$	-2.611	[-5.59; 0.37]	0.086	$\gamma$	-0.042
$\psi_1$				$\psi_1$	-0.005	[-0.01; 0.00]	$\beta_2^1$	1.256	[-3.17; 5.69]	0.579	$\alpha$	-0.003
$\psi_2$				$\psi_2$	0.005	[0.00; 0.01]	$\beta_3^1$	-0.649	[-2.02; 0.72]	0.353	$\xi_1$	0.110
							$\beta_2^2$	0.722	[-0.96; 2.41]	0.401	$\xi_2$	0.105
							$\beta_3^2$	-0.009	[-0.69; 0.67]	0.980	$\xi_3$	0.181
							$\beta_2^3$	-0.248	[-1.14; 0.64]	0.586	$\xi_4$	0.122
							$\beta_1^4$	0.477	[0.15; 0.81]	0.005	$\xi_5$	0.06
							$\beta_2^4$	-0.123	[-0.54; 0.29]	0.559	$\xi_6$	0.065
											$\xi_7$	0.071
<b>PF (N = 254)</b>												
<b>Longitudinal outcome</b>												
$\beta_1$	-0.164	[-0.45; 0.12]	0.266	$\beta_1$	1.434	[1.02; 1.85]	$\beta_1$	-2.652	[-3.67; -1.64]	<0.001	$\beta_1$	-0.394
$\beta_2$	-0.112	[-0.51; 0.29]	0.586	$\beta_2$	-0.189	[-0.68; 0.31]	$\beta_2$	0.300	[-1.10; 1.70]	0.676	$\beta_2$	-0.084
<b>Dropout</b>												
$\psi_0$				$\psi_0$	-3.231	[-5.16; -1.30]	$\beta_1^1$	-6.526	[-9.16; -3.89]	<0.001	$\gamma$	-0.003
$\psi_1$				$\psi_1$	-0.089	[-0.11; -0.06]	$\beta_2^1$	1.135	[-2.86; 5.13]	0.578	$\alpha$	-0.015
$\psi_2$				$\psi_2$	0.107	[0.07; 0.14]	$\beta_3^1$	-1.805	[-2.98; -0.63]	0.003	$\xi_1$	0.272
							$\beta_2^2$	0.362	[-1.15; 1.88]	0.640	$\xi_2$	0.270
							$\beta_1^2$	-0.474	[-1.06; 0.11]	0.115	$\xi_3$	0.474
							$\beta_2^3$	0.017	[-0.81; 0.84]	0.968	$\xi_4$	0.309
							$\beta_1^4$	0.152	[-0.16; 0.46]	0.340	$\xi_5$	0.158
							$\beta_2^4$	-0.185	[-0.61; 0.24]	0.391	$\xi_6$	0.165
											$\xi_7$	0.185
<b>PA (N = 254)</b>												
<b>Longitudinal outcome</b>												
$\beta_1$	-0.472	[-0.82; -0.13]	0.008	$\beta_1$	-0.472	[-1.23; 0.29]	$\beta_1$	1.166	[-0.30; 2.63]	0.118	$\beta_1$	-0.527
$\beta_2$	0.276	[-0.19; 0.74]	0.242	$\beta_2$	0.276	[-0.59; 1.14]	$\beta_2$	-0.501	[-2.60; 1.60]	0.640	$\beta_2$	0.285
<b>Dropout</b>												
$\psi_0$				$\psi_0$	0	[-0.20; 0.20]	$\beta_1^1$	4.582	[0.54; 8.62]	0.027	$\gamma$	-0.012
$\psi_1$				$\psi_1$	0	[-0.01; 0.01]	$\beta_2^1$	-2.505	[-8.47; 3.46]	0.411	$\alpha$	-0.003
$\psi_2$				$\psi_2$	0	[-0.01; 0.01]	$\beta_3^1$	0.281	[-1.52; 2.08]	0.759	$\xi_1$	0.086



**Table 2** Results of the longitudinal analysis of the four dimensions of interest of the EORTC QLQ-C30 (Continued)

	LMM			SM			PMM			SPM								
	Estimate	CI 95%	p	Estimate	CI 95%	p	Estimate	CI 95%	p	Estimate	CI 95%	p						
FA (N = 254)							$\beta_2^2$	0.048	[-2.15; 2.25]	0.966	$\xi_2$	0.086						
							$\beta_1^3$	0.264	[-0.60; 1.13]	0.551	$\xi_3$	0.152						
							$\beta_2^3$	0.239	[-0.93; 1.41]	0.689	$\xi_4$	0.098						
							$\beta_1^4$	-0.511	[-0.93; -0.09]	0.017	$\xi_5$	0.053						
							$\beta_2^4$	0.267	[-0.27; 0.80]	0.328	$\xi_6$	0.055						
											$\xi_7$	0.058						
Longitudinal outcome	$\beta_1$	-0.514	[-0.86; -0.17]	0.003	$\beta_1$	-2.484	[-3.12; -1.84]	<0.001	$\beta_1$	3.157	[1.67; 4.65]	<0.001	$\beta_1$	-0.399	[-0.86; 0.06]	0.089		
	$\beta_2$	0.275	[-0.18; 0.73]	0.239	$\beta_2$	0.027	[-0.59; 0.64]	0.932	$\beta_2$	-0.098	[-2.15; 1.95]	0.926	$\beta_2$	0.249	[-0.21; 0.71]	0.288		
Dropout							$\psi_0$	-0.907	[-1.27; -0.54]	<0.001	$\beta_1^1$	8.574	[4.67; 12.48]	<0.001	$\gamma$	-0.019	[-0.27; 0.23]	0.883
							$\psi_1$	0.047	[0.03; 0.06]	<0.001	$\beta_2^1$	-0.436	[-6.28; 5.41]	0.884	$\alpha$	0.006	[-0.01; 0.02]	0.299
							$\psi_2$	-0.097	[-0.12; -0.07]	<0.001	$\beta_1^2$	3.118	[1.39; 4.85]	<0.001	$\xi_1$	0.068	[-2.52; 2.66]	<0.001
											$\beta_2^2$	-1.12	[-3.29; 1.05]	0.312	$\xi_2$	0.064	[-2.55; 2.68]	<0.001
											$\beta_1^3$	0.626	[-0.20; 1.45]	0.139	$\xi_3$	0.117	[-2.48; 2.71]	<0.001
											$\beta_2^3$	-0.087	[-1.21; 1.04]	0.880	$\xi_4$	0.076	[-2.49; 2.64]	<0.001
											$\beta_1^4$	-0.737	[-1.14; -0.34]	<0.001	$\xi_5$	0.041	[-2.48; 2.56]	<0.001
										$\beta_2^4$	0.382	[-0.14; 0.90]	0.149	$\xi_6$	0.045	[-2.39; 2.48]	<0.001	
													$\xi_7$	0.049	[-2.48; 2.58]	<0.001		
Legend: LMM Linear Mixed Model, SM Selection Model, PMM Pattern-Mixture Model, SPM Shared-Parameter Model, CI Confidence Interval, QL Global Health Status, PF Physical functioning, PA Pain, FA Fatigue																		

Legend: LMM Linear Mixed Model, SM Selection Model, PMM Pattern-Mixture Model, SPW Shared-Parameter Model, CI Confidence Interval, QL Global Health Status, PF Physical functioning, PA Pain, FA Fatigue



(QL:  $\hat{\beta}_1^1 = -2.611, p = 0.086$  and PF:  $\hat{\beta}_1^1 = -6.526, p < 0.001$ ) and high levels of symptoms (PA:  $\hat{\beta}_1^1 = 4.582, p = 0.027$  and FA:  $\hat{\beta}_1^1 = 8.574, p < 0.001$ ). The estimates in this pattern were so important that they highly influenced the marginal estimates which could explain the difference in comparison with the other models.

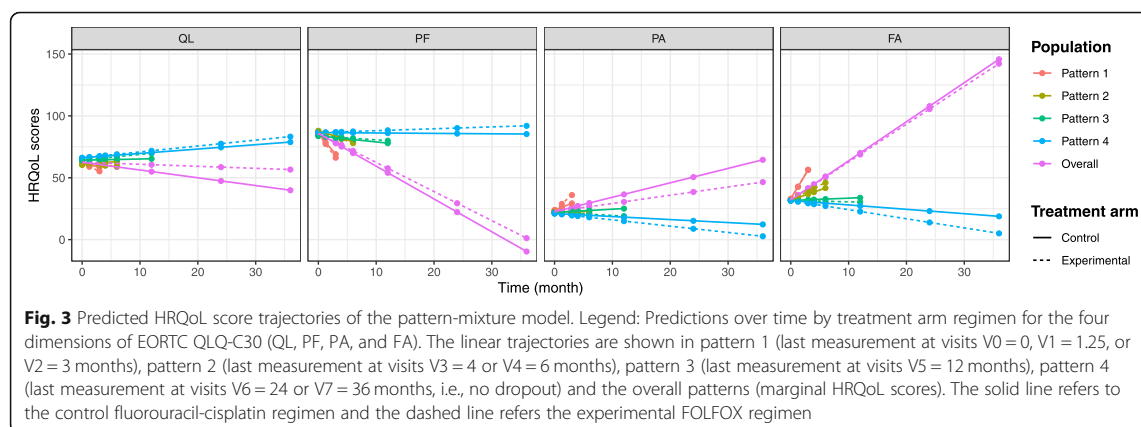
As for the treatment-by-time interaction effect, we also observed that the 95% confidence intervals for the time effect were much larger than those seen in the other three models, reflecting more uncertainty.

For scales QL and PA, the estimated effect of time in the SPM was similar to that in the LMM. No association was detected between the risk of dropout and the current HRQoL score value, which confirmed the results of a non-informative dropout already identified by the SM. In contrast with the SM, the SPM also did not detect an association between the risk of dropout and the score in the FA scale, and the estimated time effect was similar to the LMM estimate. In fact, the SPM only detected a significant association between the risk of dropout and the score in the PF scale (also found by the SM) ( $\hat{\alpha} = -0.015,$

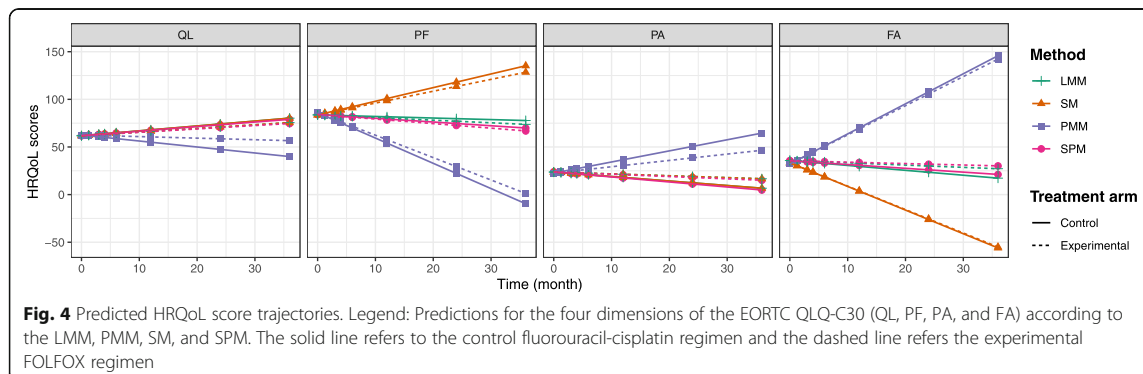
$p = 0.006$ ). In particular, a decrease of 10 points in the PF score corresponded to a risk of dropout multiplied by 1.16 (95% confidence interval: [1.00, 1.35]). The estimation of the time effect was impacted (SPM:  $\hat{\beta}_1 = -0.394, p = 0.078$  vs. LMM:  $\hat{\beta}_1 = -0.164, p = 0.266$ ). Finally, the SPM allowed a more detailed analysis of the dropout process. The baseline hazard function was high at the beginning of the study and then decreased over time for the four scales (see the  $\hat{\xi}_1, \dots, \hat{\xi}_7$  estimates). Besides this, the arm effect  $\gamma$  in the survival model was always non-significant, which suggests that there was no difference in the risk of dropout between the treatment arms.

Finally, Fig. 4 depicts how the differences between the models impacted the estimated HRQoL score trajectories.

The trajectories predicted by the PMM differed from the other models, showing poor functional capacities (QL and PF) and high levels of symptoms (PA and FA). The trajectories predicted by the SM contrasted with those of the PMM, particularly for scales PF and FA. Globally, the trajectories predicted by the SPM were consistent with those of the LMM.







## Discussion

Three approaches exist to model the joint distribution of a longitudinal outcome, such as a longitudinal score, and a dropout process: the SM, the PMM, and the SPM. In this article, we have compared them; firstly, from a methodological point of view, and secondly, when applied to data from the randomized clinical trial PRODIGE 5/ACCORD 17, which included 267 patients with advanced esophageal cancer. We have also compared the results of the three models with those obtained with the LMM.

All three approaches have different advantages and could be complementary. They also have different drawbacks and require assumptions that are untestable since they are based on unobserved data.

The PMM makes it possible to describe and study the HRQoL trajectories in each dropout pattern. In the application, the PMM revealed that the earlier the patients dropped out, the stronger their HRQoL deterioration. Besides this, by highlighting the different evolutions of HRQoL scores according to the dropout pattern, one can presume that the dropout process is informative. However, the PMM does not directly provide marginal estimates that would allow conclusions to be made for the whole population unless assumptions are made about the evolution of HRQoL trajectories after dropout. In our application we considered a simple PMM model with a linear HRQoL trajectory within each dropout pattern and a first pattern grouping patients with 1, 2 and 3 observations. It resulted in a direct and easy-to-implement formulation of the marginal estimates and implied that the HRQoL score evolution after dropout was extrapolated as an extension of the linear trajectories. This gave results that contradicted those obtained with the other models (the LMM, SM, and SPM) and with larger confidence intervals. Indeed, the first patterns including patients with few repeated measurements and a strong HRQoL deterioration highly influenced the

marginal estimates. Note that in a more complex model, making identifying assumptions would be necessary [28]; a common strategy consists in using identifying restrictions [29]. Although unverifiable, the assumptions necessary to achieve identifiability in the PMMs and obtain marginal estimates have the advantage of being explicit.

The SM and SPM are interesting approaches because they can test the mechanism of missing data through interpretable parameters obtained from the logistic regression (SM) or the Cox model (SPM). In the application, when the dropout was detected as non-informative by the SM or the SPM the results for the trajectories of HRQoL were similar to those of the LMM and led to the same conclusions. Both models detected an informative dropout in the PF dimension but only the SM detected an informative dropout in the FA dimension. The SPM results were consistent with the LMM results and had a coherent interpretation. In contrast, the SM results revealed that the probability of dropout increased with an unobserved score value corresponding to a higher level of HRQoL. It is possible that these unexpected results are the consequence of the strong assumption of a normal distribution of the complete (observed and unobserved) HRQoL score values. Indeed, it has been shown that the SM is particularly sensitive to this unverifiable assumption [24, 30].

The SPM makes also modeling assumptions. In particular, it relies on the conditional independence between the longitudinal outcome and dropout process given the random effects. The random effects are also supposed to be normally distributed. Rizopoulos et al. showed that estimation of the parameters and standard errors could be sensible to misspecification of the random effects distribution, especially when some patients have very few measurements (early dropout) [31]. Note that in this application, we considered that the risk of dropout was associated with the HRQoL score through its current value. Other association structures could be

considered, including the current slope or the random effects alone. The SPM alone is able to take into account dropout by modeling time-to-event data. Thus, unlike the PMM and the SM, the SPM treats the time-to-dropout as continuous. In our application, we used discrete dropout times corresponding to pre-specified assessment times, but the SPM would allow researchers to take into account dropouts corresponding to clinical events such as death, which can occur at any time between the HRQoL assessment times. By contrast, the use of the SPM was facilitated by the standard statistical software [27, 32–34]. Moreover, the existing programs allow for flexible models for the longitudinal outcome, more complex models for the time-to-dropout, and different association structures to capture the association between the longitudinal outcome and the time-to-dropout.

In this article, we have analyzed HRQoL data from the PRODIGE 5/ACCORD 17 clinical trial under three possible MNAR models accounting for informative dropout and the MAR corresponding model. MNAR methods, especially PMM, can also be used for sensitivity analysis to assess the robustness of the results [35].

This work has some limitations. The main objective was to compare MNAR models from a practical point of view but this does not allow to clearly decide between one model or the other. A simulation study would allow a comparison with statistical criteria by example in case of misspecification or by varying the proportion of missing data.

Longitudinal analysis of the HRQoL in the presence of missing data remains complex and unstandardized. Reviews and guidelines about reporting missing patient-reported outcome data in clinical trials have been published [36, 37]. It is recommended that the amount of missing data in each arm is reported and that the statistical methods used to handle missing data are explicitly specified. Nevertheless, there is no consensus for analyzing such data. Indeed, there is a lack of standardization and a gap between the development of statistical methods and their use in clinical trials [38, 39].

## Conclusions

This article aims to facilitate the understanding and use of such methods allowing analysis of longitudinal HRQoL data that include missing data due to dropout. Nevertheless, including in clinical trial protocol a plan to collect the reasons for non-responses would help to better characterize the missingness. Then, if informative dropout is suspected, we recommend using models that account for dropout, such as the SPM. In studies where no information is available on the reasons for missingness, the SPM can be used to confirm or invalidate the results of LMMs.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12874-020-01104-w>.

**Additional file 1: Figure S1.** Compliance in completing the EORTC QLQ-C30. Compliance in completing the entire questionnaire and for the four dimensions QL, PF, PA, and FA (ratio of the number of available questionnaires or scores to the number of expected questionnaires) at each HRQoL assessment visit (V) by treatment arm during radiochemotherapy (RT), chemotherapy (CT), and follow-up.

## Abbreviations

EORTC: European Organisation for Research and Treatment of Cancer; FA: Fatigue; HRQoL: Health-related quality of life; LMM: Linear mixed model; MAR: Missing at random; MCAR: Missing completely at random; MNAR: Missing not at random; PA: Pain; PF: Physical functioning; PMM: Pattern-mixture model; QL: Global health status; QLQ-C30: Quality of Life Questionnaire Core 30; RML: Restricted maximum likelihood; SM: Selection model; SPM: Shared-parameter model

## Acknowledgements

Not applicable

## Authors' contributions

BC performed the statistical analyses, interpretation and wrote the manuscript. CT and CM supervised this work, helped to interpret the results, and corrected the manuscript. AA and EC critically commented the manuscript. TC was the main investigator of the clinical trial (NCT00861094) and participated to the patient's inclusion. BJ and TC participated in the day-to-day running of the trial and contributed to the acquisition of the data. All authors read and approved the final manuscript.

## Funding

This work was supported by a grant from the « Institut National du Cancer (INCA 11862) » and the Region Occitanie (Program "Allocation Doctorale 2017"). (Note: The funding body had no other role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.)

## Availability of data and materials

The dataset analyzed during the current study is not publicly available due to confidentiality requirements. Data are however available from the main coordinator of the clinical trial, Pr Thierry Conroy, upon reasonable request, and with permissions of the study sponsor UNICANCER R&D.

## Ethics approval and consent to participate

UNICANCER R&D, the sponsor of the PRODIGE 5/ACCORD 17 trial (ClinicalTrials.gov Identifier: NCT00861094), provided permission for the data base access. All participants of the PRODIGE 5/ACCORD 17 trial provided written informed consent. Patient consent was not required for this study as we performed a secondary analysis of existing data.

## Consent for publication

Not applicable

## Competing interests

The authors have declared no conflicts of interest.

## Author details

<sup>1</sup>Biometrics Unit, Montpellier Cancer Institute (ICM), University of Montpellier, 208, avenue des Apothicaires, 34298 Montpellier, France. <sup>2</sup>French National Platform Quality of Life and Cancer, Montpellier, France. <sup>3</sup>Institute of Cancer Research of Montpellier (IRCM)- Inserm 1194, ICM, University of Montpellier, Montpellier, France. <sup>4</sup>Methodology and Quality of Life Unit in Oncology-Inserm UMR 1098, University Hospital of Besançon, Besançon, France. <sup>5</sup>University Bourgogne Franche-Comté, Inserm, EFS BFC, UMR 1098, Interactions Hôte-Greffon-Tumeur/Ingénierie Cellulaire et Génique, Besançon, France. <sup>6</sup>UNICANCER R&D, French Federation of Comprehensive Cancer Centres, Paris, France. <sup>7</sup>Department of Medical Oncology, Institut de

Cancérologie de Lorraine, Vandoeuvre-lès-Nancy, France. <sup>8</sup>APEMAC, équipe MICS, Université de Lorraine, Nancy, France.

Received: 27 May 2020 Accepted: 20 August 2020

Published online: 03 September 2020

## References

- Osoba D. Health-related quality of life and cancer clinical trials. *Ther Adv Med Oncol*. 2011;3:57–71.
- Rubin DB. Inference and missing data. *Biometrika*. 1976;63:581–92.
- Little RJA, Rubin DB. Statistical analysis with missing data. New York: Wiley; 1986.
- Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay CR. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Qual Life Res*. 2016;25:2977–83.
- Fairclough DL. Design and analysis of quality of life studies in clinical trials. Chapman and Hall/CRC. 2010.
- Cnaan A, Laird NM, Slator P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med*. 1997;16:2349–80.
- Fielding S, Fayers PM, Loge JH, Jordhøy MS, Kaasa S. Methods for handling missing data in palliative care research. *Palliat Med*. 2006;20:791–8.
- Hussain JA, White IR, Langan DJ, Johnson MJ, Currow DC, Torgerson DJ, et al. Missing data in randomized controlled trials testing palliative interventions pose a significant risk of bias and loss of power: a systematic review and meta-analyses. *J Clin Epidemiol*. 2016;74:57–65.
- Fairclough DL, Peterson HF, Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? *Stat Med*. 1998;17:667–77.
- DeSouza CM, Legedza ATR, Sankoh AJ. An overview of practical approaches for handling missing data in clinical trials. *J Biopharm Stat*. 2009;19:1055–73.
- Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. *Test Madr Spain*. 2009;18:1–43.
- Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. *Stat Med*. 1997;16:259–72.
- Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *J Am Stat Assoc*. 1995;90:1112–21.
- Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin*. 2004;809–34.
- Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. *Stat Med*. 2006;25:143–63.
- Dupuy J-F. Joint modeling of survival and nonignorable missing longitudinal quality-of-life data. In: Mesbah M, Cole BF, Lee M-LT, editors. *Statistical methods for quality of life studies: design, measurements and analysis*. Boston: Springer US; 2002. p. 309–22. Available from: [https://doi.org/10.1007/978-1-4757-3625-0\\_25](https://doi.org/10.1007/978-1-4757-3625-0_25).
- Michiels B, Molenberghs G, Bijnens L, Vangeneugden T, Thijs H. Selection models and pattern-mixture models to analyse longitudinal quality of life data subject to drop-out. *Stat Med*. 2002;21:1023–41.
- Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res*. 2013;23:440–59.
- Du H, Hahn EA, Cella D. The impact of missing data on estimation of health-related quality of life outcomes. *Anal Randomized Longitud Clin Trial*. 2011;1:134–44.
- Conroy T, Galais M-P, Raoul J-L, Bouché O, Gourgu-Bourgade S, Douillard J-Y, et al. Definitive chemoradiotherapy with FOLFOX versus fluorouracil and cisplatin in patients with oesophageal cancer (PRODIGES/ACCORD17): final results of a randomised, phase 2/3 trial. *Lancet Oncol*. 2014;15:305–14.
- Bascoul-Mollevi C, Gourgu S, Galais M-P, Raoul J-L, Bouché O, Douillard J-Y, et al. Health-related quality of life results from the PRODIGE 5/ACCORD 17 randomised trial of FOLFOX versus fluorouracil–cisplatin regimen in oesophageal cancer. *Eur J Cancer*. 2017;84:239–49.
- Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *JNCI J Natl Cancer Inst*. 1993;85:365–76.
- Fayers P, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A. EORTC QLQ-C30 scoring manual, European Organisation for Research and Treatment of Cancer. 3rd ed; 2001.
- Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *J R Stat Soc: Ser C: Appl Stat*. 1994;43:49–93.
- Pauler DK, McCoy S, Moynour C. Pattern mixture models for longitudinal quality of life studies in advanced stage disease. *Stat Med*. 2003;22:795–809.
- Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS. N. Y: Springer; 2000.
- Rizopoulos D. JM: an R package for the joint modelling of longitudinal and time-to-event data. *J Stat Softw Artic*. 2010;35:1–33.
- Thijs H, Molenberghs G, Michiels B, Verbeke G, Curran D. Strategies to fit pattern-mixture models. *Biostatistics*. 2002;3:245–65.
- Little RJA. Pattern-mixture models for multivariate incomplete data. *J Am Stat Assoc*. 1993;88:125–34.
- Verbeke G, Molenberghs G, Thijs H, Lesaffre E, Kenward MG. Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics*. 2001;57:7–14.
- Rizopoulos D, Verbeke G, Molenberghs G. Shared parameter models under random effects misspecification. *Biometrika*. 2008;95:63–74.
- Rizopoulos D. The R package JMBayes for fitting joint models for longitudinal and time-to-event data using MCMC. *J Stat Softw Artic*. 2016;72:1–46.
- Crowther MJ, Abrams KR, Lambert PC. Joint modeling of longitudinal and survival data. *Stata J*. 2013;13:165–84.
- Garcia-Hernandez A, Rizopoulos D. %JM: A SAS macro to fit jointly generalized mixed models for longitudinal data and time-to-event responses. *J Stat Softw Artic*. 2018;84:1–29.
- Molenberghs G, Kenward M. Missing data in clinical studies: Wiley; 2007.
- Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA*. 2013;309:814–22.
- Little RJ, D'agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012;367:1355–60.
- Bottomley A, Pe M, Sloan J, Basch E, Bonnetain F, Calvert M, et al. Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials. *Clin Trials*. 2018;15:624–30.
- Bell ML, Fiero M, Horton NJ, Hsu C-H. Handling missing data in RCTs: a review of the top medical journals. *BMC Med Res Methodol*. 2014;14:118.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



### 3.3 Modèles conjoints

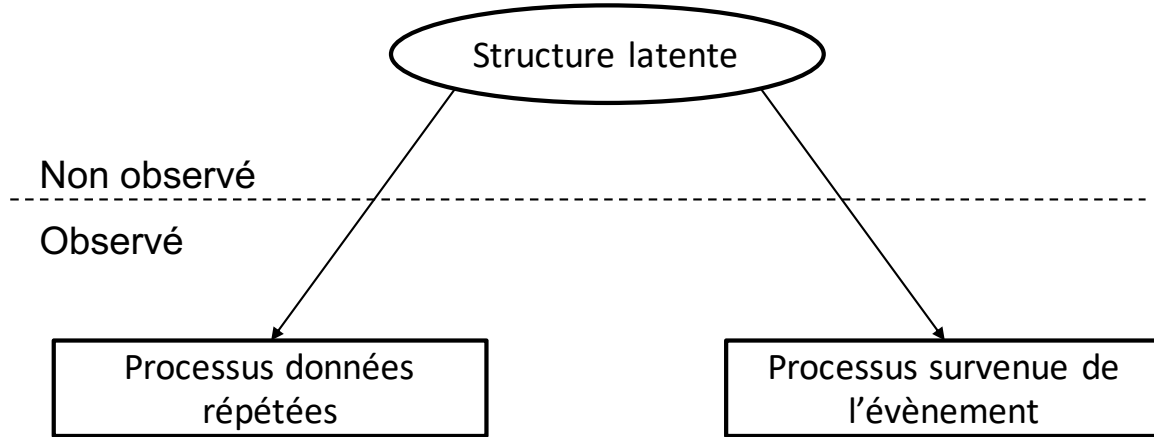
Les modèles conjoints ont été introduits vers la fin des années 1990, mais très peu utilisés jusqu'à la fin des années 2000 en raison de leur complexité calculatoire et du manque de logiciels disponibles. C'est un domaine actif de la recherche suscitant un intérêt qui a connu un développement considérable ces dernières années avec l'utilisation de nouvelles méthodes numériques et grâce à leur implémentation dans les logiciels de statistique. Les modèles conjoints sont notamment utilisés pour tenir compte d'une variable endogène (ou interne) (Kalbfleisch and Prentice, 2011, chap. 6.3) dépendante du temps et mesurée avec une erreur dans un modèle de survie, mais aussi pour modéliser une variable longitudinale affectée par des sorties d'étude informatives (Tsiatis and Davidian, 2004; Wu et al., 2011). Une des premières applications médicales des modèles conjoints concerne l'étude du lien entre l'évolution du nombre de cellules CD4 et la survie chez des patients infectés par le VIH (Faucett and Thomas, 1996). Les modèles conjoints sont aujourd'hui répandus dans différents domaines d'application. Par exemple en cancérologie, ils sont utilisés pour étudier l'évolution du marqueur de l'antigène spécifique de la prostate (PSA) conjointement à la progression du cancer de la prostate (Proust-Lima et al., 2014).

#### 3.3.1 Structure générale de la modélisation conjointe

Le principe des modèles conjoints est de considérer deux processus, l'un qui décrit la variable mesurée au cours du temps, le second qui modélise le temps jusqu'à un évènement d'intérêt, et de lier ces deux processus par une structure latente commune capable de capturer leur association. Les modèles « à effets partagés » considèrent que ce sont les effets aléatoires qui permettent de capturer la dépendance entre les deux processus (Tsiatis and Davidian, 2004; Henderson et al., 2000).

Dans cette thèse, la variable mesurée au cours du temps correspond aux scores de QdV et nous nous intéresserons uniquement aux modèles conjoints à effets aléatoires partagés (voir Figure 3.1). Ils consistent dans notre cas à modéliser le risque jusqu'à sortie d'étude comme une fonction qui dépend du score de QdV, lui-même

modélisé par un modèle linéaire mixte.



**Figure 3.1** – Représentation graphique d'un modèle conjoint constitué de deux processus (pour décrire les données répétées et la survie d'un évènement) partageant une structure d'association commune décrite au travers des effets aléatoires.

#### 3.3.2 Modèle conjoint à effets partagés

Le prédicteur linéaire du modèle linéaire mixte qui décrit le score de QdV (et plus généralement une fonction de ce prédicteur), est inclus dans les variables explicatives du modèle de type « temps jusqu'à évènement » où l'évènement correspond à la sortie d'étude (modèle de survie), comme par exemple un modèle à risques proportionnels. Les effets aléatoires sont alors utilisés pour tenir compte à la fois de la corrélation entre les mesures répétées d'un même patient, et de l'association entre les deux processus d'où la dénomination de modèles conjoints à effets partagés.

Pour une population de taille  $n$ , soit  $Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{in_i}))^\top$  le vecteur qui contient l'ensemble des  $n_i$  mesures répétées du patient  $i$  ( $i = 1, \dots, n$ ) et prenant la valeur  $Y_i(t_{ij})$  au temps  $t_{ij}$  ( $j = 1, \dots, n_i$ ). La variable de score de QdV est supposée continue et distribuée selon une distribution gaussienne.

Le modèle conjoint à effets aléatoires partagés (dans sa forme la plus classique) se décompose en deux sous-modèles : un modèle linéaire mixte pour la modélisation du score de QdV et un modèle à risques proportionnels pour modéliser le risque de sortie d'étude, partageant les mêmes effets aléatoires  $b_i$ . Le modèle

conjoint s'écrit alors :

$$\begin{cases} Y_i(t) = Y_i^*(t) + \varepsilon_i(t) \\ \quad = X_i^\top(t)\beta + Z_i^\top(t)b_i + \varepsilon_i(t) \\ h_i(t) = h_0(t) \exp \{W_i^\top \gamma + \alpha Y_i^*(t)\}, \quad t > 0 \end{cases} \quad (3.1)$$

avec le vecteur  $X_i(t)$  des variables explicatives associé aux effets fixes  $\beta$ , le vecteur  $Z_i(t)$  associé aux effets aléatoires  $b_i$  et  $\varepsilon_i(t)$  le terme d'erreur de mesure. Les effets aléatoires sont supposés suivre une loi normale d'espérance zéro et de matrice de variance-covariance  $D$  :  $b_i \sim \mathcal{N}(0, D)$  et les termes d'erreur sont supposés mutuellement indépendants, indépendants des effets aléatoires et distribués selon une loi normale d'espérance zéro et de variance  $\sigma^2$  :  $\varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$ . Dans la formulation du modèle linéaire mixte, la valeur observée du score  $Y_i(t)$  est égale à la vraie valeur  $Y_i^*(t)$  non observée à laquelle s'ajoute un terme d'erreur  $\varepsilon_i(t)$ . Pour la partie sous-modèle de survie,  $h_0(\cdot)$  désigne la fonction du risque de base,  $W_i$  le vecteur des variables explicatives non dépendantes du temps associé aux effets  $\gamma$  et enfin  $Y_i^*(t)$  la variable dépendante du temps associée au paramètre  $\alpha$ , qui quantifie l'association entre les deux processus. Il est possible de spécifier une forme paramétrique pour le risque de base  $h_0(\cdot)$  (à partir par exemple d'une distribution exponentielle, de Weibull, Gompertz ou Gamma) ou d'utiliser des formes plus flexibles basées sur des fonctions continues par morceaux ou des fonctions splines. La dépendance entre les deux processus est assurée ici au travers de la vraie valeur de score mais il existe d'autres formes de dépendance. Les formes les plus communément utilisées sont : la vraie valeur courante du score (avec le terme  $\alpha Y_i^*(t)$ ), l'évolution courante du score (avec le terme  $\alpha \frac{\partial Y_i^*(t)}{\partial t}$ ) ou les deux (avec le terme  $\alpha_1 Y_i^*(t) + \alpha_2 \frac{\partial Y_i^*(t)}{\partial t}$ ) ([Rizopoulos, 2012b](#); [Lawrence Gould et al., 2015](#)).

#### 3.3.3 Interprétation du modèle

La formulation du modèle conjoint a l'avantage de conserver les interprétations habituelles du modèle linéaire mixte et du modèle de survie lorsqu'ils sont pris séparément. En plus, le paramètre  $\alpha$  permet de mesurer la force d'association entre les deux processus sous l'hypothèse des risques proportionnels. Par exemple, en considérant le cas d'une association avec la valeur courante, le terme  $\exp(\alpha)$  s'in-

terprète comme un rapport des risques instantanés (*hazard ratio*) et correspond à l'augmentation du risque de sortie d'étude au temps  $t$  pour une augmentation d'une unité du prédicteur linéaire  $Y_i^*(t)$ .

## 3.4 Discussion et conclusion

Dans l'article de ce chapitre, nous avons présenté les trois principaux modèles pour tenir compte de sorties d'étude informatives pour l'analyse longitudinale de la qualité de vie dans les essais cliniques. Nous les avons décrits et illustrés puis nous avons montré que chaque modèle possède ses propres avantages et inconvénients.

Le PMM est l'approche certainement la plus intuitive qui consiste à modéliser la variable longitudinale conditionnellement à des profils de sorties d'étude qui se construisent par rapport aux dates de sorties d'étude observées. Théoriquement, il est possible d'envisager autant de profils qu'il y a de visites planifiées (sans compter l'inclusion). Par exemple, pour les données de l'essai clinique PRODIGE 5/ACCORD 17, ce sont 7 profils différents qui auraient pu être choisis selon les visites planifiées par le protocole de l'étude. Parfois, certains profils peuvent regrouper trop peu de patients ou trop peu de mesures, par exemple lorsqu'un profil regroupe les patients qui ont rempli le questionnaire de QdV uniquement à l'inclusion. C'est la raison pour laquelle certaines contraintes peuvent être nécessaires pour assurer l'identifiabilité des paramètres du modèle ([Verbeke and Molenberghs, 2009](#); [Thijs et al., 2002](#)) comme des relations paramétriques entre les paramètres associés aux différents profils ([Molenberghs et al., 1998](#)). Bien que certains auteurs perçoivent ces contraintes comme un inconvénient, elles peuvent être considérées comme un atout obligeant à réfléchir sur les hypothèses à formuler. De plus, ces hypothèses ont l'avantage d'être nécessairement explicites. Pour notre analyse, nous avons décidé de regrouper certaines dates de sorties d'étude pour finalement ne former que 4 profils différents. Ainsi, nous n'avons pas eu besoin d'imposer de contraintes supplémentaires pour l'estimation du modèle. Néanmoins, le principal inconvénient des PMMs est qu'il n'est en général pas possible d'estimer directement la distribution marginale des paramètres, car les pa-



ramètres sont obtenus conditionnellement aux profils, contrairement aux SMs et SPMs. Pour la modélisation des données de QdV dans un essai clinique, cela permet de connaître l’effet global du bras de traitement sur l’évolution globale de la QdV. Dans un PMM, les estimations marginales peuvent s’obtenir en extrapolant les trajectoires au-delà des sorties d’étude. Cela nous a permis de pouvoir comparer l’effet du traitement sur l’évolution globale de la QdV.

Un point important à souligner est la mise en pratique de ces modèles. Le PMM est assez simple à mettre en œuvre à partir de n’importe quel logiciel où il est possible d’estimer un LMM. En revanche, le SM nécessite une étape importante de programmation (voir le code donné en Annexe C) qui est plus difficile à adapter d’un jeu de données à l’autre. La log-vraisemblance peut devenir difficile à maximiser et surtout, nécessite des calculs d’intégrales souvent complexes, liées à l’hypothèse de normalité des données complètes. Diggle and Kenward (1994) utilisent l’algorithme du simplexe (Nelder and Mead, 1965) pour estimer leur modèle, mais nous avons choisi un algorithme d’optimisation de type Newton-Raphson qui a déjà été utilisé et est connu pour être stable et rapide (Molenberghs and Kenward, 2007). Cependant, il a été difficile, voire impossible, de faire converger l’algorithme d’optimisation pour certaines dimensions de QdV, ce qui peut compromettre les résultats obtenus. C’est un problème déjà rencontré dans la littérature où l’inverse de la matrice Hessienne n’est pas définie positive (Curran et al., 2002). Ces difficultés numériques de convergence des algorithmes d’optimisation soulignent l’importance de l’hypothèse de normalité sur les données complètes de score dans le SM, à laquelle ils sont justement particulièrement sensibles (Kenward, 1998) et qu’il est difficile de diagnostiquer. Concernant les SPMs, il existe de nombreuses implémentations dans les logiciels de statistique standards tels que R, SAS et Stata, pour faciliter l’estimation de ces modèles qui nécessitent aussi des intégrales souvent difficiles à calculer. En effet, l’écriture de la log-vraisemblance du SPM implique le calcul de deux intégrales qui n’ont pas de solution explicite, et qui doivent être approximées par des intégrations numériques. Plusieurs méthodes d’approximation numériques peuvent être utilisées, dont certaines sont implémentées dans le package JM que nous avons utilisé (Rizopoulos, 2010). Nous avons utilisé les quadratures de Gauss-Konrod (pour la fonction de survie du temps de sortie d’étude) et de Gauss-Hermite (pour les intégrales sur les effets aléatoires). En particulier pour les effets aléatoires, nous avons choisi la version



### Chapitre 3. Étude de trois approches pour tenir compte de sorties d'étude informatives

---

pseudo-adaptative de Gauss-Hermite et avons augmenté le nombre de points de quadrature en testant 3, 5, 7 et 9 points de quadrature jusqu'à obtenir des résultats stables. De plus, nous avons validé la robustesse des résultats obtenus avec la version non adaptative du même algorithme. Dans notre application, la dimension des effets aléatoires est assez faible (2), mais dès qu'on l'augmente (e.g. pour capturer des trajectoires non linéaires), il est conseillé d'utiliser la méthode de Laplace ([Rizopoulos et al., 2009](#)) ou la version pseudo-adaptative de Gauss-Hermite (qui place les nœuds de manière optimisée) pour réduire efficacement les temps de calcul ([Rizopoulos, 2012a](#)). Nous avons utilisé la méthode d'optimisation par défaut du package JM qui utilise l'algorithme EM ([Dempster et al., 1977](#)), puis en cas de non convergence un algorithme de type quasi-Newton.

L'association entre la QdV et le délai jusqu'à sortie d'étude a été estimée en utilisant la valeur courante du score de QdV dans la fonction de risque instantané de sortie d'étude. C'est généralement la structure d'association la plus naturelle et la plus utilisée dans les applications. C'est celle que nous choisirons dans la suite du manuscrit. D'autres formes d'association pourraient être envisagées, par exemple en utilisant la pente courante du score de QdV. Dans le cadre d'une analyse de données utilisant un modèle conjoint, la recherche de la meilleure structure d'association peut être réalisée en utilisant des critères statistiques de comparaison de modèles (e.g. basés sur la vraisemblance avec le critère AIC) ou de sélection de modèles ([Andrinopoulou and Rizopoulos, 2016](#)).

Alors qu'il est facile de visualiser les trajectoires de QdV dans les différents profils du PMM, l'utilisation des PMMs semble en revanche peu adaptée pour caractériser le processus à l'origine des données manquantes. Les deux autres modèles (SM et SPM) permettent de bien comprendre la dépendance aux données de QdV observées et non observées au travers des paramètres estimés d'une régression logistique (SM) ou d'un modèle de temps jusqu'à évènement (SPM). Ils font en revanche des hypothèses moins explicites que le PMM, en particulier l'hypothèse de normalité des données complètes pour le SM et l'hypothèse d'indépendance conditionnelle pour le SPM.

De manière générale, il n'est pas possible de distinguer entre un mécanisme MAR et MNAR sur la seule base des données observées et les hypothèses d'un mé-

canisme MNAR restent difficiles à tester (e.g. dépendance du mécanisme de sorties d'étude avec des données non observées). Par conséquent, une approche pour étudier l'impact de ces hypothèses consiste à effectuer des analyses de sensibilité sous diverses hypothèses MNAR. L'analyse de sensibilité s'est largement développée dans les cadres du SM et du PMM ([Thijs et al., 2002](#); [Molenberghs and Kenward, 2007](#); [Verbeke et al., 2001](#)) alors que le SPM a reçu moins d'attention sur ce point ([Njagi et al., 2014](#); [Creemers et al., 2010](#)). Très récemment, une alternative intéressante d'analyse de sensibilité dans les SPMs a été proposée et consiste à utiliser les informations disponibles sur les causes de l'absence de données ([Papageorgiou and Rizopoulos, 2020](#)) : elle est plus facile à mettre œuvre dans les logiciels existants et peut être appliquée pour étudier l'impact sur les estimations des modèles SPMs. Néanmoins, elle nécessite de connaître les causes expliquant l'absence de données qui est une information parfois difficile à obtenir et pas toujours relevée.

Nous avons vu que les raisons de sorties d'étude pouvaient être utilisées pour effectuer des analyses de sensibilité et apporter des informations intéressantes sur le mécanisme à l'origine des données manquantes. L'intégration de cette information supplémentaire dans les modèles semble difficile avec les SMs et PMMs, mais facilitée dans le SPM au travers du sous-modèle de survie (e.g. en utilisant des risques concurrents). Dans le chapitre suivant, nous nous focaliserons sur les modèles conjoints (SPMs) qui peuvent être utilisés pour prendre en compte différentes causes de sorties d'étude.



# Chapitre 4

## Modèles conjoints différenciant plusieurs causes de sorties d'étude

---

### 4.1 Introduction

Dans le chapitre précédent, les modèles conjoints se sont révélés être une approche intéressante pour la prise en compte des sorties d'étude grâce à une modélisation simultanée des scores de QdV et du temps jusqu'à sortie d'étude (modélisé par un modèle à risques proportionnels). Nous rappelons que le délai jusqu'à sortie d'étude correspond au délai depuis la première jusqu'à la dernière évaluation non-manquante de la QdV.

En oncologie, les patients dont on recueille les données de QdV sont souvent à risque de progression ou de décès. Un tel événement clinique peut induire des sorties d'étude susceptibles d'être informatives, c'est-à-dire liées à la QdV non observée ([Fairclough, 2010](#)). Par exemple, la progression de la maladie peut être accompagnée d'une augmentation des symptômes avec pour conséquence le non-remplissage du questionnaire de QdV (à la suite d'une détérioration de celle-ci). Il est également possible que le non-remplissage du questionnaire soit induit par

## **Chapitre 4. Modèles conjoints différenciant plusieurs causes de sorties d'étude**

d'autres raisons (sans rapport avec la QdV). Ainsi, il est probable qu'il existe différents types d'association entre les sorties d'étude et la QdV et que le caractère informatif ou non de la sortie d'étude puisse dépendre de sa cause.

Le modèle conjoint présenté au chapitre précédent (Chapitre 3, voir Équation 3.1) fait l'hypothèse qu'il existe une seule association, mais ne permet en aucun cas de prendre en considération différentes associations liées à différentes causes de sorties d'étude. Une solution proposée consiste à utiliser un sous-modèle à risques concurrents et de l'associer au sous-modèle pour les trajectoires de scores dans un modèle conjoint à risques concurrents. Ce modèle est une extension du modèle conjoint standard, permettant de prendre en compte plusieurs causes de sorties d'étude. Bien que cette approche soit plus intéressante car elle permet de mieux décrire le risque de sortie d'étude et son association avec la QdV, nous ignorons l'impact de ne pas prendre en compte plusieurs causes de sorties d'étude sur l'estimation des paramètres du sous-modèle qui décrit les trajectoires de scores de QdV.

Dans ce chapitre, nous proposons de comparer l'utilisation d'un modèle conjoint à risques concurrents par rapport à un modèle conjoint standard. D'abord, nous avons appliqué un modèle conjoint standard (toutes causes confondues) et un modèle conjoint à risques concurrents en considérant deux causes de sorties d'étude (liées ou non liées à un événement clinique) sur les données de l'essai clinique PRODIGE 5/ACCORD 17. Ensuite, nous avons réalisé une étude de simulations selon plusieurs scénarios afin de comparer la capacité des deux modèles à prendre en compte les sorties d'étude informatives et à estimer correctement les paramètres décrivant les trajectoires des scores de QdV. Enfin, ce chapitre sera conclu par une discussion générale sur l'utilisation de ces modèles, notamment pour l'analyse longitudinale de la QdV en présence de sorties d'étude suspectées informatives.

4.2 Article : *Joint modelling with competing risks of dropout for longitudinal analysis of health-related quality of life in cancer clinical trials*

# **Joint modelling with competing risks of dropout for longitudinal analysis of health-related quality of life in cancer clinical trials**

## **Author information**

Benjamin Cuer<sup>1,2,3</sup>, Thierry Conroy<sup>4,5</sup>, Beata Juzyna<sup>6</sup>, Sophie Gourgou<sup>1,2</sup>, Caroline Mollevi<sup>1,2,7</sup>,  
Célia Touraine<sup>1,2,7</sup>

<sup>1</sup>Montpellier Cancer Institute (ICM), University of Montpellier, Montpellier, France

<sup>2</sup>French National Platform Quality of Life and Cancer, Montpellier, France

<sup>3</sup>Institute of Cancer Research of Montpellier (IRCM) (Inserm 1194), ICM, University of Montpellier, Montpellier, France

<sup>4</sup>Institut de Cancérologie de Lorraine (Department of Medical Oncology), Vandoeuvre-lès-Nancy, France

<sup>5</sup>APEMAC (équipe MICS), Université de Lorraine, Nancy, France

<sup>6</sup>R&D UNICANCER, French Federation of Comprehensive Cancer Centres, Paris, France

<sup>7</sup>Desbrest Institute of Epidemiology and Public Health (IDESP), UMR Inserm, University of Montpellier, Montpellier, France

Corresponding author: Benjamin Cuer, Biometrics Unit, Montpellier Cancer Institute, 208, avenue des Apothicaires, 34298 Montpellier, France.

Mail: [cuerbenjamin@gmail.com](mailto:cuerbenjamin@gmail.com); ORCID: 0000-0002-5634-0704

## **Abstract**

Health-related quality of life (HRQoL) is an important endpoint in cancer clinical trials. Analysis of HRQoL longitudinal data is plagued by missing data, notably due to dropout. Joint models are increasingly receiving attention for modelling longitudinal outcomes and the time-to-dropout. However, dropout can be informative or non-informative depending on the cause. We propose using a joint model that includes a competing risks sub-model for the cause-specific time-to-dropout. We compared a competing risks joint model (CR JM) that distinguishes between two causes of dropout with a standard joint model (SJM) that treats all the dropouts equally. First, we applied the CR JM and SJM to data from 267 patients with advanced oesophageal cancer from the randomized clinical trial PRODIGE 5/ACCORD 17 to analyse HRQoL data in the presence of dropouts unrelated and related to a clinical event. Then, we compared the models using a simulation study.

We showed that the CR JM performed as well as the SJM in situations where the risk of dropout was the same whatever the cause. However, the SJM estimations were biased in the presence of both informative and non-informative dropout, impacting the HRQoL estimated parameters, whereas the CR JM estimates were unbiased.

The systematic collection of the reasons for dropout in clinical trials would facilitate the use of CR JMs, which could be a satisfactory approach to analysing HRQoL data in presence of both informative and non-informative dropout.

**Trial registration:** This study is registered with ClinicalTrials.gov, number NCT00861094.

## **Keywords:**

Joint modelling, Competing risks, Health-related quality of life, Missing data, Clinical trials, Oncology.



## 1. Introduction

Health-related quality of life (HRQoL) is an important endpoint in cancer clinical trials. HRQoL is a patient-reported outcome used to investigate the clinical benefit of new treatment strategies in oncology. This endpoint, reflecting the patients' perception of their health status, is measured by self-administered questionnaires collected longitudinally throughout the care process. Linear mixed modelling is classically used to analyse HRQoL longitudinal data because it is a flexible and powerful approach to assessing the impact of the new treatment strategy on quality of life [1, 2]. Nevertheless, HRQoL assessment may be incomplete due to missing or incomplete questionnaires. High rates of missing HRQoL data are common in clinical trials, and most studies ignore this missing data or do not report the method used to handle them. Notably, when the disease and treatment are aggressive and toxic, monotone missing data may occur from dropout due to toxicity, disease progression or death. This form of dropout can be informative, leading to a biased estimate of the treatment effect [3]. Therefore, in order to produce valid results in a longitudinal analysis, it is crucial to consider the missing data mechanism.

The missing data mechanism can be classified into three types: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). According to this taxonomy, when the missing data are unrelated to the unobserved HRQoL outcome (MCAR, MAR), the missingness is non-informative and likelihood-based inference using all available data provides valid results. In contrast, when the missing data are related to the unobserved HRQoL outcome, the missingness is informative (MNAR). In order to avoid biased results when dropout is suspected to be informative, the longitudinal and the missingness processes must be jointly modelled [4–6].

Joint modelling has received much attention in various contexts and can be used to account for informative dropout when modelling a longitudinal outcome. The idea is to define two sub-models (a linear mixed model for the longitudinal outcome and a survival model for the time-to-dropout) linked through an association structure [7]. The association between the two processes is captured using a function of the shared random effects from the linear mixed model as a covariate in the survival model. A common choice of survival model is the Cox proportional hazards model.

However, a patient may drop out before the end of the clinical trial due to various competing risks. For example, dropouts due to a clinical event may be non-random, or informative, whereas dropouts due to other causes may be uninformative. In this situation, a standard joint model (SJM) that assumes the association between dropout and HRQoL to be the same whatever the causes of dropout is likely to be inappropriate. Such modelling can produce biased estimates of the HRQoL parameters, and therefore unreliable conclusions from the predicted trajectories of the mean HRQoL score in the different treatment arms and an incorrect understanding of the association between the HRQoL outcome and the dropout process.

An alternative, more satisfactory approach would be to use a time-to-dropout model that considers competing risks, allowing the study of multiple causes of dropout that have different associations with the longitudinal HRQoL outcome. Assuming different types of dropout may also shed new light on the association between dropout and HRQoL outcome. Recent developments have extended SJMs, enabling them to consider more complex data, such as considering multiple events for the time-to-event model [8, 9]. Although underused in clinical contexts, such models appear to be of particular interest to distinguish between different types of dropout [10].

The aim of this paper is to illustrate the potential of using a competing risks joint model (CR JM) to analyse HRQoL longitudinal data in the presence of different types of dropout and investigate the impact of using a SJM instead of a CR JM in various situations. Section 2 describes the two joint models. In Section 3, we apply the models to data from 267 patients with advanced oesophageal cancer from the randomized clinical trial PRODIGE 5/ACCORD 17. Section 4 presents a simulation study evaluating the performance of the two joint models according to various scenarios and highlights the situations where the SJM may not be suitable. Finally, we conclude with a discussion in Section 5.

## 2. Models

### 2.1 General presentation and notations

The joint models used consist of two sub-models: a linear mixed model for the continuous longitudinal HRQoL score and a model for the time-to-dropout, linked by an association structure involving common parameters. The sub-model for the longitudinal outcome is common to both joint models. The two processes are assumed to be conditionally independent given the random effects. In the SJM the dropout is considered the only possible event through a survival model. In the CR JM, we consider the dropout as a multiple event through a competing risks model to distinguish between two possible causes of dropout.

Let  $\{y_i(t), t \geq 0\}$  be the process for the HRQoL score of patient  $i$  at time  $t$ . We can observe this process as long as  $t < T_i$ , where  $T_i$  is the time-to-dropout or the right-censored time (in cases of no dropout).

## 2.2 Linear mixed sub-model

We assume that the observed longitudinal HRQoL score of patient  $i$  measured at time  $t$ ,  $y_i(t)$  is composed of the unobserved true value  $y_i^*(t)$  and a residual error term  $\varepsilon_i(t)$ . More specifically, we use a linear mixed model that includes fixed effects and random subject-specific effects to take into account the repeated measurements on the same patient, that is:

$$\begin{aligned} y_i(t) &= y_i^*(t) + \varepsilon_i(t) \\ &= X_i^\top(t)\beta + Z_i^\top(t)b_i + \varepsilon_i(t) \end{aligned} \quad (1)$$

where  $X_i(t)$  and  $Z_i(t)$  denote the design matrices of the fixed effects  $\beta$  and random effects  $b_i$ , respectively. The random effects are assumed to be normally distributed with mean zero and variance-covariance matrix  $D$ , independent of the measurement errors which are assumed to be independent and normally distributed with mean zero and variance  $\sigma^2$ .

In our context of clinical trials, we assumed a linear time effect of the HRQoL score which depends on the treatment arm and random intercept and slope; that is, the following linear mixed model:

$$y_i(t) = \beta_0 + \beta_1 t + \beta_2 \{arm_i \times t\} + b_{0i} + b_{1i} t + \varepsilon_i(t) \quad (2)$$

where  $arm_i$  is the arm indicator for patient  $i$  taking a value of 0 or 1 for the control or experimental treatment arm, respectively;  $\beta = (\beta_0, \beta_1, \beta_2)$  is the vector of fixed effects, with  $\beta_0$  the intercept that represents the mean score at inclusion ( $t = 0$ ),  $\beta_1$  the slope that represents the score change by unit of time in the control arm and  $\beta_2$  the interaction parameter that represents the difference between the slopes of the experimental and control arms; and  $b_i = (b_{0i}, b_{1i})$  is the vector of random effects, with  $b_{0i}$  the random intercept that represents the individual deviations from the fixed intercept,  $b_{1i}$  the random slope that represents the individual deviations from the fixed slope, and  $D = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}$ .

### 2.3 Standard joint model (SJM)

In the SJM, we use a proportional hazards model as the sub-model for the time-to-dropout, with a hazard function expressed as:

$$h_i(t) = h_0(t) \exp\{W_i^\top \gamma + \alpha y_i^*(t)\} \quad (3)$$

where  $h_0$  is the baseline hazard function,  $W_i$  denotes the vector of covariates with  $\gamma$  the corresponding vector of coefficients,  $\alpha$  is the parameter that characterizes the association between the risk of dropout and  $y_i^*(t)$  is the current true value of the HRQoL score. Note that if the association coefficient  $\alpha = 0$ , the two sub-models are no longer linked. Thus, an estimated association coefficient which is not significantly different from zero would traduce a non-informative dropout. In this model, the instantaneous risk of dropout is multiplied by the hazard ratio  $\exp(\alpha)$  for one unit increase of the current HRQoL score.

In our context of clinical trials, we included only the treatment arm as a covariate. The hazard function reduces to:

$$h_i(t) = h_0(t) \exp\{\gamma arm_i + \alpha y_i^*(t)\} \quad (4)$$

where the baseline hazard  $h_0$  is assumed to be parametric, and  $\gamma$  is the arm effect on the risk of dropout.

### 2.4 Competing risks joint model (CR JM)

In the CR JM, we use a competing risks model as the sub-model for the time-to-dropout to consider two possible causes of dropout. We use a proportional hazards model for each competing cause  $k$  of dropout,  $k = 1, 2$ :

$$h_{ik}(t) = h_{0k}(t) \exp\{\gamma_k arm_i + \alpha_k y_i^*(t)\} \quad (5)$$

where the baseline hazard functions  $h_{0k}$  are assumed to be parametric, and  $\gamma_k$  and  $\alpha_k$  denote the cause-specific arm effects and association parameters, respectively.

In our context of clinical trials, we considered two types of dropout: dropouts unrelated ( $k=1$ ) or related ( $k=2$ ) to a clinical event such as death or disease progression. The underlying idea is that the latter are likely to be informative, in contrast with the former.

### **3. Application**

#### *3.1 Data description*

The clinical trial PRODIGE 5/ACCORD 17 was a randomized phase 2/3 trial that included 267 patients with advanced oesophageal cancer and compared the chemoradiotherapy regimens of FOLFOX ( $n=134$ ) and fluorouracil–cisplatin ( $n=133$ ). The analysis for the primary efficacy endpoint of the phase 3 in the intent-to-treat population ( $n=267$ ) showed no difference in progression-free survival between the two treatment arms [11].

The secondary endpoints were overall survival, endoscopic complete response, time-to-treatment failure, occurrence of grade 3 or 4 toxicities and quality of life. HRQoL was assessed using the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ-C30 version 3.0) [12]. The QLQ-C30 is a self-administered questionnaire composed of 30 items evaluating five functional scales, nine symptomatic scales and the global health status/quality of life (GHS/QoL) scale. For each scale, a standardized score from 0 to 100 can be calculated according to the EORTC QLQ-C30 scoring manual [13]. A high score for the functional and GHS/QoL scales reflects high functional capacities and a good level of HRQoL. Conversely, a high score for the symptom scales reflects high levels of symptoms and a poor level of HRQoL. The questionnaire was completed at eight evaluations defined in the protocol: at inclusion, during treatment at months 1.25, 3 and 4 (first evaluation of treatment efficacy) and during follow-up at months 6, 12, 24 and 36.

Four dimensions of the QLQ-C30 were pre-specified in the protocol as targeted dimensions. First, the objective was to compare GHS/QoL levels (QL scale) between treatment arms; this was followed by comparisons of physical functioning (PF scale), pain (PA scale) and fatigue (FA scale). Analysis of the HRQoL scores showed no difference between the two regimens, despite the loss of power and possible biases due to reduced compliance in completing planned questionnaires over time [14].

### *3.2 Application of the joint models*

All analyses were performed in the evaluable intent-to-treat population: the evaluable population for a given scale included patients with at least one available HRQoL score from any measurement time. We analysed the four scales of interest included as secondary endpoints for the analysis of HRQoL, namely QL, PF, PA and FA. The evaluable intent-to-treat population reduces to 252 for the QL scale (experimental arm:  $n=130$ ; control arm:  $n=122$ ) and 254 for the PF, PA and FA scales (experimental arm:  $n=131$ ; control arm:  $n=123$ ).

The purpose of this application was to analyse the evolution of the HRQoL scores taking into account two possible causes of dropout and their association with HRQoL. For monotone missing scores, we calculated the time-to-dropout as the delay between inclusion and the last HRQoL evaluation.

In order to avoid computational issues, we considered the centred HRQoL score, calculated as the HRQoL score at time  $t$  minus the mean HRQoL score at baseline ( $t = 0$ ). We applied the two joint models described in Section 2 to each of the four scales of interest: first, we applied the SJM in Equation (4), for which dropout is considered the only possible event; then, we applied the CR JM in Equation (5) to distinguish between two causes of dropout – i.e., whether

the dropouts were unrelated (cause 1) or related (cause 2) to death or disease progression. In the two models, the longitudinal trajectories followed Equation (2).

Dropouts were attributed to cause 2 if death or progression occurred between the last evaluation visit with an available score and the subsequent planned visit, and attributed to cause 1 otherwise.

The estimates of the joint models were obtained by likelihood maximization using the R package JM (R software, version 3.4.0) [15, 16]. In both joint models, we used a cubic B-splines approximation for the log baseline hazard function with five internal knots placed at the quantiles of the observed event times (QL: 1.23, 2.98, 3.99, 9.43 and 16.14 months; versus PF, PA and FA: 1.35, 3.22, 4.07, 9.66 and 16.20 months). The integrals over the random effects were approximated using a nine-point adaptive Gauss–Hermite quadrature.

### 3.3 Results

Overall, 95% of patients (QL: 240/252; PF, PA and FA: 242/254) dropped out before the last planned questionnaire and 5% of patients (QL: 12/252; PF, PA, and FA: 12/254) completed the last planned questionnaire, generating right-censored times-to-dropout. Among the total dropouts, 57% (137/240 for QL and 139/242 for PF, PA and FA) were attributed to cause 1 and 43% (103/240 for QL and 103/242 for PF, PA and FA) were attributed to cause 2. The results of the SJM and CR JM are summarized in table 1.

#### *Results of the sub-model for the time-to-dropout*

Both models found that there was no effect of the treatment arm on the risk of dropout (non-significant  $\gamma$  for the SJM and non-significant  $\gamma_1$  and  $\gamma_2$  for the CR JM) for the four analysed dimensions. The SJM detected a significant association between the current HRQoL score and the risk of dropout for three scales: QL ( $\hat{\alpha} = -0.016$ ,  $p = 0.050$ ), PF ( $\hat{\alpha} = -0.019$ ,  $p = 0.001$ ) and FA ( $\hat{\alpha} = 0.013$ ,  $p = 0.007$ ), suggesting an informative dropout. For example, a



decrease of 10 points in the current PF score corresponded to an increase of  $\exp(10 \times 0.019) = 1.21$  (95% CI = 1.11 – 1.35) in the instantaneous risk of dropout. The CR JM found a significant association between the current HRQoL score and the risk of dropout only for the PF and FA scales, where only dropout due to cause 1 was associated with the current HRQoL score: PF ( $\hat{\alpha}_1 = -0.024$ ,  $p = 0.001$ ) and FA ( $\hat{\alpha}_1 = 0.016$ ,  $p = 0.011$ ). For example, a decrease of 10 points in the current PF score corresponded to an increase of  $\exp(10 \times 0.024) = 1.27$  (95% CI = 1.11 – 1.49) in the instantaneous risk of dropout unrelated to clinical event. In fact, these results suggest that for the PF and FA scales, dropout from cause 1 (unrelated to death or progression) was informative whereas dropout from cause 2 (related to death or progression) was non-informative. By contrast, the CR JM did not detect any association between the risk of dropout and the HRQoL for the QL or PA scales.

#### *Results of the sub-model for the longitudinal outcome*

Both the SJM and CR JM found a similar and significant increase in the score over time for the QL scale (SJM:  $\hat{\beta}_1 = 0.311$ ,  $p = 0.034$ ; CR JM:  $\hat{\beta}_1 = 0.313$ ,  $p = 0.031$ ) and a decrease in the score over time for the PF scale (SJM:  $\hat{\beta}_1 = -0.512$ ,  $p = 0.017$ ; CR JM:  $\hat{\beta}_1 = -0.505$ ,  $p = 0.017$ ). In addition, neither of the two models detected a significant difference between the HRQoL score trajectories of the two treatment arms.

#### **4. Simulation study**

We conducted a simulation study to evaluate the performance of the joint model that considers competing risks of dropout. We compared the estimated parameters of the CR JM with those of the SJM (which does not account for different causes of dropout) using bias and variance criteria derived from a series of 1000 generated datasets of 500 patients equally assigned to two treatment arms.

#### 4.1 Data generation

The data generation was based on the clinical trial PRODIGE 5/ACCORD 17 (see Section 3). We considered seven planned HRQoL assessment times  $t_j, j = 1, \dots, 7$ , at: inclusion ( $t_j = 0$ ), 1.5, 3, 6, 12, 24 and 36 months. For each patient  $i$ , the actual HRQoL assessment times  $t_{ij}$  were uniformly distributed around the planned ones (i.e., more or less a value proportional to the difference between the last and the next planned assessment):  $t_{i0} = 0$  and  $t_{ij} = t_j \pm x$  where  $x = k * (t_{j-1} - t_j) * U$  with  $U \sim U(-1, 1)$  a random variable distributed on  $(-1, 1)$  and  $k = 0.5$  to concentrate the simulated value  $t_{ij}$  around  $t_j$  without overlapping with other simulated values for the same patient  $i$ .

At each HRQoL assessment time for patient  $i$ ,  $t_{ij}$ , we generated the longitudinal outcome following the linear mixed model (1) with  $\beta_0 = 0, \beta_1 = -0.5, \beta_2 = 1, \sigma = 13, \sigma_0 = 12, \sigma_1 = 1.5$  and  $\rho = -0.3$  where  $\rho = \frac{\sigma_{01}}{\sigma_0 \sigma_1}$  is the correlation coefficient between the two random effects. For each patient  $i$ , we generated the observed time-to-dropout  $T_i$  such as  $T_i = \min(T_{1i}, T_{2i}, C_i)$  with  $T_{ki}$  the times-to-dropout for each cause  $k = 1, 2$  and  $C_i$  the censoring time. The censoring time  $C_i$  was  $t_{i7}$ , the last evaluation of patient  $i$ . The two cause-specific event times were generated under the CR JM (5) following the method developed to simulate complex survival data, such as time-to-event data from a joint model, and implemented in the R package `simsurv` [17, 18]. We assumed Weibull baseline cause-specific hazard functions expressed as  $h_{01}(t) = h_{02}(t) = \phi t^{\phi-1} \exp(\gamma_0)$  with  $\phi = 3$  (shape parameter) and  $\gamma_0 = -9$  (log of the scale parameter).

#### 4.2 Estimation

The estimates of the joint models were obtained by likelihood maximization using the R package JM (R software, version 3.4.0) [15, 16]. The log baseline hazards were approximated using a cubic B-splines with one internal knot placed at the median of the observed event times.

The integrals over the random effects were approximated using a five-point adaptive Gauss–Hermite quadrature.

#### 4.3 Evaluations criteria

For each parameter  $\theta$ , we calculated the following evaluations criteria: the mean estimated parameter  $\bar{\theta}$ , the relative bias  $(\bar{\theta} - \theta)/\theta$ , the root mean square error (RMSE)  $\sqrt{(\bar{\theta} - \theta)^2}$  and the coverage rate defined as the percentage of times where  $\theta$  falls within the confidence interval of  $\bar{\theta}$ .

#### 4.4 Scenarios

We considered five scenarios by varying the strength of the association  $\alpha_k$  and the intensity of the arm effect  $\gamma_k$  on the cause-specific risk of dropout,  $k=1, 2$ . In order to facilitate comparison between scenarios, the parameters describing the HRQoL trajectories remained unchanged under all scenarios.

[insert Fig.1]

**Fig. 1** Representation of the mean longitudinal outcome from the linear mixed sub-model (a) and hazard functions from the competing risks sub-model given the mean longitudinal outcome (b) used in the five scenarios of the simulation study where Scenario 1:  $\alpha_1 = -0.05$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = -0.70$ ,  $\gamma_2 = -0.70$ ; Scenario 2:  $\alpha_1 = 0.00$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = 0.00$ ,  $\gamma_2 = -0.70$ ; Scenario 3:  $\alpha_1 = 0.05$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = 0.00$ ,  $\gamma_2 = 0.00$ ; Scenario 4:  $\alpha_1 = 0.05$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = 0.70$ ,  $\gamma_2 = -0.70$ ; and Scenario 5:  $\alpha_1 = 0.00$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = 0.70$ ,  $\gamma_2 = -0.70$

Figure 1 shows the mean trajectories of the longitudinal HRQoL outcomes and the mean hazard functions from the CR JM used to generate the data in the five scenarios.

In Scenario 1, the two causes of dropout were equivalent. The cause-specific risks of dropout were equally and negatively associated with the current true HRQoL outcome, where a decrease of 10 points in the HRQoL outcome corresponded to an increase of 1.65 in the instantaneous risk of dropout ( $\alpha_1 = \alpha_2 = -0.05$ ).

The arm effect was equal whatever the cause and corresponded to lower risks in the experimental arm ( $\gamma_1 = \gamma_2 = -0.70$ ).

Scenario 2 was identical to Scenario 1 for cause 2 ( $\alpha_2 = -0.05$  and  $\gamma_2 = -0.70$ ), whereas for cause 1 the association was null ( $\alpha_1 = 0$ ), corresponding to non-informative dropout, and the risk of dropout was the same in both arms ( $\gamma_1 = 0$ ).

In Scenario 3, dropouts from causes 1 and 2 were considered to be informative, with cause-specific risks of dropout oppositely associated with the current true HRQoL outcome ( $\alpha_1 = 0.05$  and  $\alpha_2 = -0.05$ ). There was no arm effect on the risk of dropout whatever the cause ( $\gamma_1 = \gamma_2 = 0$ ).

In Scenario 4, the dropouts were considered to be inversely informative, as in Scenario 3 ( $\alpha_1 = 0.05$  and  $\alpha_2 = -0.05$ ). There were opposite arm effects on the risk of dropout for the two causes ( $\gamma_1 = 0.70$  and  $\gamma_2 = -0.70$ ).

In Scenario 5, dropout from cause 2 was also negatively associated with the current true HRQoL outcome ( $\alpha_2 = -0.05$ ) but dropout from cause 1 was treated as non-informative ( $\alpha_1 = 0$ ). Again, there were opposite arm effects on the risk of dropout for the two causes of dropout ( $\gamma_1 = 0.70$  and  $\gamma_2 = -0.70$ ).

#### 4.5 Simulation results

Table 2 describes dropout occurrence in the five different scenarios. In Scenario 1, we observed all planned measures of the longitudinal outcome (right-censored time-to-dropout) in 12.7% of the patients, whereas in the other scenarios all patients (or almost all) dropped out. Dropouts

due to causes 1 and 2 were equally distributed in Scenarios 1 and 3, and more unbalanced in Scenarios 2, 4 and 5, where more dropouts from cause 1 occurred. Table 3 provides the mean results of the 1000 simulations. Figure 2 gives a graphical representation of the distribution of the estimated  $\beta_1$ ,  $\beta_2$  and association parameters.

[insert Fig. 2]

**Fig. 2** Boxplots of the distributions of the association parameters,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  for the standard joint model (SJM) and the competing risks joint model (CR JM) in the presence of dropouts from two causes, C1 and C2, for the five scenarios considered in the simulation study

In Scenario 1, the parameters of the hazard functions were the same whatever the cause. Thus, the CR JM used to generate the data can reduce to a SJM with, in particular,  $\alpha = \alpha_1 = \alpha_2$  and  $\gamma = \gamma_1 = \gamma_2$ . The estimates of the hazard parameters were unbiased with both the SJM and CR JM ( $\bar{\alpha} = -0.050$ ,  $\bar{\alpha}_1 = -0.050$ ,  $\bar{\alpha}_2 = -0.050$  and  $\bar{\gamma} = -0.696$ ,  $\bar{\gamma}_1 = -0.691$ ,  $\bar{\gamma}_2 = -0.699$ ). The slope ( $\beta_1$ ) and interaction ( $\beta_2$ ) parameters of the longitudinal sub-model were also unbiased estimated and with coverage rates close to 95% for both the SJM and CR JM. The RMSE were similar in both models. Finally, in Scenario 1, for which a competing risks sub-model was not necessarily useful, both models performed similarly.

We will now comment on the results for Scenarios 2, 3, 4 and 5 obtained from the SJM for the association parameter, the parameter of the arm effect on the risk of dropout, and the longitudinal sub-model parameters. In these scenarios, the estimates from the CR JM were close to the simulated values, with a coverage rate close to 95%, and had smaller biases and RMSE than the SJM. In Scenarios 2, 3, 4 and 5, the SJM gave an estimate for the association parameter  $\alpha$  that was between the two simulated values for each cause,  $\alpha_1$  and  $\alpha_2$ . In Scenarios 3 and 4, where there were opposite cause-specific associations, the SJM estimated an association that was between  $\alpha_1$  and  $\alpha_2$ , suggesting no association with the longitudinal outcome. In Scenarios

2 and 5, where there was no association for cause 1 and a negative association of -0.05 for cause 2, the SJM estimated an association closer to  $\alpha_1$  than to  $\alpha_2$ :  $\bar{\alpha} = -0.017$  in Scenario 2 and  $\bar{\alpha} = 0.007$  in Scenario 4.

In a similar way to that of the association parameter, the SJM estimated a  $\gamma$  parameter between the simulated values for each cause,  $\gamma_1$  and  $\gamma_2$ . However, in all the scenarios where the arm had an effect on at least one cause of dropout – that is, in Scenarios 2, 4 and 5 – the estimated value was not in the middle of  $\gamma_1$  and  $\gamma_2$  but was closer to  $\gamma_1$ , even when the cause-specific effects were opposite. In the latter cases, the SJM suggested a risk of dropout that was a little more important in the experimental arm than in the control arm (Scenario 4:  $\bar{\gamma} = 0.152$ , Scenario 5:  $\bar{\gamma} = 0.213$ ), whereas this risk was much more important for dropouts due to cause 1 ( $\gamma_1 = 0.7$ ), and, in contrast, less important for dropouts due to cause 2 ( $\gamma_2 = -0.7$ ).

When the effect parameters of the time-to-dropout sub-model differed according to the dropout causes of dropout, SJM tended to incorrectly estimate the estimations of the longitudinal sub-model ( $\beta_1$  and  $\beta_2$ ).

Using the SJM, the relative bias for  $\beta_1$  varied from -11.5% in Scenario 3 to -18.9% in Scenario 4, and for  $\beta_2$  from -8.3% in Scenario 2 to -24.6% in Scenario 4. By comparison, the maximal relative bias using the CR JM was -0.5% (for both  $\beta_1$  and  $\beta_2$ ) in Scenario 3. In the worst case – that is, in Scenario 4 – the coverage rates from the SJM were 87% for  $\beta_1$  and 71% for  $\beta_2$ . The bias of  $\beta_1$  corresponded in Scenarios 2, 3, 4 and 5 to an overestimation of the HRQoL level in the control arm and the bias for  $\beta_2$  to an underestimation of the HRQoL improvement in the experimental arm compared with the control arm. The consequence of the misestimation of  $\beta_1$  and  $\beta_2$  on the predicted HRQoL score trajectories in the two arms can be appreciated in figure 3.

[insert Fig. 3]

**Fig. 3** Simulated trajectories of the health-related quality of life (HRQoL) score and predicted trajectories from the standard joint model (SJM) and competing risks joint model (CR JM) by treatment arm for the five scenarios considered in the simulation study where the CR JM-predicted trajectories (blue lines) are superimposed on the simulated trajectories (black lines)

The SJM-predicted trajectory is above the simulated trajectory in the control arm and below the simulated trajectory in the experimental arm. The CR JM-predicted trajectories are superimposed on the simulated trajectories.

## 5. Discussion

This article focuses on the analysis of longitudinal HRQoL data in clinical trials where patients are likely to drop out during treatment or follow-up. We compared the use in these situations of a CR JM that distinguishes between two causes of dropout with a SJM that treats all the dropouts equally.

In the simulation study, we showed that the CR JM performed as well as the SJM in situations where the risk of dropout was the same whatever the cause. However, in situations where two different risks of dropout coexisted, the SJM provided biased estimates. The estimates of the effect parameter on the risk of dropout were between the values of the cause-specific parameters. In particular, when the two types of dropout were informative and oppositely associated with the current true HRQoL outcome, the SJM estimated a null association parameter, suggesting non-informative dropout. The misspecification of the time-to-dropout sub-model that did not distinguish between the dropouts resulted in biased estimates of the HRQoL parameters. In particular, the difference between the slopes of the HRQoL trajectories of the two arms was underestimated.

In our application, we found the same pattern as in the simulation study: the parameters estimated by the SJM were between the two cause-specific parameters estimated by the CR JM. The CR JM captured supplementary information about the dropout and revealed that only one cause of dropout was associated with the current HRQoL score. Unexpectedly, it was not dropout related to death or progression that was associated with the HRQoL outcome. This may be due to a lack of reliability concerning the cause of dropout. Indeed, the cause of dropout was constructed *a posteriori*, since the protocol did not include collecting the reasons for dropout. Another explanation is that a non-negligible proportion of dropout unrelated to death or progression could actually be informative. Death and progression are generally well documented, but other events, not documented, could be related to HRQoL. The systematic collection of the reasons for missing HRQoL data in clinical trials would allow a better attribution of the cause of dropout and facilitate the use and the interpretation of CR JMs. Recommendations exist to limit the occurrence of missing data and some authors encourage documenting why the data are missing [19, 20]. This could guide the choice of analysis strategy while only 50% of clinical trials use adequate methods for handling missing data [21, 22].

In conclusion, the competing risks joint modelling is a satisfactory approach to distinguishing between informative and non-informative dropouts in order to obtain valid results for the longitudinal analysis of HRQoL in clinical trials. Moreover, it offers a better understanding of the association between the HRQoL outcome and the dropout process than a standard joint modelling. Nevertheless, the cause of dropout can be unknown or unreliable in practice. Collecting the reasons for missing data in clinical trials would allow a better attribution of the cause of dropout and facilitate the use and interpretation of competing risks joint models.



## References

- [1] Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; 963–974.
- [2] Chen H, Cohen P. Using individual growth model to analyze the change in quality of life from adolescence to adulthood. *Health and Quality of Life Outcomes* 2006; 4: 10.
- [3] Bell ML, Kenward MG, Fairclough DL, et al. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *Bmj* 2013; 346: e8668.
- [4] Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; 1: 465–480.
- [5] Ibrahim JG, Chu H, Chen LM. Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. *Journal of Clinical Oncology* 2010; 28: 2796–2801.
- [6] Cuer B, Mollevi C, Anota A, et al. Handling informative dropout in longitudinal analysis of health-related quality of life: application of three approaches to data from the esophageal cancer clinical trial PRODIGE 5/ACCORD 17. *BMC Medical Research Methodology* 2020; 20: 223.
- [7] Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; 809–834.
- [8] Elashoff R, Li G, Li N. *Joint Modeling of Longitudinal and Time-to-Event Data*. CRC Press, 2016.
- [9] Williamson PR, Kolamunnage-Dona R, Philipson P, et al. Joint modelling of longitudinal and competing risks data. *Statistics in Medicine* 2008; 27: 6426–6438.
- [10] Kolamunnage-Dona R, Powell C, Williamson PR. Modelling variable dropout in randomised controlled trials with longitudinal outcomes: application to the MAGNETIC study. *Trials* 2016; 17: 222.
- [11] Conroy T, Galais M-P, Raoul J-L, et al. Definitive chemoradiotherapy with FOLFOX versus fluorouracil and cisplatin in patients with oesophageal cancer (PRODIGE5/ACCORD17): final results of a randomised, phase 2/3 trial. *The Lancet Oncology* 2014; 15: 305–314.
- [12] Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology. *JNCI: Journal of the National Cancer Institute* 1993; 85: 365–376.
- [13] Fayers P, Aaronson NK, Bjordal K, et al. *EORTC QLQ-C30 scoring manual*. 3rd ed. European Organisation for Research and Treatment of Cancer, 2001.
- [14] Bascoul-Mollevi C, Gourgou S, Galais M-P, et al. Health-related quality of life results from the PRODIGE 5/ACCORD 17 randomised trial of FOLFOX versus fluorouracil–cisplatin regimen in oesophageal cancer. *European Journal of Cancer* 2017; 84: 239–249.

- [15] Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software, Articles* 2010; 35: 1–33.
- [16] Rizopoulos D. *Joint models for longitudinal and time-to-event data: With applications in R*. Chapman and Hall/CRC, 2012.
- [17] Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. *Statist Med* 2013; 32: 4118–4134.
- [18] Crowther MJ, Lambert PC. Simulating complex survival data. *The Stata Journal* 2012; 12: 674–687.
- [19] Little RJ, D’agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine* 2012; 367: 1355–1360.
- [20] Calvert M, Blazeby J, Altman DG, et al. Reporting of Patient-Reported Outcomes in Randomized Trials: The CONSORT PRO Extension. *JAMA* 2013; 309: 814–822.
- [21] Stockler MR, Hilpert F, Friedlander M, et al. Patient-reported outcome results from the open-label phase III AURELIA trial evaluating bevacizumab-containing therapy for platinum-resistant ovarian cancer. *Journal of Clinical Oncology* 2014; 32: 1309.
- [22] Powney M, Williamson P, Kirkham J, et al. A review of the handling of missing longitudinal outcome data in clinical trials. *Trials* 2014; 15: 237.

## Article tables

Table 1: Results of the longitudinal analysis of the HRQoL scores centred on the four dimensions of interest of the EORTC QLQ-C30, taking into account dropout using the standard joint model (SJM) and the competing risks joint model (CR JM) that distinguishes between dropouts unrelated (cause 1) or related (cause 2) to a clinical event.

		SJM		CR JM	
		Estimate [CI 95%]	<i>p</i>	Estimate [CI 95%]	<i>p</i>
<b>QL (n=252)</b>					
Linear mixed sub-model for the longitudinal outcome					
	$\beta_1$	0.311 [0.02; 0.60]	0.034	0.313 [0.03; 0.60]	0.031
	$\beta_2$	0.166 [-0.21; 0.54]	0.383	0.170 [-0.20; 0.54]	0.363
Survival sub-model for the time-to-dropout					
	$\gamma$	-0.019 [-0.28; 0.24]	0.884		
	$\gamma_1$			-0.050 [-0.39; 0.29]	0.772
	$\gamma_2$			0.079 [-0.43; 0.59]	0.763
	$\alpha$	-0.016 [-0.03; 0.00]	0.050		
	$\alpha_1$			-0.016 [-0.04; 0.00]	0.126
	$\alpha_2$			0.001 [-0.03; 0.03]	0.951
<b>PF (n=254)</b>					
Linear mixed sub-model for the longitudinal outcome					
	$\beta_1$	-0.512 [-0.93; -0.09]	0.017	-0.505 [-0.92; -0.09]	0.017
	$\beta_2$	0.078 [-0.41; 0.56]	0.753	0.075 [-0.41; 0.56]	0.763
Survival sub-model for the time-to-dropout					
	$\gamma$	-0.050 [-0.31; 0.21]	0.699		
	$\gamma_1$			-0.087 [-0.42; 0.25]	0.614
	$\gamma_2$			0.083 [-0.43; 0.59]	0.751
	$\alpha$	-0.019 [-0.03; -0.01]	0.001		
	$\alpha_1$			-0.024 [-0.04; -0.01]	0.001
	$\alpha_2$			0.012 [-0.01; 0.03]	0.258
<b>PA (n=254)</b>					
Linear mixed sub-model for the longitudinal outcome					
	$\beta_1$	-0.214 [-0.58; 0.15]	0.247	-0.207 [-0.57; 0.16]	0.262
	$\beta_2$	-0.210 [-0.69; 0.27]	0.392	-0.237 [-0.72; 0.24]	0.334
Survival sub-model for the time-to-dropout					
	$\gamma$	-0.069 [-0.32; 0.19]	0.599		
	$\gamma_1$			-0.087 [-0.42; 0.25]	0.610
	$\gamma_2$			0.071 [-0.44; 0.59]	0.787
	$\alpha$	0.008 [0.00; 0.02]	0.185		
	$\alpha_1$			0.008 [-0.01; 0.02]	0.342
	$\alpha_2$			0.000 [-0.02; 0.02]	0.992
<b>FA (n=254)</b>					
Linear mixed sub-model for the longitudinal outcome					
	$\beta_1$	-0.223 [-0.56; 0.12]	0.200	-0.222 [-0.57; 0.13]	0.215
	$\beta_2$	-0.235 [-0.70; 0.22]	0.316	-0.238 [-0.70; 0.22]	0.313
Survival sub-model for the time-to-dropout					
	$\gamma$	-0.033 [-0.29; 0.22]	0.797		
	$\gamma_1$			-0.068 [-0.40; 0.27]	0.693
	$\gamma_2$			0.073 [-0.44; 0.58]	0.778
	$\alpha$	0.013 [0.00; 0.02]	0.007		
	$\alpha_1$			0.016 [0.00; 0.03]	0.011
	$\alpha_2$			-0.007 [-0.03; 0.01]	0.471

CI, confidence interval; EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire; FA, fatigue; PA, pain; PF, physical functioning; QL, global health status/quality of life

Table 2: Mean event rates (%) in the 1000 simulated data sets.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
No dropout, %	12.70	0.18	0.00	0.00	0.04
Dropout from cause 1, %	43.71	57.59	49.96	59.00	62.92
Dropout from cause 2, %	43.59	42.23	50.04	40.99	37.04

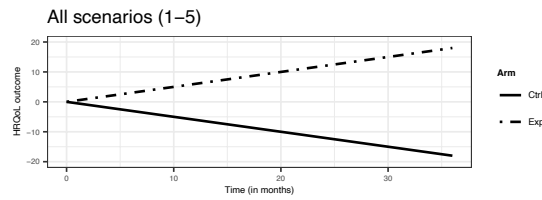
Table 3: Results on simulated data based on 1000 replications per scenario according to the standard joint model (SJM) and the competing risks joint model (CR JM). For each parameter: true value (Value), mean estimate (Mean), relative bias (Rel. bias) in percentage, root mean square error (RMSE) and 95% coverage rate (Cov. rate) in percentage.

Scenario		SJM					CR JM			
		Value	Mean	Rel. Bias (%)	RMSE	Cov. Rate (%)	Mean	Rel. bias (%)	RMSE	Cov. Rate (%)
1	$\beta_1$	-0.5	-0.494	-1.1	0.147	94	-0.497	-0.6	0.147	94
	$\beta_2$	1	1.000	0.0	0.167	96	0.999	-0.1	0.167	96
	$\alpha$	-	-0.050	0.7	0.005	95	-	-	-	-
	$\alpha_1$	-0.05	-	-	-	-	-0.050	0.0	0.007	94
	$\alpha_2$	-0.05	-	-	-	-	-0.050	-0.3	0.006	94
	$\gamma$	-	-0.696	-0.6	0.126	94	-	-	-	-
	$\gamma_1$	-0.7	-	-	-	-	-0.691	-1.3	0.161	95
	$\gamma_2$	-0.7	-	-	-	-	-0.699	-0.1	0.159	95
2	$\beta_1$	-0.5	-0.434	-13.3	0.160	92	-0.499	-0.3	0.144	94
	$\beta_2$	1	0.917	-8.3	0.194	93	0.999	-0.1	0.177	95
	$\alpha$	-	-0.017	-	-	-	-	-	-	-
	$\alpha_1$	0	-	-	-	-	0.000	0.0	0.004	95
	$\alpha_2$	-0.05	-	-	-	-	-0.050	0.2	0.007	96
	$\gamma$	-	-0.298	-	-	-	-	-	-	-
	$\gamma_1$	0	-	-	-	-	0.004	0.4	0.132	94
	$\gamma_2$	-0.7	-	-	-	-	-0.699	-0.2	0.174	94
3	$\beta_1$	-0.5	-0.443	-11.5	0.157	92	-0.497	-0.5	0.148	94
	$\beta_2$	1	0.886	-11.4	0.215	89	0.995	-0.5	0.184	96
	$\alpha$	-	0.000	-	-	-	-	-	-	-
	$\alpha_1$	0.05	-	-	-	-	0.051	2.6	0.007	96
	$\alpha_2$	-0.05	-	-	-	-	-0.051	2.7	0.008	96
	$\gamma$	-	0.005	0.5	0.101	94	-	-	-	-
	$\gamma_1$	0	-	-	-	-	0.003	0.3	0.165	94
	$\gamma_2$	0	-	-	-	-	0.009	0.9	0.162	95
4	$\beta_1$	-0.5	-0.406	-18.9	0.178	87	-0.498	-0.4	0.149	93
	$\beta_2$	1	0.754	-24.6	0.319	71	0.997	-0.3	0.194	95
	$\alpha$	-	0.007	-	-	-	-	-	-	-
	$\alpha_1$	0.05	-	-	-	-	0.051	2.2	0.007	95
	$\alpha_2$	-0.05	-	-	-	-	-0.051	2.7	0.008	95
	$\gamma$	-	0.152	-	-	-	-	-	-	-
	$\gamma_1$	0.7	-	-	-	-	0.709	1.3	0.153	94
	$\gamma_2$	-0.7	-	-	-	-	-0.702	0.2	0.182	95
5	$\beta_1$	-0.5	-0.424	-15.1	0.165	91	-0.498	-0.4	0.146	94
	$\beta_2$	1	0.886	-11.4	0.218	90	0.998	-0.2	0.185	95
	$\alpha$	-	-0.016	-	-	-	-	-	-	-
	$\alpha_1$	0	-	-	-	-	0.000	0.0	0.004	95
	$\alpha_2$	-0.05	-	-	-	-	-0.050	0.1	0.008	94
	$\gamma$	-	0.213	-	-	-	-	-	-	-
	$\gamma_1$	0.7	-	-	-	-	0.713	1.8	0.126	94
	$\gamma_2$	-0.7	-	-	-	-	-0.715	2.2	0.200	95

## Article figures

Figure 1: Representation of the mean longitudinal outcome from the linear mixed sub-model (a) and hazard functions from the competing risks sub-model given the mean longitudinal outcome (b) used in the five scenarios of the simulation study. Scenario 1:  $\alpha_1 = -0.05$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = -0.70$ ,  $\gamma_2 = -0.70$ ; Scenario 2:  $\alpha_1 = 0.00$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = 0.00$ ,  $\gamma_2 = -0.70$ ; Scenario 3:  $\alpha_1 = 0.05$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = 0.00$ ,  $\gamma_2 = 0.00$ ; Scenario 4:  $\alpha_1 = 0.05$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = 0.70$ ,  $\gamma_2 = -0.70$ ; and Scenario 5:  $\alpha_1 = 0.00$ ,  $\alpha_2 = -0.05$ ,  $\gamma_1 = 0.70$ ,  $\gamma_2 = -0.70$ .

### a. Longitudinal sub-model



### b. Competing risks sub-model

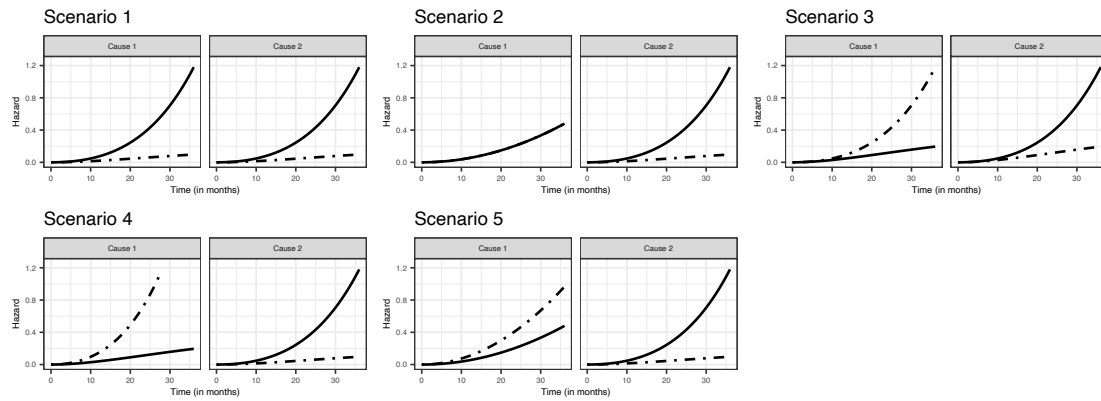


Figure 2: Boxplots of the distributions of the association parameters,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  for the standard joint model (SJM) and the competing risks joint model (CR JM) in the presence of dropouts from two causes, C1 and C2, for the five scenarios considered in the simulation study.

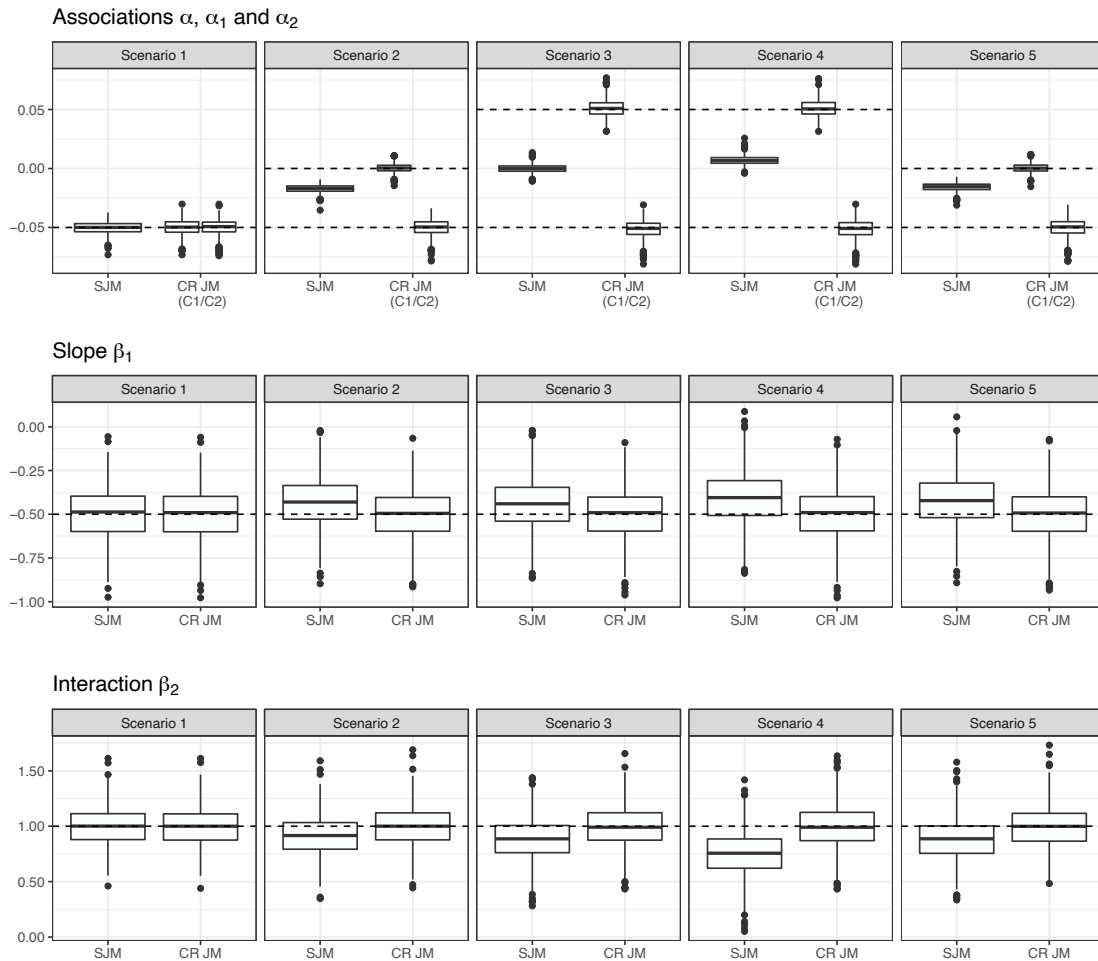
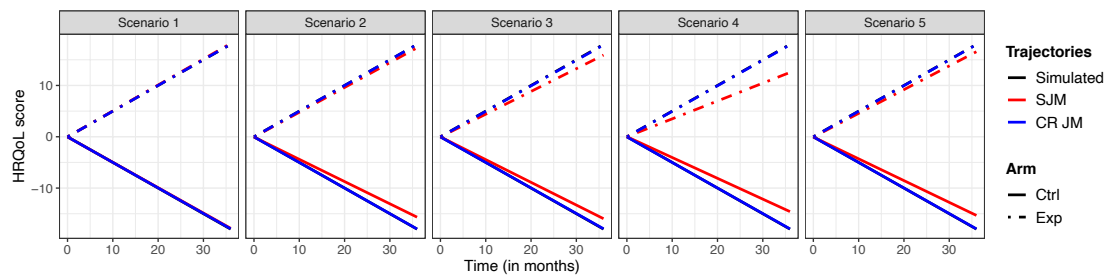


Figure 3: Simulated trajectories of the health-related quality of life (HRQoL) score and predicted trajectories from the standard joint model (SJM) and competing risks joint model (CR JM) by treatment arm for the five scenarios considered in the simulation study. The CR JM-predicted trajectories (blue lines) are superimposed on the simulated trajectories (black lines).





## Chapitre 4. Modèles conjoints différenciant plusieurs causes de sorties d'étude

**Tableau 4.1** – Attribution des causes de sorties d'étude (cause 1 : non liée à un évènement clinique, cause 2 : liée à un évènement clinique) en fonction de l'apparition de l'évènement clinique pendant l'évaluation de la QdV (O : observé, X : manquant).

	Visite Inclusion	Visite 1.25 mois	Visite 3 mois	Visite 4 mois	Visite 6 mois	Visite 12 mois	Visite 24 mois	Visite 36 mois	Evènement Clinique	Cause
Cas 1	O	O	O	O	O	X	X	X	Décès à $t = 26$	1
Cas 2	O	O	O	O	O	X	X	X	Décès à $t = 10$	2
Cas 3	O	O	O	O	O	X	X	X	Progression à $t = 6$	2
Cas 4	O	O	O	O	O	X	X	X	Décès à $t = 13$	2

### 4.3 Discussion et conclusion

Nous avons analysé les données de QdV de l'essai clinique PRODIGE 5/ACCORD 17 pour 4 dimensions du QLQ-C30 par un modèle conjoint standard et un modèle conjoint à risques concurrents. Nous avons également mené une étude de simulations afin d'étudier la performance de ces modèles selon différents scénarios et ainsi identifier les situations où le modèle conjoint standard semble moins performant. Les estimations des paramètres caractérisant la trajectoire du score de QdV dans les deux groupes de traitement ont été comparées entre les deux modèles et interprétées en relation avec les estimations des paramètres caractérisant les risques de sorties d'étude.

Grâce aux données qui ont été collectées pendant l'essai, nous avons considéré deux causes de sorties d'étude en fonction qu'elles soient liées ou non à un évènement clinique. Lorsque la progression de la maladie ou le décès précèdent l'évaluation manquante de la QdV, nous avons désigné les sorties d'étude comme étant « liées à un évènement clinique » (cause 2) alors que les autres sorties d'étude ont été désignées comme étant « non liées à un évènement clinique » (cause 1). Comme illustré dans le Tableau 4.1, nous avons fait l'hypothèse que les raisons de sorties d'étude sont attribuées à un évènement clinique (progression de la maladie ou décès) lorsque celui-ci survient entre la dernière évaluation disponible de QdV et la suivante prévue par le protocole (dans le cas d'un décès, cas 2) ou lorsque celui-ci survient à la dernière évaluation disponible (dans le cas d'une progression, cas 3). Enfin, lorsqu'un patient ne répond pas au questionnaire de QdV à la visite prévue par le protocole, puis décède dans les jours qui suivent (cas 4), la sortie d'étude est alors également attribuée à la cause 2.

L'étude de simulations que nous avons réalisée nous a permis de conclure que le modèle conjoint à risques concurrents est aussi performant (pour ce qui est du biais et de la variance) que le modèle conjoint standard lorsqu'il n'est pas nécessaire de considérer deux causes de sorties d'étude (scénario 1 de l'étude de simulations). Plus précisément, l'étude de simulations montre que les deux modèles conjoints estiment sans biais, à la fois les paramètres du sous-modèle du temps jusqu'à sortie d'étude et du sous-modèle pour le score mesuré au cours du temps. L'utilisation d'un modèle conjoint à risques concurrents n'apporte pas d'information supplémentaire en comparaison avec le modèle conjoint standard avec des sorties d'études non informatives quelle que soit la cause. Par conséquent, l'évolution du score de QdV est sensiblement la même dans ces deux modèles. Sur les dimensions QL et PA de l'essai clinique PRODIGE 5/ACCORD 17, où les paramètres d'association ne sont significatifs pour aucune des causes révélant des sorties d'étude non informatives, les deux modèles conjoints donnent des résultats similaires sur les paramètres de QdV.

Lorsqu'il existe différentes associations de la QdV avec le risque de sortie d'étude en fonction de la cause, le modèle conjoint à risques concurrents semble être un modèle performant. En effet, en présence d'associations positive et négative (scénarios 3 et 4 de l'étude de simulations) ou nulle et négative (scénarios 2 et 5 de l'étude de simulations), ce modèle donne des estimations sans biais sur les paramètres d'association. Le modèle conjoint standard estime un seul paramètre d'association qui se situe de façon attendue entre les associations relatives à chaque cause. Le modèle conjoint à risques concurrents permet de faire la distinction entre les causes, là où le modèle conjoint standard ne le permet pas. Ensuite, l'étude de simulations montre que les paramètres du sous-modèle pour les trajectoires de score sont biaisés dans le modèle conjoint standard du fait de la mauvaise prise en compte de l'association entre les deux sous-modèles. Dans les scénarios considérés, le modèle conjoint standard sous-estime le paramètre de pente dans le bras contrôle et surestime le paramètre de pente dans le bras expérimental. Enfin, lorsque les sorties d'étude sont associées à la QdV de manière opposée (scénarios 3 et 4 de l'étude de simulations), le modèle conjoint standard suggère des sorties d'étude non informatives en estimant une association proche de zéro entre les deux sous-modèles (phénomène de « compensation » des deux causes). C'est dans cette dernière situation que les biais sur les trajectoires de QdV sont les plus importants,

## **Chapitre 4. Modèles conjoints différenciant plusieurs causes de sorties d'étude**

en particulier dans le scénario 4 où il y a aussi un effet opposé du bras sur les risques de sortie d'étude.

Finalement, ce travail met en évidence qu'il peut être important que la prise en compte de l'association en QdV et sortie d'étude dépende de la cause de sortie d'étude au risque d'induire des biais sur le sous-modèle pour le score de QdV et donc sur les trajectoires de QdV. Le modèle conjoint à risques concurrents permet d'obtenir des estimations sans biais pour les deux sous-modèles en comparaison au modèle conjoint standard. De plus, le modèle conjoint à risques concurrents permet de décrire avec précision les risques de sorties d'étude en permettant de distinguer des causes informatives et non informatives, et en quantifiant le degré d'association d'une augmentation ou d'une diminution de la QdV avec le risque de sortir d'étude, en fonction de la cause considérée. Dans les données de QdV de l'essai clinique PRODIGE 5/ACCORD 17, aucune association significative entre les sorties d'étude et le score courant de QdV n'a été détectée lorsque les sorties d'étude sont liées à un événement clinique. Ce constat est assez étonnant puisque ce sont généralement des patients dans un état de santé dégradé pour lesquels nous suspectons des sorties d'études informatives. Pourtant, l'étude de simulations a bien montré que le modèle conjoint à risques concurrents était capable de distinguer des sorties d'étude informatives et non informatives. En fait, malgré l'attribution des causes de sorties d'étude telles que décrites ci-dessus, il reste difficile de bien distinguer les patients qui sortent d'étude en lien avec la dégradation de leur état de santé. Dans l'essai clinique PRODIGE 5/ACCORD 17, aucune information supplémentaire n'a été collectée pour comprendre les raisons de sorties d'étude. Les événements cliniques comme la progression ou le décès étaient en revanche bien documentés, comme c'est généralement le cas. Nous sommes donc partis de l'*a priori* selon lequel les sorties d'études précédant un événement clinique seraient probablement informatives, mais il est possible qu'une part non négligeable des sorties d'étude *a priori* considérées comme non informatives puissent l'être : par exemple, la décision du patient de ne plus répondre au questionnaire ou bien le choix de l'équipe médicale de ne plus transmettre de questionnaire à un patient trop affaibli par la maladie. Ces décisions peuvent être en réalité la conséquence d'une détérioration de la QdV, et plus généralement de l'état de santé du patient, même si aucune progression n'a été observée. Ainsi, le recueil systématique des raisons de non-remplissage des questionnaires permettrait une meilleure

compréhension et attribution de leur cause et faciliterait ainsi l'utilisation des modèles conjoints à risques concurrents.

Dans ce chapitre, nous nous sommes concentrés sur la modélisation et la prise en compte des sorties d'étude par un modèle conjoint à risques concurrents. Plus précisément, ce travail s'est consacré à l'utilisation d'un sous-modèle à risques concurrents pour le temps jusqu'à sortie d'étude. Nous avons vu que ce modèle conjoint pouvait être une alternative intéressante par rapport à un modèle conjoint standard et qu'il y avait des avantages à utiliser un tel modèle tant au niveau de l'interprétation que de l'estimation des paramètres. Après avoir exploré dans le cadre d'une modélisation conjointe du score de QdV et du temps jusqu'à sortie d'étude une alternative au sous-modèle pour le temps jusqu'à sortie d'étude, nous allons investiguer dans le chapitre suivant des alternatives au sous-modèle du score de QdV.



# Chapitre 5

## **Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude**

---

L'objectif de ce chapitre est de proposer des alternatives de modélisation dans le cadre d'une modélisation conjointe des scores de QdV et des sorties d'étude. Dans un premier temps, nous modéliserons les trajectoires de QdV par des évolutions non linéaires, puis dans un second temps, nous explorerons la modélisation du score de QdV en tant que variable ordinale plutôt que continue.

### **5.1 Prise en compte des trajectoires non linéaires**

#### **5.1.1 Introduction**

Pour modéliser les trajectoires de QdV en présence de sorties d'étude, nous avons vu que les modèles conjoints étaient particulièrement adaptés. Nous avons proposé d'utiliser un modèle à risques concurrents pour décrire les sorties d'étude afin de mieux décrire le processus de sorties d'étude et par conséquent mieux tenir

## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

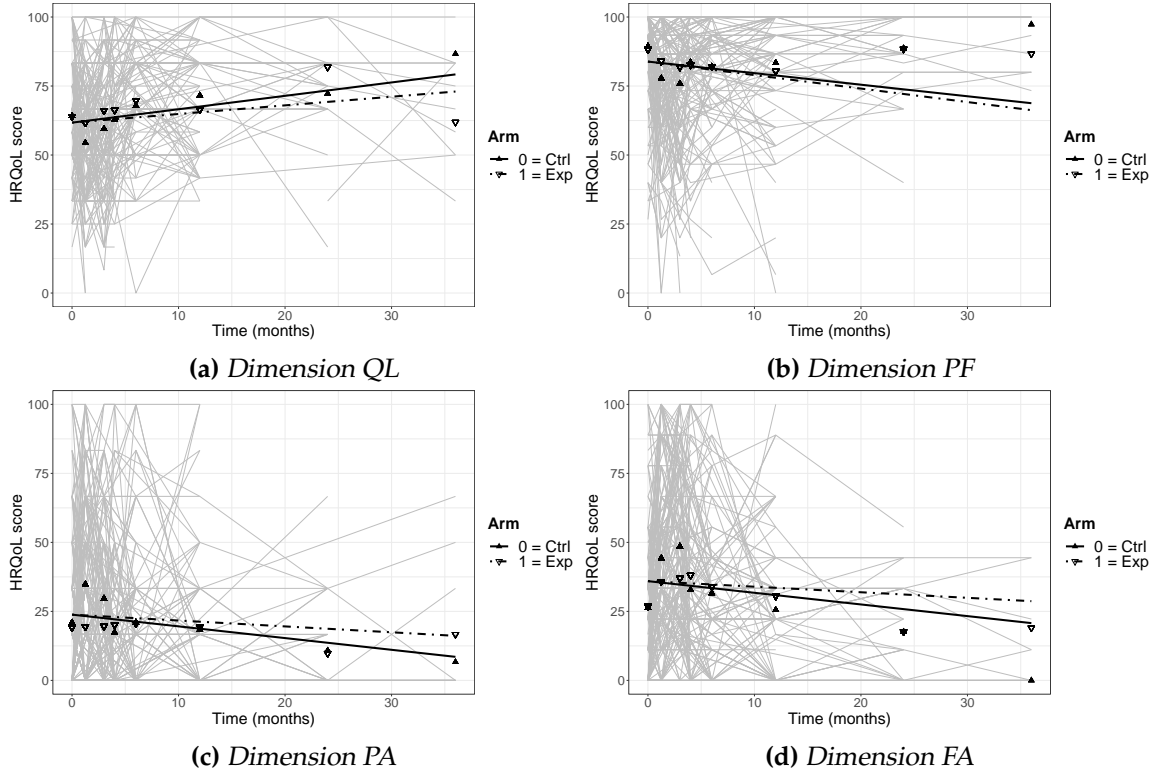
---

compte du mécanisme qui en découle (voir Chapitre 4). Nous avons jusqu'alors toujours considéré un modèle conjoint avec des évolutions linéaires, en utilisant le sous-LMM tel que décrit dans l'équation (1.3). Cependant, il serait intéressant d'étudier une modélisation plus adaptée permettant de modéliser des trajectoires non linéaires du score de QdV.

Dans le cas des données de l'essai clinique PRODIGE 5/ACCORD 17, et comme en témoigne la Figure 5.1, nous remarquons que les trajectoires individuelles (en gris) et les scores moyens (en noir) n'évoluent pas de façon linéaire, aussi bien pour le statut global de santé et les capacités physiques (QL, PF) que pour les dimensions symptomatiques (PA, FA). En effet, les trajectoires de QdV semblent montrer une diminution de la QdV lorsque les patients sont sous traitement (diminution des capacités physique ou du statut global de santé et une augmentation des symptômes) alors qu'on observe une augmentation de la QdV après la fin du traitement (augmentation des capacités physique ou du statut global de santé et une diminution des symptômes). Par conséquent, l'hypothèse de linéarité des trajectoires du sous-LMM des scores de QdV dans le modèle conjoint (en noir) ne semble pas la plus optimale. Une solution alternative consiste à envisager des modèles beaucoup plus flexibles basés sur des évolutions non linéaires en utilisant un polynôme d'ordre supérieur (e.g. effet quadratique) ou par des splines ([Perperoglou et al., 2019](#)). Ces approches permettent de capturer des possibles évolutions non linéaires, mais il faut veiller à l'identifiabilité des modèles lorsque le nombre de paramètres à estimer augmente ([Fairclough, 2010](#)). Généralement, les splines sont plus utilisées en raison de leur flexibilité et de certaines de leurs propriétés locales naturelles et numériques que les polynômes n'ont pas ([Ruppert et al., 2003](#)), même si les résultats peuvent être plus difficiles à interpréter cliniquement. Grâce aux développements pour approximer numériquement et plus rapidement les intégrales dans l'écriture de la vraisemblance d'un modèle conjoint ([Rizopoulos, 2012a](#)), il a été possible d'utiliser des modèles plus flexibles pour capturer des trajectoires avec des évolutions non linéaires par des approches basées sur les splines naturelles ou les B-splines ([Ding and Wang, 2008](#); [Brown et al., 2005](#); [Rizopoulos and Ghosh, 2011](#)), là où une trajectoire linéaire serait trop restrictive.

Afin de continuer de prendre en compte la problématique des sorties d'études informatives, nous allons utiliser un modèle conjoint combinant un sous-modèle

## 5.1. Prise en compte des trajectoires non linéaires



**Figure 5.1** – Trajectoires individuelles (en gris), moyennes observées des scores et trajectoires prédites par le modèle conjoint linéaire dans le bras contrôle (symbole triangulaire vers le haut et courbe pleine) et le bras expérimental (symbole triangulaire vers le bas et courbe en pointillé).

de survie pour le temps jusqu'à sortie d'étude et un sous-modèle pour les trajectoires non linéaires du score de QdV. Ce dernier semble mieux adapté pour prendre en compte les évolutions du score par rapport à un sous-modèle avec des trajectoires linéaires. L'objectif est alors de comparer les deux modèles conjoints, avec des trajectoires linéaires ou non linéaires pour les scores de QdV.

### 5.1.2 Modèles conjoints avec des splines

Pour une population de taille  $n$ , soit  $Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{in_i}))^\top$  le vecteur qui contient l'ensemble des  $n_i$  mesures répétées du patient  $i$  ( $i = 1, \dots, n$ ) où le score de QdV prend la valeur  $Y_i(t_{ij})$  au temps  $t_{ij}$  ( $j = 1, \dots, n_i$ ). La variable de score de QdV est supposée continue et distribuée selon une distribution gaussienne. En



## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

---

considérant un sous-modèle d'évolutions linéaires ou non linéaires, cela nous permet de définir les deux différents modèles conjoints appelés respectivement JM linéaire et JM splines. Le modèle conjoint JM linéaire est le modèle qui a été utilisé jusqu'à maintenant, composé d'un sous-modèle linéaire mixte à ordonnée et pente aléatoires (voir Figure 5.1).

### Modèle conjoint avec sous-modèle d'évolutions non linéaires (JM splines)

Pour modéliser les trajectoires non linéaires des scores de QdV, nous utilisons un modèle à splines naturelles cubiques pour les effets fixes et les effets aléatoires. Le modèle mixte s'écrit alors :

$$\begin{aligned} Y_i(t) &= Y_i^*(t) + \varepsilon_i(t) \\ &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})B_1(t) + (\beta_2 + b_{2i})B_2(t) + (\beta_3 + b_{3i})B_3(t) \\ &\quad + \beta_4\{B_1(t) \times \text{arm}_i\} + \beta_5\{B_2(t) \times \text{arm}_i\} + \beta_6\{B_3(t) \times \text{arm}_i\} + \varepsilon_i(t), \end{aligned}$$

où  $\text{arm}_i$  prend les valeurs 0 et 1 pour le bras contrôle et expérimental et  $B_k(t); k = 1, 2, 3$  sont les éléments de la base de splines avec 3 degrés de liberté pour des splines cubiques avec deux nœuds internes.

Le sous-modèle pour le temps jusqu'à sortie d'étude est un modèle à risques proportionnels (voir Chapitre 3). C'est le même qui est utilisé dans les deux modèles conjoints, où la fonction de risque instantané s'écrit :

$$h_i(t) = h_0(t) \exp \{ \gamma \text{arm}_i + \alpha Y_i^*(t) \} \quad (5.1)$$

avec  $h_0$  la fonction de risque de base,  $\gamma$  l'effet du bras et  $\alpha$  l'association entre le risque de sortie d'étude et  $Y_i^*(t)$  le score courant de QdV au temps  $t$ . Afin de permettre de la flexibilité dans la modélisation, le logarithme du risque de base  $h_0$  est approximé par une fonction de B-splines.

### Comparaison entre les bras pour les trajectoires de scores

Pour connaître l'effet du bras sur les trajectoires de scores de QdV dans le modèle JM linéaire, nous avons déjà vu qu'il suffit de s'intéresser au terme d'interaction, ce qui revient à tester les hypothèses :

$$\mathcal{H}_0 : \beta_2 = 0$$

$$\mathcal{H}_1 : \beta_2 \neq 0.$$

En revanche, pour le modèle JM splines, cela revient à tester les hypothèses :

$$\mathcal{H}_0 : \beta_4 = \beta_5 = \beta_6 = 0$$

$$\mathcal{H}_1 : \beta_4 \neq 0 \text{ ou } \beta_5 \neq 0 \text{ ou } \beta_6 \neq 0.$$

Pour tester l'hypothèse nulle  $\mathcal{H}_0$  contre l'hypothèse alternative  $\mathcal{H}_1$ , il est possible d'utiliser le test du rapport de vraisemblance défini par :

$$\text{LR} = -2 \left\{ \ell(\hat{\theta}_0) - \ell(\hat{\theta}_1) \right\}$$

où  $\ell(\cdot)$  est la fonction de log-vraisemblance du modèle avec  $\hat{\theta}_0$  et  $\hat{\theta}_1$  les vecteurs des estimateurs du maximum de vraisemblance sous les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$ .

### 5.1.3 Application sur les données de l'essai clinique PRODIGE 5 / ACCORD 17

#### Résultats

L'objectif est d'appliquer les deux modèles conjoints décrits (avec une évolution linéaire ou non linéaire) pour la modélisation des scores. Nous avons choisi 5 nœuds internes pour la fonction B-splines qui approxime le logarithme du risque de base  $h_0$ , placés aux quantiles des temps de sortie d'étude observés (QL : 1, 3, 4, 9, 16 mois ; PF, PA et FA : 1, 3, 4, 10, 16 mois). Concernant le sous-modèle avec des évolutions non linéaires, nous avons choisi 2 nœuds internes placés à 1.1 et 3.5 mois (correspondant à 33.3% et 66.7% des temps de suivi observés) pour les

## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

---

effets fixes et les effets aléatoires ce qui offre déjà une grande flexibilité dans les formes des trajectoires de scores. Pour chacun des modèles conjoints, nous avons choisi 5 points de quadrature adaptative de Gauss-Hermite pour l'approximation des intégrales de la vraisemblance du modèle. Le nombre de nœuds choisi est assez faible, mais l'utilisation de la méthode adaptative de Gauss-Hermite permet de réduire le nombre de nœuds et le temps de calcul qui peut être long (complexité des intégrales à calculer en raison du sous-modèle à splines pour le score).

Les estimations des deux modèles conjoints, à savoir le modèle conjoint (JM linéaire) et le modèle conjoint avec splines (JM splines) sont reportées dans le Tableau 5.2. Quelle que soit la modélisation utilisée, il n'y a pas d'effet du bras  $\hat{\gamma}$  sur le risque de sortie d'étude pour les quatre dimensions étudiées et aucune association n'est détectée pour la dimension PA entre le score courant de QdV et les sorties d'étude. En revanche, les trois associations détectées par le modèle JM linéaire pour les autres dimensions (QL, PF et FA) sont confirmées par le modèle JM splines, avec une association plus faible pour l'échelle PF (JM linéaire :  $\hat{\alpha} = -0.019, p = 0.001$  et JM splines :  $\hat{\alpha} = -0.014, p = 0.002$ ) mais plus forte pour les échelles QL (JM linéaire :  $\hat{\alpha} = -0.016, p = 0.050$  et JM splines :  $\hat{\alpha} = -0.020, p = 0.002$ ) et FA (JM linéaire :  $\hat{\alpha} = 0.013, p = 0.005$  et JM splines :  $\hat{\alpha} = 0.014, p < 0.001$ ). Le paramètre d'association semble être ajusté avec la modélisation non linéaire des scores, en comparaison avec la modélisation linéaire. Par exemple, la diminution de 10 points du score de la dimension QL est associée à un risque plus élevé de sortir d'étude qui passe de 17% (HR = 1.17 [1.00 ; 1.38]) à 22% (HR = 1.22 [1.07 ; 1.38]).

Pour la modélisation du score, il est plus difficile de comparer les estimations des coefficients car les coefficients de régression ne s'interprètent pas de la même façon dans les deux sous-modèles. Les coefficients des splines pour les trajectoires de score n'ayant pas une interprétation immédiate, nous utilisons des représentations graphiques pour tracer les trajectoires moyennes du score de QdV au cours du temps dans chaque bras de traitement (voir Figure 5.2).

Avec des trajectoires linéaires, les scores sont égaux uniquement à l'inclusion, mais évoluent ensuite sans jamais pouvoir être de nouveaux égaux entre les deux

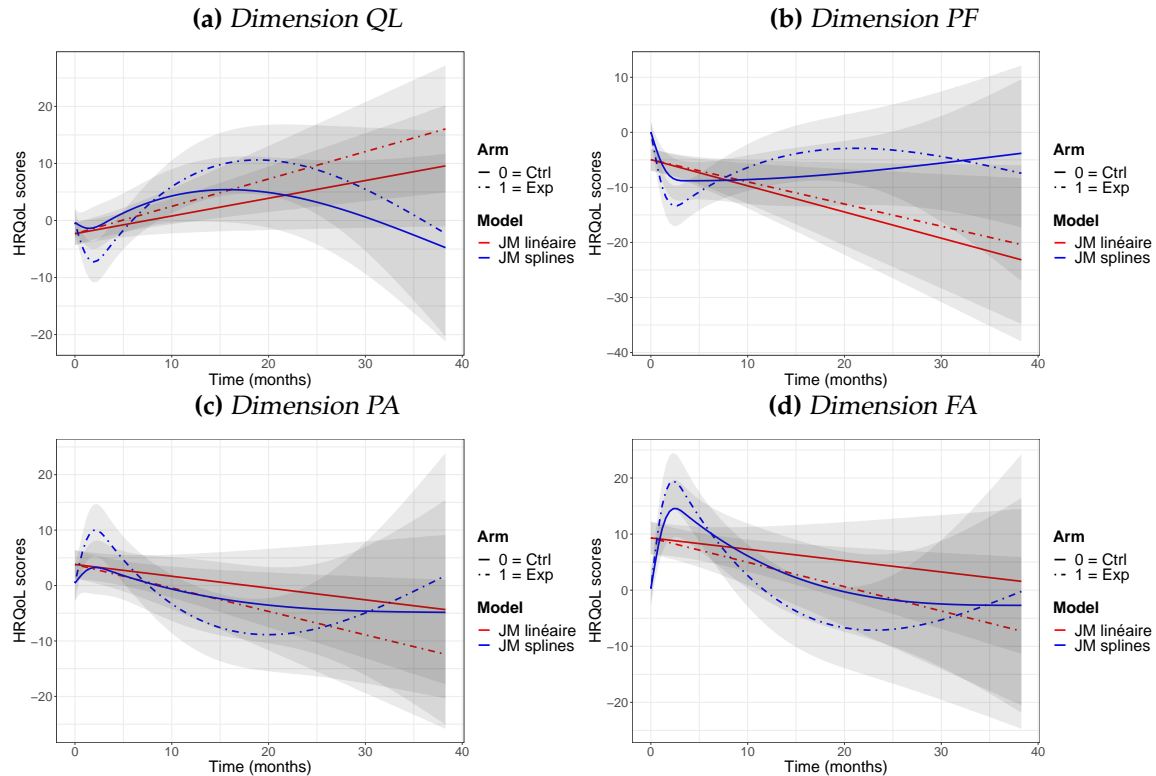
## 5.1. Prise en compte des trajectoires non linéaires

---

bras de traitement. Avec des trajectoires non linéaires, l'effet du traitement sur la QdV peut être décrit plus précisément. Nous remarquons que les trajectoires de QdV entre les deux bras de traitement se croisent entre 5 et 10 mois (fin de période de traitement) pour les 4 dimensions. Nous retrouvons un effet délétère du traitement expérimental sur la QdV à court terme (pendant la période de traitement), mais qui semble bénéfique à plus long terme (pendant le suivi). En effet, pendant la période de traitement (entre l'inclusion et le quatrième mois) la radiochimiothérapie est associée à une diminution des capacités fonctionnelles (PF) et du statut global de santé ainsi qu'à une augmentation des symptômes (PA, FA). Nous remarquons que cette détérioration de la QdV semble être plus marquée dans le bras expérimental (FOLFOX) dont les toxicités sont connues pour être plus importantes. Ensuite, la QdV semble s'améliorer et se stabiliser au cours du suivi pour les quatre dimensions. Même si la QdV a été fortement impactée par le traitement FOLFOX, nous remarquons que pour les quatre dimensions étudiées, les patients du bras expérimental semblent ressentir une amélioration de QdV au cours du suivi par rapport aux patients du bras contrôle. Cette interprétation plus fine de l'évolution de la QdV au cours du traitement et de la prise en charge n'est pas possible avec des trajectoires linéaires. Nous remarquons que les intervalles de confiance (IC) à 95% (tracés en gris) des trajectoires sont assez larges. Cette incertitude sur les estimations peut s'expliquer en raison de la baisse du nombre de mesures collectées (e.g décès des patients, sorties d'étude ou plus généralement données manquantes), notamment à mesure que l'on avance dans le suivi des patients. Les deux modélisations ne détectent pas pour autant de différence entre les deux bras de traitement pour les quatre dimensions, malgré la prise en compte de la non-linéarité des trajectoires (voir Tableau 5.1 pour les valeurs du test du rapport de vraisemblance). Néanmoins, nous remarquons qu'une certaine tendance se dégage pour la dimension QL en utilisant JM splines ( $p = 0.10$ ) par rapport à JM linéaire ( $p = 0.27$ ), sans pour autant conclure à une différence entre les deux bras de traitement. Même si la modélisation linéaire (JM linéaire) est moins flexible, elle paraît néanmoins robuste pour détecter une différence d'évolution entre les deux bras de traitement.

## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

**Figure 5.2** – Trajectoires prédites des scores de QdV modélisés par deux modèles conjoints, JM linéaire (en rouge) et JM splines (en bleu).



### Analyses de sensibilité

Nous avons fait varier le nombre de nœuds internes des splines du sous-modèle pour le score. L'augmentation du nombre de nœuds (3, 5 ou 7 nœuds) n'a pas montré de différences sur les estimations des paramètres du modèle conjoint. De même, l'augmentation du nombre de points de quadrature (5, 7 ou 9 points de quadratures) du modèle conjoint n'a pas montré de différence sur les estimations des paramètres. Ainsi, les résultats semblent robustes avec les paramètres choisis (nombre de nœuds et de points de quadrature).

### Validité des hypothèses des modèles conjoints

Lorsqu'il s'agit d'utiliser des modèles statistiques en pratique pour l'analyse de données, il est important de pouvoir vérifier l'adéquation du modèle aux données

## 5.1. Prise en compte des trajectoires non linéaires

**Tableau 5.1** – Effet du bras de traitement sur l'évolution du score de QdV dans un modèle conjoint (JM linéaire) ou avec des splines (JM splines) donné par le test du rapport de vraisemblance (LR).

Dimension	JM linéaire		JM splines	
	LR	<i>p</i>	LR	<i>p</i>
QL	1.20	0.27	6.26	0.10
PF	0.73	0.40	3.95	0.27
PA	0.06	0.80	5.63	0.13
FA	2.17	0.14	2.59	0.46

pour garantir la pertinence du modèle formulé et les hypothèses associées. Les outils classiques pour valider les hypothèses d'un modèle (e.g. hypothèse de normalité des données observées ou spécification du modèle) sont généralement basés sur l'étude des résidus (spécifiques aux patients ou marginaux par rapport aux effets aléatoires). Les principales propriétés et caractéristiques des résidus à la fois pour les modèles linéaires mixtes (Verbeke and Molenberghs, 2009; Santos Nobre and da Motta Singer, 2007) (Verbeke and Molenberghs, 2009) et les modèles de survie (Cox and Snell, 1968; Therneau et al., 1990) ont déjà beaucoup été étudiées dans la littérature, mais beaucoup moins dans le contexte des modèles conjoints (Dobson and Henderson, 2003).

Une particularité des modèles conjoints est que le processus qui décrit les données répétées peut être affecté par les réalisations d'événements décrites par le second processus (ici les sorties d'étude). Dans cette situation, les résidus pour la partie longitudinale calculés à partir des données observées ne sont pas représentatifs de la population dans sa totalité et l'interprétation des graphiques classiques peut en être faussée (Verbeke et al., 2008). Par exemple, une tendance systématique observée dans le graphique des résidus marginaux peut être attribuée à une mauvaise spécification du modèle, mais aussi à des résidus marginaux impactés par la présence de sorties d'études informatives. Pour contourner cette difficulté en présence de sorties d'études informatives et obtenir des validations graphiques des hypothèses du modèle, Rizopoulos et al. (2010) ont proposé d'imputer les valeurs manquantes sur la base du modèle appliqué aux données observées (toutes les mesures observées comme s'il n'y avait pas eu de sorties d'étude). Ainsi en utilisant les techniques d'imputations multiples pour tenir compte de l'incertitude des

## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

**Tableau 5.2** – Estimations des paramètres pour les deux modèles conjoints avec des trajectoires de score de QdV linéaire (JM linéaire) ou non linéaire (JM splines).

Dimension		JM linéaire			JM splines		
		Est	SE	<i>p</i>	Est	SE	<i>p</i>
QL							
Score	$\beta_0$	-2.318	1.034	0.025	-0.370	1.221	0.762
	$\beta_1$	0.311	0.146	0.033	12.57	6.038	0.037
	$\beta_2$	0.169	0.188	0.369	2.101	4.940	0.671
	$\beta_3$				-3.675	8.235	0.655
	$\beta_4$				25.74	6.067	0.000
	$\beta_5$				2.027	5.141	0.693
	$\beta_6$				2.031	9.058	0.823
	Sortie d'étude	$\gamma$	-0.016	0.131	0.901	-0.035	0.132
$\alpha$		-0.016	0.008	0.050	-0.020	0.006	0.002
PF							
Score	$\beta_0$	-4.96	1.021	<0.001	0.038	0.998	0.969
	$\beta_1$	-0.475	0.201	0.018	-5.517	5.354	0.303
	$\beta_2$	0.072	0.240	0.763	-13.76	4.024	0.001
	$\beta_3$				-0.023	6.802	0.997
	$\beta_4$				8.191	5.483	0.135
	$\beta_5$				-15.60	5.069	0.002
	$\beta_6$				-0.895	9.877	0.928
	Sortie d'étude	$\gamma$	-0.05	0.130	0.704	-0.049	0.130
$\alpha$		-0.019	0.006	0.001	-0.014	0.005	0.002
PA							
Score	$\beta_0$	3.801	1.326	0.004	0.440	1.667	0.792
	$\beta_1$	-0.213	0.185	0.250	-6.334	7.743	0.413
	$\beta_2$	-0.211	0.245	0.389	-2.145	5.756	0.709
	$\beta_3$				-6.802	10.11	0.501
	$\beta_4$				-24.45	7.351	0.001
	$\beta_5$				1.394	6.071	0.818
	$\beta_6$				-3.902	11.08	0.725
	Sortie d'étude	$\gamma$	-0.069	0.131	0.600	-0.061	0.130
$\alpha$		0.008	0.006	0.191	0.007	0.005	0.117
FA							
Score	$\beta_0$	9.315	1.480	<0.001	0.287	1.588	0.857
	$\beta_1$	-0.202	0.177	0.253	-9.504	7.558	0.209
	$\beta_2$	-0.232	0.238	0.330	10.66	5.604	0.057
	$\beta_3$				-10.03	9.481	0.290
	$\beta_4$				-26.03	7.586	0.001
	$\beta_5$				10.70	6.462	0.098
	$\beta_6$				-10.39	12.23	0.396
	Sortie d'étude	$\gamma$	-0.032	0.131	0.809	-0.034	0.130
$\alpha$		0.013	0.005	0.005	0.014	0.004	<0.001

## 5.1. Prise en compte des trajectoires non linéaires

---

valeurs imputées ([Gelman et al., 2005](#)), il est possible de calculer les résidus marginaux obtenus à partir des données complétées. Les imputations multiples font généralement l'hypothèse que les patients ont des mesures planifiées à des visites fixes (évaluations à des temps spécifiés par le protocole par exemple). Cependant, nous avons suivi la méthodologie adoptée par ([Rizopoulos, 2012b](#)) qui consiste à supposer que les patients ont des mesures à des visites aléatoires (évaluations à des temps différents), ce qui est notre cas dans l'application.

### Etude des résidus

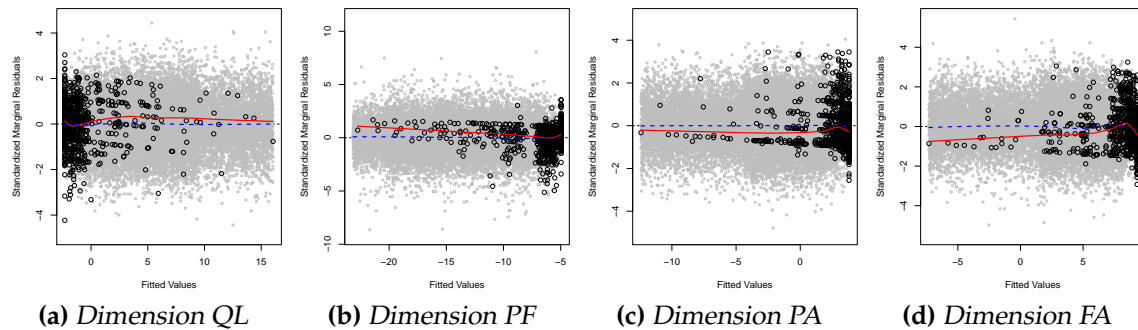
Nous avons étudié les résidus marginaux des modèles conjoints présentés pour étudier les hypothèses de chacun des modèles. Les résidus marginaux sont tracés en distinguant les résidus obtenus à partir des données observées (en noir) et les résidus sur les données complétées (en gris, après  $M = 20$  imputations multiples) pour le modèle conjoint avec des trajectoires linéaires (voir Figure 5.3 et non linéaires (voir Figure 5.4).

Les résidus marginaux standardisés du modèle conjoint montrent une tendance systématique (courbe rouge) par exemple pour la dimension PF où plus les valeurs prédites sont faibles, plus les résidus sont positifs, ce qui suggère que la formulation de la matrice de design associée aux effets fixes n'est pas adaptée. Avec le JM splines (partie droite), cette tendance systématique est plus faible, voire proche de zéro. Ceci est particulièrement vrai pour les dimensions PF, QL et FA. L'utilisation du modèle conjoint splines permet donc de corriger la mauvaise spécification du JM linéaire en effaçant cette tendance systématique grâce à la flexibilité de la modélisation utilisée. Les résidus obtenus après imputations multiples ne montrent pas de tendance particulière (courbe en bleu et en pointillé proche de zéro) ce qui suggère un faible impact des sorties d'étude informatives pour le calcul des résidus pour les deux modèles conjoints.

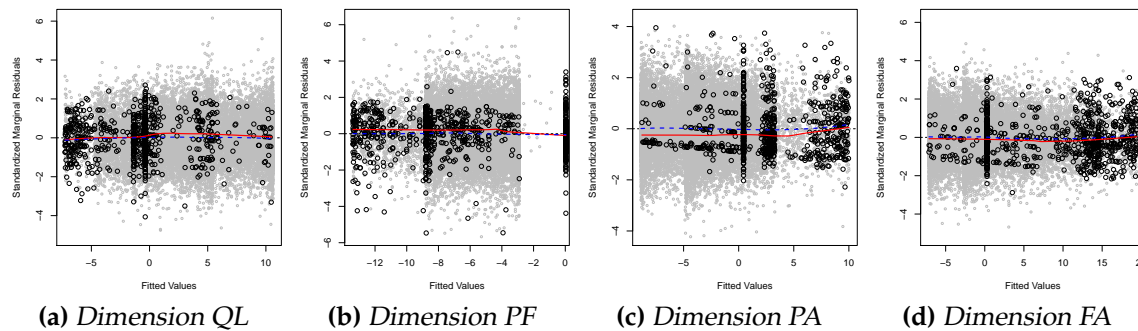
D'autres résidus pour le sous-modèle de temps jusqu'à sortie d'étude peuvent être calculés (résidus de Cox-Snell et résidus martingales) ([Cox and Snell, 1968](#); [Therneau et al., 1990](#)). Ils sont donnés en Annexe D car l'intérêt réside plutôt dans la modélisation des données de score. Nous avons vu que le modèle conjoint avec



## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude



**Figure 5.3** – Résidus marginaux obtenus pour  $M = 20$  imputations avec JM linéaire pour les 4 dimensions analysées.



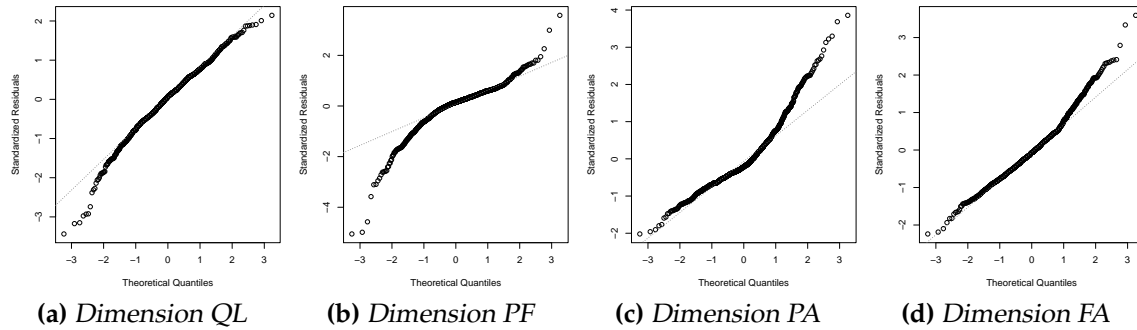
**Figure 5.4** – Résidus marginaux obtenus pour  $M = 20$  imputations avec JM splines pour les 4 dimensions analysées.

splines était à privilégier en étudiant les résidus marginaux pour le sous-modèle du score. Cependant, l'hypothèse de normalité des données de score (qui est une hypothèse fondamentale pour sous-LMM) n'a pas encore été testée. Nous allons donc détailler sur le modèle conjoint à splines l'étude de cette hypothèse.

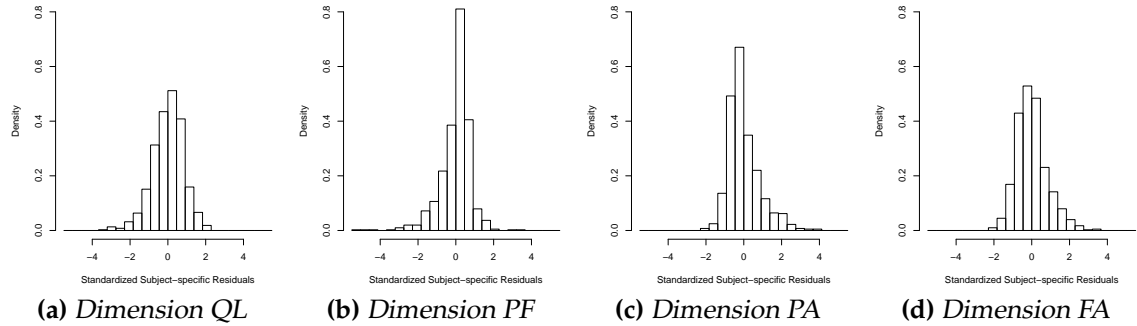
### Etude de la normalité des scores

La modélisation des scores dans les modèles conjoints repose jusqu'alors sur l'hypothèse de continuité et de normalité des scores. Cependant, l'hypothèse de continuité est difficile à croire de par la construction du calcul de score. L'hypothèse de normalité peut être étudiée en traçant l'histogramme des résidus standardisés pour chaque observation et le diagramme quantile-quantile (Q-Q plot) (voir Figure 5.5 et Figure 5.6). Visuellement, les résidus obtenus ne semblent pas

## 5.1. Prise en compte des trajectoires non linéaires



**Figure 5.5** – Q-Q plot des résidus standardisés du modèle conjoint avec splines pour vérifier l'adéquation à l'hypothèse de normalité.



**Figure 5.6** – Histogrammes des résidus standardisés du modèle conjoint avec splines pour vérifier l'hypothèse de normalité des résidus.

toujours se répartir selon une distribution gaussienne (voir la Figure 5.5). Nous remarquons une certaine asymétrie dans la distribution des résidus comparative-ment à la loi normale (en particulier la dimension PA). Ce constat est le même en représentant les Q-Q plots, où les résidus ne se répartissent pas toujours au-tour de la courbe en pointillés (voir la Figure 5.6). Bien que les modèles linéaires mixtes soient robustes à l'hypothèse de normalité, ces outils de diagnostic sou-lignent l'écart à cette hypothèse des résidus du modèle, et *a fortiori* de la distribu-tion des données de scores.

### 5.1.4 Conclusion

Nous avons utilisé jusqu'alors un modèle conjoint avec des trajectoires linéaires pour les données de scores de QdV. Afin de donner un meilleur ajustement aux

## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

---

données observées, nous avons étendu cette modélisation en utilisant des splines pour estimer des trajectoires non linéaires de QdV.

L'utilisation de splines offre une grande flexibilité, mais nécessite de choisir certains paramètres de réglages ([Perperoglou et al., 2019](#)). Les splines cubiques sont habituellement utilisées car elles permettent d'ajuster de nombreuses formes de trajectoires. Alors que l'emplacement des nœuds pour des splines naturelles cubiques ne semble pas vraiment crucial dans la plupart des situations, le choix du nombre de nœuds est en revanche plus difficile à définir ([Harrell Jr, 2015](#); [Stone and Koo, 2015](#)). Il y a un compromis à trouver car l'augmentation du nombre de nœuds peut engendrer un sur-ajustement des données (overfitting) et accroître la variance, tandis que la diminution du nombre de nœuds peut aboutir à une fonction peu flexible ayant tendance à diminuer la qualité des estimations. Pour choisir le nombre de nœuds dans l'application, nous avons comparé plusieurs modèles et utilisé le critère d'information d'Akaike (AIC) pour trouver celui qui donnait le meilleur ajustement aux données. Pour les 4 dimensions de QdV, ce sont les modèles avec seulement 2 nœuds qui ont le plus faible AIC, ce qui nous a permis de choisir ce nombre de nœuds. Une façon correcte de les placer sans *a priori* particulier est de les choisir aux différents quantiles (e.g. avec un seul nœud placé à la médiane) de façon à ce que les nœuds soient espacés de manière équidistante. Sinon, une autre approche populaire pour faciliter le choix du nombre de nœuds consiste à inclure un nombre relativement élevé de nœuds, puis de pénaliser certains coefficients de splines en utilisant des splines pénalisées ([Andrinopoulou et al., 2018](#); [Paul H. C. Eilers and Marx, 1996](#)).

Sur l'application, l'utilisation de splines dans un modèle conjoint n'a pas mis en évidence de différence d'évolution de QdV entre les deux bras de traitement, confirmant ainsi les résultats du modèle conjoint linéaire. Néanmoins, la modélisation conjointe avec des splines semble améliorer l'estimation (ajustement de certains paramètres, notamment le terme d'association) et l'adéquation du modèle aux données. Grâce à cette modélisation plus fine, il est possible d'obtenir une interprétation plus détaillée de l'évolution du score de QdV. Ainsi, l'interprétation clinique des résultats est enrichie, là où des trajectoires linéaires étaient certainement trop restrictives. Toutefois, une difficulté qui reste inchangée est l'hypothèse de continuité et de normalité des scores de QdV des modèles conjoints. Cette hy-

pothèse ne semble pas adaptée puisque les valeurs de scores ne peuvent prendre qu'un nombre fini de valeurs. Par exemple la dimension PA, composée de deux questions à quatre modalités prend sept valeurs de scores possibles. De ce constat, il paraît nécessaire d'envisager d'autres types de modélisation car l'hypothèse de normalité semble compromise et non des plus adaptées pour la modélisation des scores de QdV.

## 5.2 Prise en compte du score en tant que donnée ordinaire

### 5.2.1 Introduction

L'utilisation des scores de QdV est une façon simple pour refléter le niveau de QdV des patients. De par leur nature, les données de QdV font référence à un ordre naturel exprimé par les différents niveaux de score calculés. Les scores obtenus et calculés à partir de la moyenne des réponses aux items ne peuvent prendre qu'un nombre fini de valeurs possibles, bornées entre 0 et 100. La plupart des analyses traitent les scores de QdV comme des données continues, quel que soit le nombre de catégories pour l'échelle considérée, en supposant que les données de scores se répartissent selon une distribution gaussienne, qui est une hypothèse rarement vérifiée en pratique (voir 5.1.3). Pour se rapprocher de la nature des données de score de QdV, nous allons explorer la modélisation des scores en tant que variable catégorielle ordinaire. Dans ce contexte, nous avons besoin de définir une famille de modèles plus large que les LMMs qui ont été jusqu'alors utilisés.

Un cadre alternatif pour l'analyse des données de mesures répétées est l'utilisation de modèles linéaires mixtes généralisés – GLMMs pour Generalized Mixed Linear Models ([Breslow and Clayton, 1993](#)). Ils sont apparus comme la combinaison naturelle entre deux grandes familles de modèles, les modèles linéaires mixtes et les modèles linéaires généralisés – GLMs pour Generalized Linear Models) ([Nelder and Mead, 1965](#)). Les GLMMs forment donc une classe de modèles enrichie en matière d'effets aléatoires par rapport aux GLMs et élargie en matière de lois par

rapport aux LMMs.

Les GLMMs sont donc assez similaires aux LMMs sur la façon dont ils se construisent et nous donnons quelques ouvrages dans lesquels ils sont amplement décrits ([Brown and Prescott, 2014](#); [McCulloch and Searle, 2001](#); [Faraway, 2016](#); [Demidenko, 2013](#)). Nous en rappelons ici les éléments les plus importants. Soit  $Y_i$  le vecteur de la variable longitudinale de l'individu  $i$  ( $i = 1, \dots, n$ ) et  $Y_i(t)$  la valeur prise au temps  $t$ . Conditionnellement au vecteur des effets aléatoires  $b_i$ , les vecteurs  $Y_i$  sont supposés indépendants entre eux et de distribution appartenant à la famille exponentielle ([McCullagh and Nelder, 1989](#)) ; cette famille de loi regroupe un grand nombre de lois usuelles (binomiale, Poisson, normale, gamma, et d'autres).

L'espérance de la variable à expliquer (conditionnellement aux effets aléatoires) est reliée aux variables explicatives par l'intermédiaire d'une fonction de lien :

$$g(E[Y_i(t) | b_i]) = \eta_i(t)$$

où  $g(\cdot)$  une fonction de lien monotone et  $\eta_i(t)$  le prédicteur linéaire au temps  $t$  tel que :

$$\eta_i(t) = X_i^\top(t)\beta + Z_i^\top(t)b_i$$

avec  $X_i(t)$  et  $Z_i(t)$  les deux matrices contenant les variables explicatives associées aux vecteurs d'effets fixes  $\beta$  et aléatoires  $b_i \sim \mathcal{N}(0, D)$ . Puisque la loi normale appartient à la famille exponentielle, nous remarquons que le LMM est bien un cas particulier de GLMM en utilisant la fonction identité comme fonction de lien.

Le calcul de la vraisemblance des GLMMs peut devenir complexe en raison de certaines propriétés calculatoires des LMMs que les GLMMs n'ont pas. En particulier, les GLMMs impliquent le calcul d'intégrales qui n'ont pas de solution analytique, contrairement aux LMMs, grâce à l'hypothèse de normalité à la fois sur la variable à expliquer et sur les effets aléatoires et au fait qu'il existe une relation linéaire directe entre eux ([Stroup, 2012](#), chap. 4). En général, l'estimation des GLMMs nécessite des méthodes numériques adaptées afin d'évaluer et de maximiser la fonction de vraisemblance (ou de quasi-vraisemblance) ([Brown and Prescott,](#)

2014; Faraway, 2016; Demidenko, 2013).

Les GLMMs sont des modèles couramment utilisés en raison de leur large champ d'application, qui peuvent servir aussi bien pour la modélisation de données binaires, de données discrètes quantitatives, et plus généralement, là où l'hypothèse de normalité n'est pas des plus adaptées, comme pour les distributions de scores de QdV.

Pour modéliser des données ordinales, trois familles de modèles de régression logistique prédominent (Stroup, 2012; Agresti, 2010) :

- les modèles cumulatifs, aussi appelés modèles cumulatifs à cotes proportionnelles (*proportional-odds cumulative models*), ou encore modèles de régression logistique basés sur les probabilités cumulées ;
- les modèles adjacents, aussi appelés modèles de régression logistique basés sur les catégories adjacentes ;
- les modèles à ratio de continuation (*continuation ratio model*), aussi appelés modèles séquentiels, ou encore modèles de régression logistique à rapports continus.

Chaque modèle possède des spécificités, ce qui fait qu'il n'existe pas une façon unique d'analyser des données ordinales en tant que telle.

Une problématique qui reste commune dans notre cas quelle que soit la modélisation utilisée est la présence de données manquantes dans les données de QdV. Comme pour les analyses avec des LMMs, les estimations peuvent être fortement impactées par la présence de données manquantes, en particulier à causes des sorties d'études informatives (Blanchin et al., 2011; de Bock et al., 2016). Nous avons vu qu'une façon de tenir compte efficacement des données manquantes est d'utiliser un modèle conjoint. Cependant, les modèles conjoints pour tenir compte des données manquantes informatives sont proposés, en grande majorité, pour tenir compte de l'association entre une variable continue gaussienne mesurée au cours du temps et le temps jusqu'à sortie d'étude.

Actuellement, l'utilisation de modèles conjoints pour traiter des données ordi-

nales reste très limitée (Alsefri et al., 2020). La plupart de ces modèles conjoints sont développés pour une application spécifique et peuvent être difficiles à estimer (Armero et al., 2016). Les modèles à ratios de continuation sont décrits comme étant plus faciles à estimer, même dans le cas d'un modèle conjoint (Andrinopoulou et al., 2014; Andrinopoulou and Rizopoulos, 2016). Les auteurs ont proposé ce modèle en bayésien à partir d'un sous-GLMM pour données binaires. Ils ont construit leur procédure d'estimation basée sur des techniques MCMC et ont utilisé WinBUGS ou JAGS pour estimer le modèle conjoint. Cependant, il reste assez difficile de réinvestir leur travail car même avec leurs codes, disponibles sur demande, cela demanderait un effort non négligeable de programmation pour les utiliser sur de nouvelles données. Ainsi, nous avons plutôt étudié comment il est possible d'estimer ce modèle conjoint à partir des logiciels existants.

Dans un premier temps, nous allons décrire le modèle conjoint utilisé, à savoir un modèle conjoint avec sous-GLMM de type ratios de continuation, dans notre contexte de modélisation de données ordinales de score de QdV avec prise en compte des sorties d'étude. Nous détaillerons ensuite la méthode utilisée pour son estimation et l'appliquerons finalement à l'analyse de la dimension PA des données de QdV de l'essai clinique PRODIGE 5/ACCORD 17.

### 5.2.2 Modèle conjoint avec sous-GLMM à ratios de continuation

Afin de tenir compte d'une variable répétée ordinale dans un modèle conjoint, il est nécessaire de remplacer le sous-LMM par un sous-GLMM pour données ordinales. La suite de ce travail s'intéresse à un modèle conjoint composé : d'un sous-modèle à ratios de continuation pour la modélisation du score de QdV, et d'un sous-modèle à risques proportionnels pour le temps jusqu'à sortie d'étude.

Soit  $Y_i(t)$  la valeur du score de QdV du patient  $i$  au temps  $t$ . Les données de scores sont supposées pouvoir prendre  $K$  valeurs possibles parmi l'ensemble ordonné des catégories  $\{1, \dots, K\}$ .

Le modèle à ratios de continuation est généralement utilisé chez des patients qui passent progressivement d'un état de santé à l'autre pour comprendre ce qui distingue les patients qui ont atteint une catégorie donnée, de ceux qui n'ont pas

## 5.2. Prise en compte du score en tant que donnée ordinale

encore atteint cette catégorie (Harrell Jr, 2015; Fagerland and Hosmer, 2016; Bender and Benner, 2000). Ce modèle est basé sur les probabilités conditionnelles :

$$\omega_{ik}(t) = \Pr(Y_i(t) = k \mid Y_i(t) \leq k).$$

Nous supposons un modèle à ordonnée à l'origine et pente aléatoires avec la fonction de lien *logit* :

$$\text{logit}(\omega_{ik}(t)) = \log\left(\frac{\omega_{ik}(t)}{1 - \omega_{ik}(t)}\right) = \lambda_k + \eta_i(t), \quad k = 2, \dots, K \quad (5.2)$$

avec  $\lambda_k$  les seuils qui représentent l'ordonnée à l'origine de la catégorie  $k$  ( $K - 1$  coefficients seulement sont estimés par souci d'identifiabilité) et où  $\eta_i(t)$ , le prédicteur linéaire, est donné par :

$$\eta_i(t) = \beta_1 t + \beta_2 (t \times \text{arm}_i) + b_{0i} + b_{1i} t,$$

où les effets du temps et d'interaction du temps avec le bras de traitement ne dépendent pas de la catégorie  $k$  (hypothèse de proportionnalité).

Nous avons utilisé la fonction de lien *logit* (basée sur la fonction de répartition d'une loi logistique) mais d'autres fonctions de lien existent comme le lien *probit* (basé sur la fonction de répartition d'une loi normale centrée réduite) ou plus rarement le lien *log-log* (basé sur la fonction de répartition d'une distribution de Gumbel) (Agresti, 2010; Lunn et al., 2001; O'Connell, 2006). La fonction de lien *logit* est généralement la plus populaire parce que la plus simple à interpréter notamment en utilisant des odds ratios (ORs). En effet, on peut écrire aussi :

$$\text{logit}(\omega_{ik}(t)) = \log\left(\frac{\Pr(Y_i(t) = k)}{\Pr(Y_i(t) < k)}\right).$$

Cette écriture se rapproche de la régression logistique binaire, selon laquelle une catégorie serait formée des patients pour lesquels  $Y_i(t) = k$  et l'autre catégorie des patients pour lesquels  $Y_i(t) < k$ .



## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

---

### Interprétation du sous-GLMM

L'interprétation des coefficients  $\beta_1$  et  $\beta_2$  se fait par le biais d'odds ratios (ORs) pour des patients qui partagent les mêmes effets aléatoires. L'OR associé à l'augmentation d'une unité de temps pour le bras contrôle, est donné par :

$$\exp(\beta_1) = \frac{\Pr(Y_i(t+1) = k \mid \text{arm}_i = 0) / \Pr(Y_i(t+1) < k \mid \text{arm}_i = 0)}{\Pr(Y_i(t) = k \mid \text{arm}_i = 0) / \Pr(Y_i(t) < k \mid \text{arm}_i = 0)}.$$

De même, l'OR associé à l'augmentation d'une unité de temps pour le bras expérimental, est donné par :

$$\exp(\beta_1 + \beta_2) = \frac{\Pr(Y_i(t+1) = k \mid \text{arm}_i = 1) / \Pr(Y_i(t+1) < k \mid \text{arm}_i = 1)}{\Pr(Y_i(t) = k \mid \text{arm}_i = 1) / \Pr(Y_i(t) < k \mid \text{arm}_i = 1)}.$$

Ces ORs sont valables quelle que soit la catégorie de score  $k$ .

### Modélisation du temps jusqu'à sortie d'étude

Le sous-modèle pour le temps jusqu'à sortie d'étude est un modèle à risques proportionnels (voir équation Équation 5.1) :

$$h_i(t) = h_0(t) \exp \{ \gamma \text{arm}_i + \alpha \eta_i(t) \} \quad (5.3)$$

avec  $h_0(\cdot)$  la fonction de risque de base,  $\gamma$  l'effet du bras et  $\alpha$  l'association entre le risque de sortie d'étude et  $\eta_i(t)$  le prédicteur linéaire du sous-GLMM évalué au temps  $t$ .

### Estimation

Pour l'estimation de ce modèle conjoint, nous avons utilisé le package `JMbayes` qui permet d'utiliser un sous-GLMM à la place d'un sous-LMM. Cependant, ce package ne peut actuellement pas traiter le cas d'un sous-GLMM à  $K > 2$  catégories (et donc *a fortiori* le cas d'un sous-GLMM à  $K > 2$  catégories ordonnées), c'est-à-dire qu'il permet actuellement de prendre en compte qu'un modèle de régression

logistique simple.

Un des avantages du modèle à ratios de continuation est que sa vraisemblance peut être réécrite comme un produit de lois qui correspond à  $K - 1$  régressions logistiques (Agresti, 2010; Armstrong and Sloan, 1989; Berridge and Whitehead, 1991). Pour estimer le modèle, il faut réorganiser les données d'une certaine façon : Harrell Jr (2015, chap. 13) ont décrit comment procéder et ont même proposé une solution pour le faire avec la fonction `cr.setup` du package `rms` (Harrell Jr, 2013). Les observations (lignes) sont dupliquées en fonction du nombre de régressions logistiques qui vont être appliquées à chaque patient. Au total,  $K - 1$  régressions logistiques sont appliquées : une sur chaque sous-population de patients  $i$  tels que  $Y_i \leq k$ , pour  $k = 2, \dots, K$ . Pour chaque ligne associée à l'observation  $Y_i$ , on introduit la nouvelle variable binaire  $Y_i^*$  égale à 1 si  $Y_i = k$  et 0 si  $Y_i < k$ .

Un exemple de la transformation des données est montré dans le Tableau 5.3 avec  $K = 4$  catégories de score. La première régression logistique est appliquée à la sous-population de patients  $i$  tels que  $Y_i \leq 4$ . On a :

$$(Y_i^*(t) \mid Y_i \leq 4) = \begin{cases} 1 & \text{si } Y_i(t) = 4 \\ 0 & \text{si } Y_i(t) < 4 \end{cases}$$

La deuxième régression logistique est appliquée à la sous-population de patients  $i$  tels que  $Y_i \leq 3$ . On a :

$$(Y_i^*(t) \mid Y_i \leq 3) = \begin{cases} 1 & \text{si } Y_i(t) = 3 \\ 0 & \text{si } Y_i(t) < 3 \end{cases}$$

Enfin, la troisième régression logistique est appliquée à la sous-population de patients  $i$  tels que  $Y_i \leq 2$  et on obtient :

$$(Y_i^*(t) \mid Y_i \leq 2) = \begin{cases} 1 & \text{si } Y_i(t) = 2 \\ 0 & \text{si } Y_i(t) = 1 \end{cases}$$

Le modèle peut être finalement estimé sur la variable binaire  $Y^*$ . En notant la

## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

**Tableau 5.3** – Exemple de la réorganisation des observations pour transformer la variable ordinale  $Y$  en une nouvelle variable binaire  $Y^*$  lorsque  $K = 4$ .

id	visite	$Y$	id	pop	visite	$Y^*$
1	1	2	1	$\leq 2$	1	1
			1	$\leq 3$	1	0
			1	$\leq 4$	1	0
1	2	2	1	$\leq 2$	2	1
			1	$\leq 3$	2	0
			1	$\leq 4$	2	0
1	3	1	1	$\leq 2$	3	0
			1	$\leq 3$	3	0
			1	$\leq 4$	3	0
2	1	4	2	$\leq 4$	1	1
2	2	3	2	$\leq 3$	1	1
			2	$\leq 4$	1	0

probabilité conditionnelle associée à cette variable :

$$\omega_{ik}^*(t) = \Pr(Y_i^*(t) = 1 \mid Y_i \leq k),$$

le modèle à ratios de continuation prend la forme :

$$\text{logit}(\omega_{ik}^*(t)) = \lambda_k \text{pop}_k + \beta_1 t + \beta_2 (t \times \text{arm}_i) + b_{0i} + b_{1i} t$$

où  $\text{pop}_k$  est l'indicatrice d'appartenance à la sous-population de patients  $i$  tels que  $Y_i \leq k$  pour  $k = 2, \dots, K$  et  $\lambda_k$  l'ordonnée à l'origine pour chaque régression logistique sur  $Y_i^*$ .

Nous avons utilisé ce recodage pour estimer les modèles à ratios de continuation dans le modèle conjoint. En ce qui concerne le sous-modèle du temps jusqu'à sortie d'étude, le package `JMbayes` suppose que le risque de base est modélisé par des B-splines. Nous avons validé cette façon d'estimer le modèle conjoint par simulations (voir Annexe E). Les valeurs estimées des paramètres correspondaient bien aux valeurs simulées, pour  $K = 4$  catégories.

### 5.2.3 Application sur les données de l'essai clinique PRODIGE 5/ACCORD 17

#### Catégories de scores

Nous avons déjà évoqué que le nombre de modalités selon chaque dimension permettait d'obtenir seulement certaines valeurs de scores, selon les recommandations du calcul des scores par l'EORTC. Par exemple, dans le cas de la dimension PA, le score peut prendre 7 catégories différentes (0, 16.66, 33.33, 50, 66.66, 83.33, 100). En fonction de la dimension considérée, le nombre de catégories peut être élevé (avec QL : 13, PF : 21, PA : 7, FA : 13 catégories observées). Même s'il n'existe pas, à notre connaissance, de règle sur le nombre de catégories maximal dans un modèle ordinal, un tel modèle ne fait sens qu'avec un nombre relativement faible de catégories. Dans notre contexte, un modèle ordinal n'aura finalement du sens que pour les dimensions uni-item (4 catégories différentes) ou à 2 items comme PA (7 catégories différentes) mais pas au-delà. C'est la raison pour laquelle nous analysons seulement les scores obtenus pour la dimension PA, qui est la dimension avec le moins de catégories, et pour laquelle l'hypothèse de normalité semble la moins satisfaisante (voir 5.1.3).

#### Résultats sur l'application

Les résultats de l'application sur les données de l'essai clinique PRODIGE 5/ACCORD 17 pour la dimension PA sont donnés dans le Tableau 5.4. Nous avons également ajouté les résultats du modèle conjoint linéaire (JM linéaire, présenté en section 5.1) afin de mettre en perspective les résultats obtenus avec le modèle conjoint à ratios de continuation (appelé JM ordinal).

Les estimations de JM ordinal montrent qu'il n'y a probablement pas d'effet du temps ( $\hat{\beta}_1$  proche de 0, avec  $\hat{\beta}_1 = -0.117, P = 0.266$ ). Il n'y a pas non plus d'effet du traitement ( $\hat{\beta}_2$  proche de 0 avec  $\hat{\beta}_2 = 0.041, P = 0.739$ ). Ces résultats sont cohérents avec ceux obtenus par JM linéaire ( $\hat{\beta}_1 = -0.213, p = 0.250$  et  $\hat{\beta}_2 = -0.211, p = 0.389$ ). Ensuite, l'effet du bras sur le risque de sortie d'étude est

## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

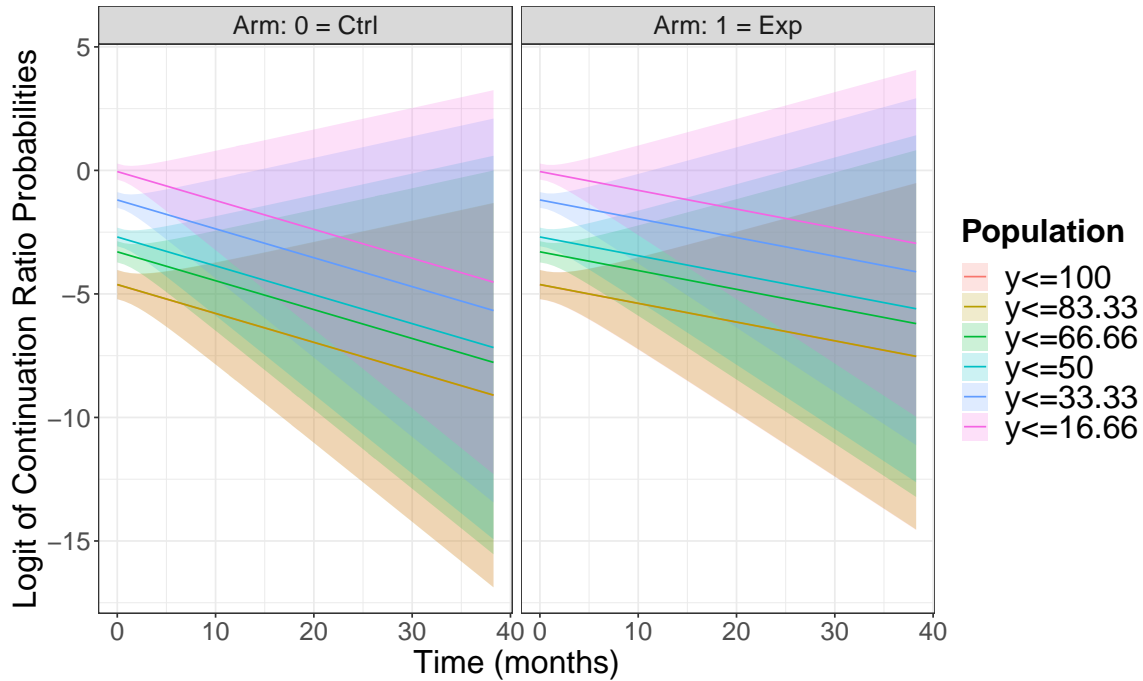
**Tableau 5.4** – Estimations des paramètres du modèle conjoint avec des trajectoires de score de QdV linéaire (JM linéaire, déjà présenté) et du modèle conjoint pour données ordinales avec sous-modèle à ratios de continuation (JM ordinal).

Dimension	JM linéaire			JM ordinal		
	Est	SE	$p$	Est	SE	$P$
<b>PA</b>						
Score						
$\lambda_2$	-	-	-	-0.047	0.006	0.766
$\lambda_3$	-	-	-	-1.196	0.005	<0.001
$\lambda_4$	-	-	-	-2.694	0.006	<0.001
$\lambda_5$	-	-	-	-3.296	0.009	<0.001
$\lambda_6$	-	-	-	-4.62	0.01	<0.001
$\lambda_7$	-	-	-	-4.625	0.01	<0.001
$\beta_0$	3.801	1.326	0.004	-	-	-
$\beta_1$	-0.213	0.185	0.250	-0.117	0.007	0.266
$\beta_2$	-0.211	0.245	0.389	0.041	0.008	0.739
Sortie d'étude						
$\gamma$	-0.069	0.131	0.600	-0.074	0.003	0.565
$\alpha$	0.008	0.006	0.191	0.051	0.001	0.002

facilement comparable d'un modèle à l'autre avec des coefficients dont les estimations sont très proches (JM ordinal :  $\hat{\gamma} = -0.074$  et JM linéaire :  $\hat{\gamma} = -0.069$ ). Les estimations des coefficients d'association sont quant à elles différentes d'un modèle à l'autre (JM ordinal :  $\hat{\alpha} = 0.051$  et JM linéaire :  $\hat{\alpha} = 0.008$ ) puisque la QdV est modélisée différemment dans les deux modèles. Le risque de sortie d'étude ne dépend pas du bras de traitement dans les deux modèles alors que JM ordinal montre que le paramètre d'association est significatif ( $\hat{\alpha} = 0.051$ ,  $P = 0.002$ ). L'absence d'association dans JM linéaire ( $\hat{\alpha} = 0.008$ ,  $p = 0.191$ ) pourrait éventuellement s'expliquer par le fait que le sous-LMM n'est pas adapté.

Finalement, les résultats obtenus sur la dimension PA restent assez cohérents avec ceux obtenus par les précédents modèles conjoints (voir Tableau 5.2), sans distinguer de différence entre les deux bras de traitement. Cependant, le modèle conjoint à ratios de continuation est le seul modèle conjoint à avoir détecté un terme d'association pour cette dimension. Ainsi, il pourrait être intéressant d'analyser les dimensions avec peu de catégories par ce type de modèle pour apporter un nouvel éclairage sur les données.

## 5.2. Prise en compte du score en tant que donnée ordinale



**Figure 5.7** – Estimation des logits de ratios de continuation pour l’analyse des scores de la dimension PA.

### 5.2.4 Conclusion

Nous nous sommes intéressés aux GLMMs, qui est le cadre utilisé pour modéliser des données qui ne respectent pas forcément l’hypothèse de normalité. Nous avons donc supposé le score comme une variable ordinale plutôt que continue et considéré un modèle à ratios de continuation en tant que sous-GLMM dans un modèle conjoint. Le modèle à ratios de continuation estime la probabilité d’être dans une catégorie de score, étant donné d’avoir déjà été dans cette catégorie ou dans une catégorie supérieure. Des effets aléatoires (ordonnée à l’origine et pente aléatoires) ont été introduits de façon à tenir compte des données répétées.

Nous avons utilisé la propriété des modèles à ratios de continuation selon laquelle il est possible d’estimer ce modèle à partir de régressions logistiques binaires. Cette réécriture du modèle avait déjà été utilisée pour un modèle conjoint, mais sans considérer de pente aléatoire par soucis de convergence ([Andrinopoulou et al., 2014](#)). Les résultats sur la dimension PA montrent que ni le temps ni le bras de traitement ne modifie de manière significative les probabilités d’avoir

## Chapitre 5. Modélisations alternatives des scores de QdV dans un modèle conjoint en présence de sorties d'étude

---

de forts symptômes de douleur par rapport à des symptômes de douleurs moins importants.

Ce travail a fait l'hypothèse que l'effet du temps et du traitement sont constants quelles que soient les catégories (les coefficients du modèle sont indépendants de la catégorie). Le nombre de paramètres à estimer dans le sous-GLMM est donc raisonnable. Comme la dimension PA a déjà 7 niveaux de catégories, nous n'avons pas testé cette hypothèse pour limiter le nombre de paramètres à estimer.

L'interprétation du modèle à ratios de continuation n'est pas facile. D'autres modèles tels que les modèles cumulatifs sont plus intuitifs quant à leur interprétation. Les modèles cumulatifs permettent de comparer la probabilité d'être égale ou inférieure à une catégorie, par rapport à la probabilité d'être dans une catégorie plus élevée. A titre illustratif, nous avons ajouté les résultats sur la dimension PA avec un modèle cumulatif (voir Annexe F) sans tenir compte des sorties d'étude. Les résultats obtenus restent cohérents et montrent que ni le temps ni le traitement n'a un impact sur les probabilités modélisées. De plus, les modèles cumulatifs présentent des propriétés intéressantes telles que l'interprétation via une variable latente ([Anderson and Philips, 1981](#); [Masyn et al., 2014](#)) qui est particulièrement utile pour l'étude de la QdV. La variable latente permet de prendre en compte le concept de la QdV comme une variable continue et non directement observable, pour laquelle seulement quelques valeurs de scores peuvent être obtenues selon les recommandations du calcul du score de l'EORTC. Il est possible de tracer les trajectoires prédites de la variable latente dans chaque bras de traitement, ce qui facilite la visualisation et l'interprétation du modèle via cette quantité (voir Figure F.1).

Il aurait été intéressant de pouvoir considérer un modèle conjoint avec un sous-modèle cumulatif pour les catégories de scores. D'abord, pour tenir compte des sorties d'étude dans l'analyse puisque nous avons vu dans le modèle à ratios de continuation qu'il existait une association entre la dimension PA et le risque de sortie d'étude, puis pour faciliter également l'interprétation des coefficients du modèle. Le terme d'association aurait également une interprétation plus simple : il quantifierait l'augmentation (ou la diminution) du risque de sortie d'étude associé à l'évolution de la variable latente.

## 5.3 Discussion

Dans ce dernier chapitre, nous avons proposé des alternatives de modélisation dans le cadre d'une modélisation conjointe des scores de QdV et des sorties d'étude.

Dans un premier temps nous avons utilisé des évolutions non linéaires pour les trajectoires de scores de QdV grâce à un sous-LMM avec des splines. Cette modélisation plus flexible a permis de mettre en évidence une diminution de la QdV notamment pendant la période de traitement, où la QdV semble fortement impactée par celui-ci. Cette modélisation donne ainsi une interprétation plus fine sur l'évolution des scores, même si elle n'aboutit pas à une conclusion différente en ce qui concerne l'effet global du traitement sur la qualité de vie. Cependant, il est probable que dans d'autres applications, le modèle avec des splines puisse détecter une différence non mise en évidence avec des trajectoires linéaires. Même si ce modèle permet un meilleur ajustement aux données, il fait toujours l'hypothèse de normalité sur les résidus. Il a été montré que cette hypothèse peut être mise en défaut, notamment pour les dimensions ne considérant qu'un ou deux items (et donc avec peu de catégories) comme la dimension PA.

Pour répondre à cette problématique, nous avons exploré, dans un second temps, la modélisation du score de QdV en tant que variable ordinale plutôt que continue. Nous avons utilisé un modèle à ratios de continuation comme sous-GLMM du modèle conjoint. Sur la dimension PA, le modèle a détecté une association avec la QdV et le risque de sortie d'étude. Aucun autre modèle conjoint jusqu'alors utilisé n'avait détecté d'association pour cette dimension. Bien que facile à estimer, ce modèle reste cependant difficile à interpréter. Les modèles conjoints avec un sous-modèle cumulatif pour les données ordinales sont encore aujourd'hui très peu étudiés ([Armero et al., 2016](#); [Chakraborty and Das, 2010](#); [Li et al., 2010](#); [Luo, 2014](#)). Pourtant, ils présentent de bonnes propriétés d'interprétation, notamment avec la représentation du modèle via une variable latente. Les difficultés d'estimation sont certainement un frein à leur utilisation et à l'implémentation dans les logiciels, mais ils pourraient être une alternative pour analyser les dimensions présentant peu de catégories de score ([Barbieri et al., 2017](#)).





## Conclusion et perspectives

---

La QdV apparaît comme un critère majeur de plus en plus utilisé dans les essais cliniques en oncologie pour évaluer l'efficacité d'une nouvelle stratégie thérapeutique ou d'un nouveau traitement. Elle est évaluée par le biais d'auto-questionnaires transmis aux patients à différentes visites au cours de leur prise en charge. En Europe, le questionnaire le plus couramment utilisé est le QLQ-C30 de l'EORTC qui évalue des dimensions fonctionnelles, symptomatiques et le statut global de santé. Ces questionnaires sont souvent en partie incomplets (items mal ou non renseignés), voire complètement manquants (non complétés, non rendus ou non transmis au patient). En général, l'analyse longitudinale des scores de QdV est réalisée via des modèles linéaires mixtes (LMMs). Cependant, ces modèles peuvent donner des estimations biaisées en présence de données manquantes. Ainsi, la prise en compte des données manquantes est à l'heure actuelle un enjeu important et une problématique majeure dans l'analyse longitudinale de la QdV dans les essais thérapeutiques en oncologie.

L'objectif de ce travail était d'explorer les méthodes statistiques disponibles et adéquates pour tenir compte des données manquantes, d'identifier les modèles les plus pertinents et d'étudier comment ils pourraient améliorer la prise en compte des données manquantes afin de proposer des recommandations au moment de l'analyse de la QdV ou plus précocement dès la rédaction du protocole.

### 6.1 Contributions par rapport aux objectifs initiaux

Dans cette thèse, nous nous sommes concentrés uniquement sur le cas des données manquantes de type monotone (assimilées à des sorties d'étude, au sens de la QdV) ; elles sont très présentes dans les données de QdV (e.g. suite au décès ou à la progression de la maladie) et leur prise en compte représente un véritable enjeu au moment de l'analyse.

Tout d'abord, nous avons fait un état de l'art de la littérature existante sur les données manquantes, et rappelé les trois mécanismes de données manquantes (MCAR, MAR et MNAR). Lorsque le mécanisme de données manquantes est supposé MCAR ou MAR (non informatif), des modèles basés sur la vraisemblance des données (comme les LMMs) peuvent être utilisés. En présence d'un mécanisme de type MNAR (informatif), il est alors nécessaire de modéliser simultanément la variable d'intérêt et le mécanisme de données manquantes au risque d'obtenir des résultats biaisés. Cet état de l'art de la littérature nous a permis d'orienter la suite de notre travail vers ce type de modélisation. Pour tenir compte de ces sorties d'étude suspectées informatives, différents modèles ont pu être identifiés : les modèles de sélection, les modèles de mélange de profils et enfin les modèles conjoints à effets partagés. Ces trois modèles sont décrits au début du manuscrit, dans le Chapitre 2. Ce travail a abouti à la rédaction d'un chapitre de livre, permettant de mon point de vue de diffuser ces approches et inciter les lecteurs à les utiliser dans le contexte des essais cliniques.

Parmi les trois modèles présentés, nous avons sélectionné le plus pertinent à considérer dans notre contexte sur la base d'arguments méthodologiques et pratiques. D'une part, les modèles de sélection peuvent être difficiles à appréhender et nécessitent généralement un travail assez conséquent de programmation. Ils modélisent seulement la probabilité de sortie d'étude à chaque visite, sans tenir compte du délai d'apparition. D'autre part, les modèles de mélange de profils n'estiment pas directement les quantités d'intérêts et la sélection des différents profils peut être difficile à réaliser et peut engendrer des problèmes de non-identifiabilité. Les modèles conjoints à effets partagés sont en revanche plus intéressants sur plusieurs aspects. D'abord, ils tiennent compte du délai d'apparition

## 6.1. Contributions par rapport aux objectifs initiaux

---

des sorties d'étude avec un sous-modèle de temps jusqu'à évènement. Ensuite, ce sous-modèle est assez simple d'interprétation : le terme d'association permet de juger si les sorties d'étude sont informatives ou non et l'effet du bras permet de savoir si le risque de sortie d'étude est différent en fonction du traitement. Les modèles conjoints à effets partagés sont aussi facilement comparables aux LMMs puisqu'ils sont aussi constitués d'un sous-LMM. Enfin, l'utilisation de ces modèles est facilitée par l'existence de nombreuses implémentations dans les logiciels de statistique standards tels que R, SAS et Stata. Pour ces raisons, les modèles conjoints à effets partagés (appelés aussi modèles conjoints) constituent donc une alternative intéressante au LMM seul pour la prise en compte des sorties d'étude potentiellement informatives. L'ensemble de ce travail est traité dans le Chapitre 3 du manuscrit, et a été publié dans la revue *BMC Medical Research Methodology* afin de proposer ces solutions pour l'analyse de la QdV dans les essais cliniques en présence de sorties d'étude.

La suite du manuscrit s'est alors concentrée exclusivement sur les modèles conjoints. Nous avons examiné séparément les deux sous-modèles du modèle conjoint afin de proposer des améliorations de chacun d'eux, adaptées au contexte de la QdV.

Tout d'abord, nous nous sommes intéressés au sous-modèle pour le temps jusqu'à sortie d'étude. Ce travail s'inspire des données de l'essai clinique PRODIGE 5/ACCORD 17 pour des patients atteints d'un cancer de l'œsophage localement avancé. Ces patients, pour lesquels on recueille les données de QdV, sont souvent à risque de progression ou de décès. Or, un tel évènement clinique peut induire des sorties d'étude susceptibles d'être informatives (ou « plus informatives ») par rapport aux autres sorties d'étude. Il est probablement trop restrictif de considérer que les sorties d'étude sont toutes associées de la même façon à la QdV. Nous avons donc utilisé un modèle conjoint qui permet de différencier les causes de sorties d'étude grâce à un sous-modèle à risques concurrents. Nous l'avons comparé au modèle conjoint standard sur les données de score de QdV de l'essai puis sur des données simulées. Ce travail a permis notamment de justifier son utilisation, dès que le caractère informatif ou non de la sortie d'étude dépend de sa cause ou, plus généralement, lorsque le degré et/ou le sens de l'association entre QdV et risque de sortie d'étude dépend de sa cause. En plus de donner des estimations non bia-

## Chapitre 6. Conclusion et perspectives

---

sées, le modèle conjoint à risques concurrents permet une interprétation plus riche du mécanisme à l'origine des sorties d'étude. Finalement, ce travail souligne les avantages de considérer un tel modèle tant au niveau de l'interprétation, que de l'estimation des paramètres (Chapitre 4).

Après avoir étudié le sous-modèle du temps jusqu'à sortie d'étude, nous nous sommes ensuite intéressés au sous-modèle qui décrit les trajectoires du score de QdV (Chapitre 5).

Dans un premier temps, il nous a semblé intéressant de considérer que les scores puissent évoluer autrement que selon une trajectoire linéaire. Nous avons assoupli cette hypothèse en introduisant de la non-linéarité sur les trajectoires de score grâce à un sous-LMM utilisant des splines. Les coefficients estimés sont plus nombreux et plus difficilement interprétables. Par exemple, il n'y a plus un coefficient, mais des coefficients associés à l'effet du traitement. Par conséquent, des tests basés sur le rapport de vraisemblance doivent être utilisés pour tester l'effet du traitement sur l'évolution de la QdV. Il est conseillé de représenter les trajectoires de QdV prédites par le modèle afin de visualiser les évolutions au cours du temps dans les deux bras de traitement. Néanmoins, l'utilisation des splines offre une plus grande flexibilité de modélisation et souvent un meilleur ajustement aux données. L'interprétation clinique des résultats est aussi enrichie par rapport à des trajectoires linéaires. Pour ces multiples raisons, le sous-modèle avec des splines est à privilégier. Cependant, il reste un LMM et fait donc une hypothèse de normalité et de continuité sur les scores de QdV qui n'est pas forcément appropriée aux données rencontrées dans notre contexte. En effet, les scores ne peuvent prendre qu'un nombre fini de valeurs ordonnées, sont bornés entre 0 et 100, et leur répartition pour certaines dimensions peut être fortement asymétrique.

C'est la raison pour laquelle, nous avons considéré dans un second temps, la variable de score en tant que variable ordinale plutôt que continue. Cette modélisation semble plus adaptée à la façon dont les scores sont construits, en particulier pour les dimensions construites à partir de 1 ou 2 items. Par exemple, la dimension PA constituée de 2 items du QLQ-C30 ne donne lieu qu'à 7 valeurs possibles de score. Nous avons proposé d'utiliser un modèle conjoint constitué d'un sous-GLMM de type ratio de continuation pour la variable ordinale de score. L'esti-

mation de ce modèle peut être réalisée avec des logiciels existants, ce qui rend ce modèle assez attractif. Avec ce modèle conjoint, c'est la première fois qu'une association entre le score de QdV pour cette dimension et le risque de sortie d'étude a été détectée, ce qui souligne la pertinence de son utilisation. Les estimations du sous-modèle à ratio de continuation ne sont pas les plus facilement interprétables et d'autres modèles plus populaires, comme le modèle cumulatif, pourraient le substituer, mais sont plus difficiles à mettre en œuvre dans un contexte de modélisation conjointe.

## 6.2 Perspectives du travail de thèse

Une hypothèse fondamentale des LMMs repose sur la normalité de la variable réponse. L'hypothèse de continuité sur les scores de QdV ne semble pas la plus adaptée en raison du nombre de valeurs prises par certaines dimensions (en particulier pour les dimensions construites à partir de 1 ou 2 items). De plus, le choix d'une distribution normale n'est pas le plus approprié pour certaines distributions des scores de QdV avec par exemple une forte asymétrie dans leur répartition. Les données de QdV présentent des effets plafond ou plancher, en raison de la propriété bornée du score calculé. Ces limites méthodologiques, au-delà de la prise en compte des sorties d'étude informatives, posent la question de la qualité des résultats issus des LMMs et soulèvent des perspectives intéressantes de travail pour améliorer la modélisation des données de QdV ([Barbieri et al., 2016](#)).

Il serait envisageable de transformer les données de score pour se rapprocher d'une distribution normale, afin d'analyser les valeurs de scores transformées par les outils utilisés dans cette thèse. Les transformations possibles pourraient s'inspirer des travaux qui utilisent les processus à classes latentes sur des données de tests psychométriques dans l'étude de la démence ([Philipps et al., 2014](#)).

Une autre solution consisterait à abandonner l'hypothèse de normalité au profit d'autres sous-modèles de type GLMM pour lesquels le terme d'erreur est supposé suivre une distribution continue non gaussienne, en s'appuyant sur les travaux de [Rizopoulos \(2019\)](#). On trouve par exemple la loi de Student qui est une loi symétrique et la loi Beta qui prend des formes assez flexibles et surtout dont le do-

## Chapitre 6. Conclusion et perspectives

---

maine définition est un intervalle borné, ce qui est une propriété adaptée pour des scores de QdV compris entre 0 et 100. La loi Beta a d'ailleurs déjà été utilisée pour l'analyse de scores de QdV mais pas dans le contexte d'une modélisation conjointe ([Hunger et al., 2012](#)). Une perspective serait donc d'explorer ces modèles dans le cadre des modèles conjoints.

Pour les dimensions à 1 ou 2 items, il serait plus intéressant de traiter le score comme une variable ordinale ([Bascoul-Mollevis et al., 2020](#)) plutôt que continue. Une suite logique de ce travail serait de travailler sur le modèle cumulatif qui semble le plus adéquat pour prendre en compte l'aspect ordinal des données de QdV. En effet, l'utilisation de la variable latente sous-jacente qu'il modélise est un véritable atout pour la visualisation et l'interprétation des résultats ([Barbieri et al., 2017](#)). Il est possible de tracer les trajectoires prédites de la variable latente dans chaque bras de traitement. Une des perspectives serait la mise en œuvre d'un modèle cumulatif dans le contexte d'un modèle conjoint. De tels modèles ont finalement été que très peu considérés ([Armero et al., 2016](#); [Chakraborty and Das, 2010](#); [Li et al., 2010](#)) et ne sont généralement pas disponibles dans les logiciels pour les utiliser facilement.

Nous avons abordé dans l'ensemble de ce travail uniquement la question des données manquantes monotones. En pratique, des données manquantes intermittentes surviennent également pendant le recueil des questionnaires de QdV. Généralement, les données manquantes intermittentes ne sont pas prises en compte. Cela revient à supposer finalement qu'elles sont de type MCAR ou MAR. Une perspective possible de ce travail serait de vérifier ces hypothèses sur les données manquantes intermittentes, et éventuellement proposer des approches pour tenir compte.

Plus généralement pour l'analyse de la QdV, il serait intéressant de prendre en compte son aspect multidimensionnel (un score par dimension). Actuellement, le score de QdV est analysé dimension par dimension, ce qui pose en outre le problème de la multiplicité des tests. Lorsque des dimensions spécifiques sont définies dans le protocole comme étant d'un intérêt particulier (dans l'essai clinique analysé, 4 dimensions), il serait probablement plus pertinent de les analyser simultanément afin de pouvoir conclure globalement sur la QdV relative à l'ensemble de

ces dimensions. Théoriquement, il est possible d'utiliser des extensions de modèle conjoint en considérant un sous-GLMM multivarié pour considérer la QdV comme une variable multidimensionnelle pour répondre à cette question ([Papageorgiou et al., 2019](#); [Hickey et al., 2016](#)) tout en tenant compte des sorties d'étude. L'estimation de tels modèles peut devenir complexe en raison de l'augmentation de la dimension des effets aléatoires et des intégrales à calculer. Pour cela, des solutions logicielles existent comme les packages `joinerML` ([Hickey et al., 2018](#)) et `JMBayes` ([Rizopoulos, 2016](#)) où les modèles peuvent être explorés dans un contexte fréquentiste ou bayésien.

Enfin, il nous paraît important de continuer de participer à la diffusion des méthodes présentées au sein de la communauté scientifique (congrès, articles) pour encourager leur utilisation dans les essais cliniques.

## 6.3 Conclusion générale

Il existe actuellement des recommandations sur le report des résultats lorsque des mesures reportées par le patient (PROs) sont utilisées dans les essais cliniques ([Calvert et al., 2013, 2018](#)). En ce qui concerne les données manquantes, il est recommandé d'inclure dans le protocole les stratégies pour minimiser leur apparition, reporter précisément le nombre de données manquantes et clairement documenter les modèles utilisés pour en tenir compte. Dans ce contexte, le projet européen SISAQOL, mené par l'EORTC, a été lancé afin de proposer des recommandations pour l'analyse de la QdV dans les essais cliniques en oncologie ([Bottomley et al., 2018](#); [Coens et al., 2020](#)). Les premiers travaux ont permis d'obtenir quelques indications supplémentaires sur le traitement des données manquantes. Par exemple, il est recommandé d'utiliser au moins deux approches pour évaluer l'effet des données manquantes sur l'analyse et ainsi s'assurer de la robustesse des résultats obtenus. Cependant, aucune méthode n'a clairement été mise en évidence pour guider le choix de celle-ci.

En conclusion, ce travail de thèse a permis de proposer plusieurs stratégies possibles pour l'analyse longitudinale de la QdV en présence de sorties d'étude potentiellement informatives. Ces approches devraient être mises en œuvre en plus



## Chapitre 6. Conclusion et perspectives

---

des approches classiques pour s'assurer, à minima, de la robustesse des résultats obtenus. Deux principales recommandations se dégagent finalement de nos travaux. D'une part, il faudrait proposer l'utilisation des modèles conjoints pour éviter d'obtenir des résultats biaisés par la présence de données manquantes au moment de l'analyse longitudinale de la QdV dans les essais thérapeutiques. D'autre part, il faudrait s'assurer, dès la conception du protocole, du recueil systématique et standardisé des raisons de sorties d'étude pour faciliter et guider les éventuelles analyses qui pourraient être réalisées par la suite.





# Annexe A

## **Questionnaire EORTC QLQ-C30**

---

## Annexe A. Questionnaire EORTC QLQ-C30

ENGLISH



### EORTC QLQ-C30 (version 3)

We are interested in some things about you and your health. Please answer all of the questions yourself by circling the number that best applies to you. There are no "right" or "wrong" answers. The information that you provide will remain strictly confidential.

Please fill in your initials:

--	--	--	--

Your birthdate (Day, Month, Year):

--	--	--	--	--	--	--	--

Today's date (Day, Month, Year):

31							
----	--	--	--	--	--	--	--

	Not at All	A Little	Quite a Bit	Very Much
1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?	1	2	3	4
2. Do you have any trouble taking a <u>long</u> walk?	1	2	3	4
3. Do you have any trouble taking a <u>short</u> walk outside of the house?	1	2	3	4
4. Do you need to stay in bed or a chair during the day?	1	2	3	4
5. Do you need help with eating, dressing, washing yourself or using the toilet?	1	2	3	4

#### During the past week:

	Not at All	A Little	Quite a Bit	Very Much
6. Were you limited in doing either your work or other daily activities?	1	2	3	4
7. Were you limited in pursuing your hobbies or other leisure time activities?	1	2	3	4
8. Were you short of breath?	1	2	3	4
9. Have you had pain?	1	2	3	4
10. Did you need to rest?	1	2	3	4
11. Have you had trouble sleeping?	1	2	3	4
12. Have you felt weak?	1	2	3	4
13. Have you lacked appetite?	1	2	3	4
14. Have you felt nauseated?	1	2	3	4
15. Have you vomited?	1	2	3	4
16. Have you been constipated?	1	2	3	4

Please go on to the next page

**During the past week:**

	Not at All	A Little	Quite a Bit	Very Much
17. Have you had diarrhea?	1	2	3	4
18. Were you tired?	1	2	3	4
19. Did pain interfere with your daily activities?	1	2	3	4
20. Have you had difficulty in concentrating on things, like reading a newspaper or watching television?	1	2	3	4
21. Did you feel tense?	1	2	3	4
22. Did you worry?	1	2	3	4
23. Did you feel irritable?	1	2	3	4
24. Did you feel depressed?	1	2	3	4
25. Have you had difficulty remembering things?	1	2	3	4
26. Has your physical condition or medical treatment interfered with your <u>family</u> life?	1	2	3	4
27. Has your physical condition or medical treatment interfered with your <u>social</u> activities?	1	2	3	4
28. Has your physical condition or medical treatment caused you financial difficulties?	1	2	3	4

**For the following questions please circle the number between 1 and 7 that best applies to you**

29. How would you rate your overall health during the past week?

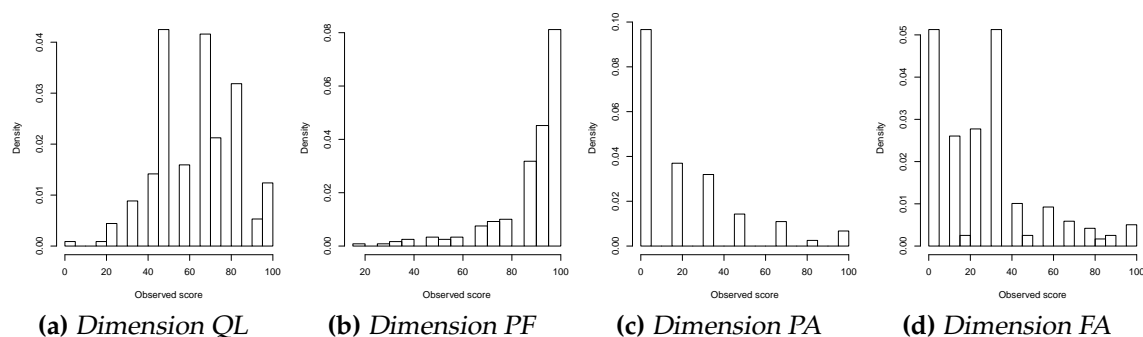
1      2      3      4      5      6      7  
Very poor      Excellent

30. How would you rate your overall quality of life during the past week?

1      2      3      4      5      6      7  
Very poor      Excellent



# Distributions des scores de QdV



**Figure B.1** – Histogrammes des scores de QdV à l'inclusion pour les données de l'essai clinique PRODIGE 5/ACCORD 17.





# Code pour le modèle de Diggle et Kenward

---

Dans le Chapitre 2, nous avons introduit le modèle de sélection de façon assez générale. Nous revenons plus en détail sur la mise en pratique de ce modèle.

Le modèle de sélection a notamment été utilisé dans l'article du Chapitre 3. En reprenant les notations de cet article, soit  $Y_i$  le vecteur des données de QdV et  $D_i$  la variable de sortie d'étude. On introduit en supplément la notation  $Y_i^{\text{obs}}$  qui est le vecteur des données observées de QdV et  $Y_i^{\text{mis}}$  qui est le vecteur des données manquantes. En utilisant la décomposition du modèle de sélection, la distribution des données observées sous l'hypothèse d'un mécanisme MNAR s'écrit ([Molenberghs and Kenward, 2007](#)) :

$$\begin{aligned} f(Y_i^{\text{obs}}, D_i) &= \int f(Y_i, D_i) dY_i^{\text{mis}} \\ &= \int f(Y_i) \times f(D_i | Y_i) dY_i^{\text{mis}} \end{aligned}$$

Cette écriture montre comment intervient la dépendance des données non observées sur le mécanisme à l'origine des sorties d'étude, avec un modèle marginal pour  $Y_i$  et un modèle pour le mécanisme de sortie d'étude conditionnellement aux

## Annexe C. Code pour le modèle de Diggle et Kenward

---

valeurs prises par  $Y_i$ .

Le modèle de [Diggle and Kenward \(1994\)](#) propose de combiner un LMM avec une régression logistique pour la modélisation des sorties d'étude. Afin de partager le travail réalisé pour l'estimation de ce modèle, nous mettons à disposition le code R qui permet le calcul de la vraisemblance selon ce modèle. Elle a ensuite été optimisée en utilisant la fonction `nlm` pour l'estimation des paramètres du maximum de vraisemblance.

```
#####
# FCT TO CALCULATE LIKELIHOOD SM -----
#####

# Function to compute the likelihood from the Selection model
# (see Diggle and Kenward, 1994)

# Assume we have patient $i = 1, \cdots, N$
# where a longitudinal outcome Y is measured at planned visit $j = 1, \cdots, n$.
# Data required to have a specific form to work detailed here.
# Data is the complete data (i.e. containing all observations - missing or not) with N*n rows.
# If the longitudinal outcome is missing, its value is 'NA'.
# The data must include the dropout indicator (T/F)
# for each patient to indicate if the patient is dropout or not
# The data must include the $D_i$ variable indicating the occasion at which dropout occurs
# We make the convention that $D_i = n+1$ for a complete sequence

# 'parameter' : a list with components where:

# 'beta'    representing the fixed-effects of the LMM (e.g. beta0, beta1, beta2)
# 'd'       representing the random var and cov effects (e.g. sd(b0), sd(b1), cov(b0, b1))
# 'sigma2'  representing the estimated within-group error variance
# 'psi'     representing the initialization of psi parameter in the logistic regression

# 'arg' : list of arguments to declare the variable name in the dataframe where:

# 'fit'     an object inheriting from class lme.
# 'y'       a character string indicating the outcome variable in the LMM (e.g. 'QL2')
# 'id'      a character string indicating the grouping variable (e.g. 'npat')
# 'visit'   a character string indicating the time variable in the LMM (e.g. 'visit')
# dropout   a character string indicating the indicator for dropout variable
# 'di'      a character string indicating the variable at which dropout occurs
# n.random  a number indicating the dimension of random effects used in the LMM

nllik <- function(parameter = arg$parameter, arg = arg){
```

---

```

#####
# OBTAIN ARGUMENTS
#####

# See description of the parameters
fit      <- arg$fit
y        <- arg$y
visit    <- arg$visit
id       <- names(fit$groups)
data     <- fit$data                # specific data with specific form to work
nind     <- length(unique(data[[id]])) # number of subject
N        <- nrow(data)              # N * nrow
n.random <- arg$n.random
dropout  <- arg$dropout
di       <- arg$di

# To verify if the planned visits are exactly the same
time     <- unique(data[[visit]])   # will be use to construct the var/cov matrix V
# Calculate the number of planned visits
maxvisit <- max(tapply(X = data[[visit]], INDEX = data[[id]], length))

#####
# MODEL PARAMETER
#####

parameter <- unlist(parameter)

# manipulate string of the name coefficient
# extract the name before '.' to obtain the name of coeff
string    <- sub("\\\\.\\*", "", names(unlist(arg$parameter)))

# Obtain the idx
idx.beta  <- which(str_detect(string, "beta") == T)
idx.d     <- which(str_detect(string, "d") == T)
idx.phi   <- which(str_detect(string, "phi") == T)
idx.sigma2 <- which(str_detect(string, "sigma2") == T)
idx.psi   <- which(str_detect(string, "psi") == T)
idx.tau2  <- which(str_detect(string, "tau2") == T)
idx.rho   <- which(str_detect(string, "rho") == T)

# values for the parameters to be optimized over
beta      <- NULL
d         <- NULL
phi       <- NULL
sigma     <- NULL
rho       <- NULL
tau2      <- NULL
sigma2    <- NULL

beta      <- if(length(idx.beta) > 0){parameter[idx.beta]}      # fixed-effects
d         <- if(length(idx.d) > 0){parameter[idx.d]}           # random var and cov effects
sigma2    <- if(length(idx.sigma2) > 0){parameter[idx.sigma2]}  # variance within-group error

```

## Annexe C. Code pour le modèle de Diggle et Kenward

---

```
psi      <- if(length(idx.psi) > 0){parameter[idx.psi]}      # logistic parameter
tau2     <- if(length(idx.tau2) > 0){parameter[idx.tau2]}    # tau*H_i where H_i
rho      <- if(length(idx.rho) > 0){parameter[idx.rho]}      # correlation H

# *** Statistical remark ***
# General form for Covariance matrix
# V_i = d Jn + sigma^2 In + tau^2 H_i (see SAS decomposition or Molenberghs, 2007)
# with Jn n*n matrix with 1, In n*n identity matrix and H_i autocorrelation matrix with rho
# However, we assuming independent error (tau^2 = 0)
# it reduces the expression of V_i

# Add names to the coefficients
names(beta) <- if(!is.null(beta)){names(unlist(arg$parameter))[idx.beta]}
names(d)    <- if(!is.null(d)){names(unlist(arg$parameter))[idx.d]}
names(rho)  <- if(!is.null(rho)){names(unlist(arg$parameter))[idx.rho]}
names(sigma2) <- if(!is.null(sigma2)){names(unlist(arg$parameter))[idx.sigma2]}
names(psi)  <- if(!is.null(psi)){names(unlist(arg$parameter))[idx.psi]}
names(tau2) <- if(!is.null(tau2)){names(unlist(arg$parameter))[idx.tau2]}

#####
# MEAN XB
#####

model.matf <- model.matrix.lm(fit, data = data, na.action = "na.pass")
mu         <- model.matf %*% beta
# mean
mu         <- data.frame(mu = mu, obs = as.numeric(rownames(mu)))

#####
# VARIANCE
#####

Z <- designZ(fit = fit) # design matrix RE

# construct variance-covariance matrix
V <- VarCov.lmm(d = d,
               tau2 = tau2,
               sigma2 = sigma2,
               rho = rho,
               hij = time, Z = Z, n.random = n.random)

# Matrix must be a square matrix
if(ncol(V) != nrow(V) | ncol(V) != N){
  print("Matrice Var/Cov non carree")
}
rownames(V) <- c(1:nrow(V))
rownames(V) <- c(1:nrow(V))

# Add constraints on the Std parameters >=0
```

---

```

# Test if at least one terme is negative
isEmpty <- function(x) {
  return(length(x)==0)
}

cond.d      <- any( (sapply(d[1:n.random], sign) == -1) == T)
cond.sigma2 <- any( (sapply(sigma2, sign) == -1) == T)
if(!isEmpty(rho)){
  cond.rho    <- rho < -1 | rho > 1
  if(cond.rho == T){
    return(1e10)
  }
}

# Penalize likelihood if not respect
if (cond.d == T | cond.sigma2 == T){
  return(1e10) # large value because we cant to minimise the -log likelihood
}

# IDX COMPLETERS VS DROPOUT
idxcompleter <- data[data[[dropout]] == F,][c(id, di, dropout)]
idxcompleter <- data.frame(idxcompleter, obs = as.numeric(rownames(idxcompleter)))

idxdropout  <- data[data[[dropout]] == T,][c(id, di, dropout)]
idxdropout  <- data.frame(idxdropout, obs = as.numeric(rownames(idxdropout)))

#####
# LIKELIHOOD DECOMPOSITION
#####

# Fct
integr <- function(yd, psi, mu, sigma, lastobs){
  n <- length(psi) # MCAR/MAR/MNAR
  g <- c(exp( psi[1:n] %*% (c(1, lastobs, yd)[1:n]) ) )
  if(g > 1e10){
    g <- 1e10
  }
  g <- g / (1+g)
  int <- g*dnorm(x = yd, mean = mu, sd = sigma)
  return(int)
}

# Gauss (L1)
ldensity1 <- function(yobs, ntime, muobs, sigmaobs){
  if(length(yobs) == 1){
    l11 <- dnorm(x = yobs, mean = muobs, sd = sqrt(sigmaobs), log = T)
  }
  if(length(yobs) > 1){
    l11 <- dmnorm(x = yobs, mean = muobs, sigma = sigmaobs, log = T)
  }
  return(l11)
}

```

## Annexe C. Code pour le modèle de Diggle et Kenward

---

```
# Logistic (L2)
ldensity2 <- function(x, y){
  # If missing data, compute 0 for the product
  y[which(is.na(y))] <- 0
  x[which(is.na(x))] <- 0
  n <- length(x) # MCAR/MAR/MNAR
  xy <- x[1:n] %*% y[1:n]
  g <- exp( xy )
  if(g > 1e10){
    g <- 1e10
  }
  g <- g / (g+1)
  if(g == 1){
    g <- g - 0.0001
  }
  ll <- log(1-g)
  return(ll)
}

# COMPLETERS

llc <- 0 # 1klh completers

# each contribution (i)
for(ind in unlist( unique(idxcompleter[id]) ) ){
  obs      <- idxcompleter$obs[idxcompleter[id] ==ind] # idx obs for ind
  ni       <- length(obs)                             # number obs for ind
  yobs     <- unlist(data[y])[obs]                     # for ind
  muobs    <- mu$mu[mu$obs %in% obs]
  idx.obs  <- as.numeric(rownames(V)) %in% obs
  Vobs     <- V[idx.obs, idx.obs]

  # ll2 logistique
  ll2 <- 0
  for(i in 2 : ni){
    ll2 <- cbind(ll2, ldensity2(x = psi, y = c(1, yobs[i-1], yobs[i])))
    ll2 <- sum(ll2) # sum obs for ind
  }

  # ll1 gauss
  ll1 <- 0
  # use mvnorm
  ll1 <- ldensity1(yobs = yobs, ntime = ni, muobs = muobs, sigmaobs = Vobs)
  # update 1klh completers
  llc <- cbind(llc, ll1 + ll2) # 1klh
}

# DROPOUT (di < ni)

lld <- 0 # 1klh completers
```

---

```

# each contribution (i)
for(ind in unlist(unique(idxdropout[id]))) ){
  di      <- unique(idxdropout$di[idxdropout[id] ==ind]) # time dropout
  obstheo <- idxdropout$obs[idxdropout[id] ==ind]        # obs if no dropout
  ntheo   <- length(obstheo)
  ytheo   <- unlist(data[y])[obstheo]
  obs     <- obstheo[!is.na(ytheo)]
  ni      <- length(obs)                                # number obs for ind
  yobs    <- unlist(data[y])[obs]                        # obs for ind
  muobs   <- mu$mu[mu$obs %in% obstheo]                  # before
  Vobs    <- V[obstheo, obstheo]                        # before

  ## l12 logistique
  l12 <- 0
  # if only one measurement: no action
  if(ni >= 2){
    for(i in 2 : ni){
      l12 <- cbind(l12, ldensity2(x = psi, y = c(1, yobs[i-1], yobs[i])))
      l12 <- sum(l12)
    }
  }
  ## l11 gauss
  l11 <- 0
  l11 <- ldensity1(yobs = yobs, ntime = ni,
    muobs = muobs[1:ni], sigmaobs = Vobs[1:ni, 1:ni])

  ## l13 marginal
  muobsdi <- muobs[di]
  vi11 <- Vobs[c(1:ni), c(1:ni)]
  vi22 <- Vobs[di, di]
  vi12 <- Vobs[c(1:ni), di]
  vi21 <- Vobs[di, c(1:ni)]
  ecurr <- muobsdi + vi21%% solve(vi11)%%(yobs - muobs[1:ni])
  vcurr <- vi22 - vi21%% solve(vi11) %% vi12

  # Integrate on the interval: Romberg method
  l13 = 0
  if(vcurr > 0){
    borninf <- c(ecurr - 2*sqrt(vcurr))
    bornsup <- c(ecurr + 2*sqrt(vcurr))
    # Not accurate - not used
    # int      <- integrate(integr, lower= borninf, upper=bornsup,
    #                        psi = psi ,
    #                        mu = ecurr, sigma = sqrt(vcurr),
    #                        lastobs = yobs[ni])
    # intval    <- int$value
    int      <- romberg(integr, a= borninf, b=bornsup,
                        psi = psi ,
                        mu = ecurr, sigma = sqrt(vcurr),
                        lastobs = yobs[ni])
    intval    <- int$value
    if(intval == 0){

```



## Annexe C. Code pour le modèle de Diggle et Kenward

---

```
        intval <- intval + 1e-5
      }
      l13 <- log(intval)
    }
    lld <- rbind(lld, l11 + l12 + l13) # lklh
  } # end loop

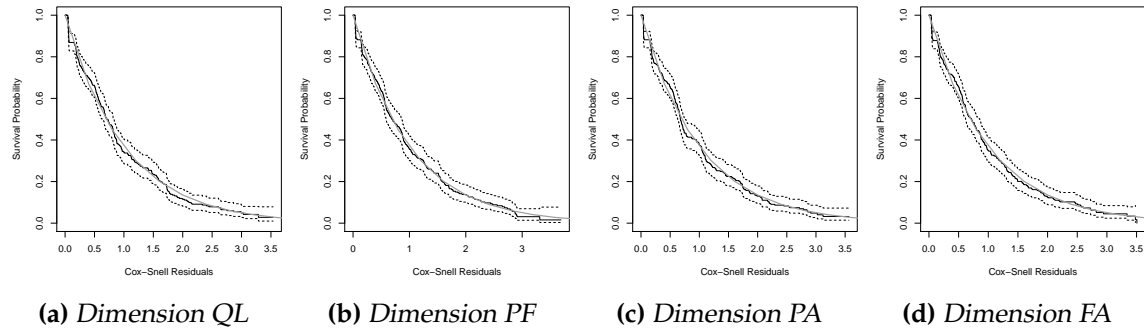
# Likelihood completers
ll  <- sum(llc, lld)
res <- - ll # - (Log)-Vraisemblance du modele
return(res)
}
```

# Annexe D

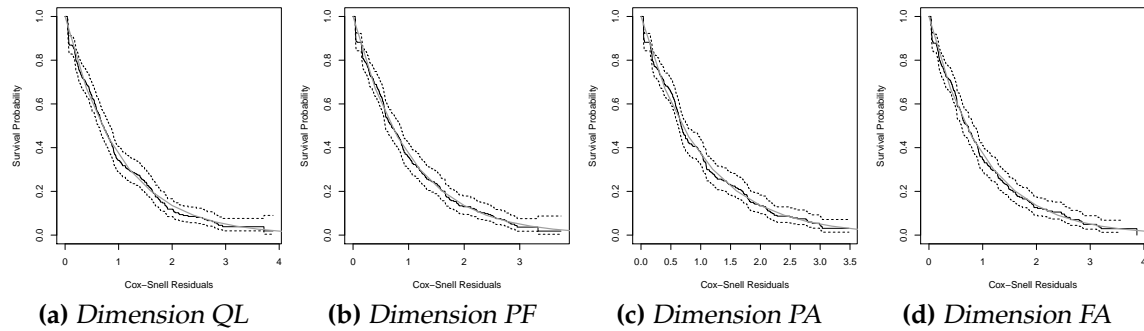
## **Etude des résidus du sous-modèle pour le temps jusqu'à sortie d'étude**

---

## D.1 Résidus de Cox-Snell

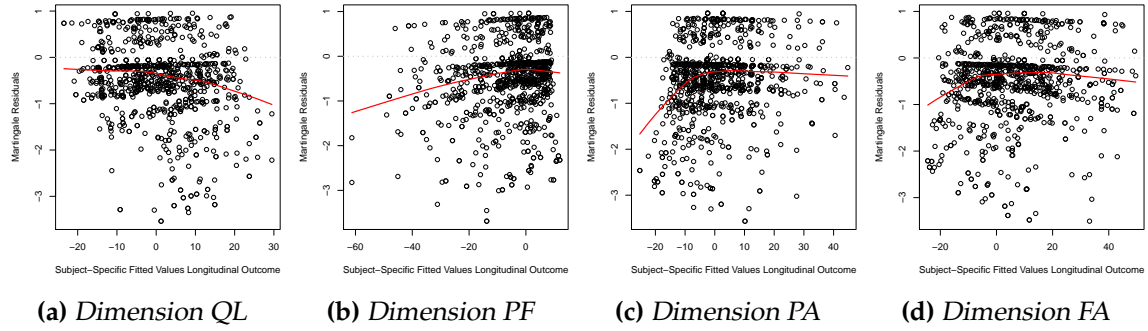


**Figure D.1** – Résidus de Cox-Snell du modèle conjoint linéaire. En noir est tracée la fonction de survie de Kaplan-Meier des résidus avec l'intervalle de confiance à 95% et en gris la fonction de survie théorique distribuée selon une loi exponentielle de paramètre 1.

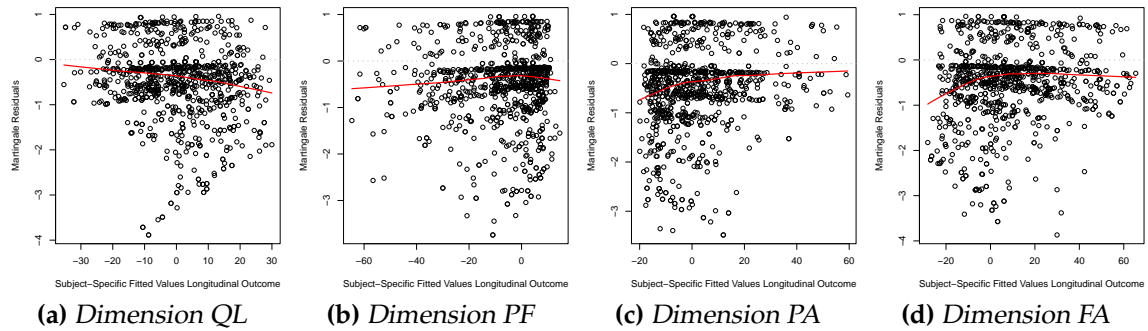


**Figure D.2** – Résidus de Cox-Snell du modèle conjoint à splines. En noir est tracée la fonction de survie de Kaplan-Meier des résidus avec l'intervalle de confiance à 95% et en gris la fonction de survie théorique distribuée selon une loi exponentielle de paramètre 1.

## D.2 Résidus martingales



**Figure D.3** – Résidus martingales du modèle conjoint linéaire avec la courbe de lissage en rouge estimée par une régression non-paramétrique locale.



**Figure D.4** – Résidus martingales du modèle conjoint à splines avec la courbe de lissage en rouge estimée par une régression non-paramétrique locale.



## Résultats de simulations pour le modèle conjoint à ratio de continuation

---

Nous avons mené une étude de simulation pour valider la méthode d'estimation du modèle conjoint à ratio de continuation. Nous avons généré 500 jeux de données selon le modèle Équation 5.3 avec 1000 patients répartis équitablement dans les deux bras de traitement.

La simulation des données a été inspirée des données de l'essai clinique PRODIGE 5/ACCORD 17. Les paramètres ont été choisis à partir des valeurs obtenues sur l'application du modèle pour la dimension PA. Pour réduire le temps de calcul et faciliter l'estimation du modèle, nous avons regroupé certaines catégories : les 7 catégories de score de la dimension PA (0, 16.66, 33.33, 50, 66.66, 83.83, 100) ont été regroupées selon les valeurs les plus proches en seulement 4 catégories (0, 33.33, 66.67, 100).

Le modèle est estimé avec le package `JMbayes` par méthode MCMC en choisissant 50000 itérations dont 5000 utilisées pour la période de chauffe. Les résultats moyens obtenus sur les 500 simulations sont résumés dans le Tableau E.1. Nous avons calculé pour chaque paramètre du modèle la valeur moyenne des estimations ainsi que son écart-type. L'étude de simulation montre que les paramètres

## Annexe E. Résultats de simulations pour le modèle conjoint à ratio de continuation

---

**Tableau E.1** – Résultats de simulations pour l'estimation du modèle conjoint à ratio de continuation en donnant la valeur moyenne estimée et l'écart-type des estimations pour chaque paramètre simulé.

Coefficient	Valeur simulée	Moyenne	Ecart-type
$\lambda_1$	-0.86	-0.81	1.26
$\lambda_2$	-1.76	-1.76	1.22
$\lambda_3$	3.44	3.56	1.13
$\beta_1$	0.44	0.42	0.23
$\beta_2$	-0.30	-0.30	0.17
$\rho_{01}$	-0.12	-0.10	0.08
$\sigma_0$	1.96	2.06	1.04
$\sigma_1$	1.80	1.81	0.17
$\alpha$	-0.02	-0.02	0.01
$\gamma$	-0.08	-0.09	0.09

sont relativement bien estimés.

# Modèle cumulatif pour données ordinales

---

## Modèle cumulatif

Parmi les modèles pour données ordinales, le plus connu et le plus utilisé est vraisemblablement le modèle cumulatif, introduit et rendu populaire par [McCullagh and Nelder \(1989\)](#). Il est basé sur les probabilités cumulées :

$$\omega_{ik}(t) = \Pr(Y_i(t) \leq k).$$

Les modèles cumulatifs de type logit comparent la probabilité d'obtenir au moins une certaine valeur de score (correspondant à la catégorie  $k$ ) avec la probabilité d'obtenir un score plus grand que cette même valeur. Pour modéliser les mesures répétées de score, nous supposons un modèle cumulatif à ordonnée à l'origine et pente aléatoire qui s'écrit :

$$\begin{aligned} \text{logit}(\omega_{ik}(t)) &= \log\left(\frac{\Pr(Y_i(t) \leq k)}{\Pr(Y_i(t) > k)}\right) = \lambda_k - \eta_i(t) \quad k = 1, \dots, K-1 \\ &= \lambda_k - (\beta_1 t + \beta_2 (t \times \text{arm}_i) + b_{0i} + b_{1i} t) \end{aligned}$$



où  $\beta_1$  est l'effet du temps et  $\beta_2$  l'effet d'interaction entre le bras de traitement et le temps, et  $b_i = (b_{0i}, b_{1i})$  les effets aléatoires supposés suivre une loi normale d'espérance zéro et de matrice de variance-covariance  $D = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}$ . Comme pour le modèle à ratio de continuation, les seuls coefficients qui dépendent de la catégorie sont les coefficients  $\lambda_k$  qui correspondent à des ordonnées à l'origine pour chaque catégorie de réponse. Ils sont contraints d'être ordonnés :

$$-\infty = \lambda_0 < \lambda_1 < \dots < \lambda_{K-1} < \lambda_K = +\infty.$$

Le signe négatif devant le prédicteur linéaire  $\eta_i(t)$  vise à faciliter l'interprétation du modèle, en faisant en sorte qu'à mesure que les variables explicatives augmentent (e.g. le temps), la probabilité d'obtenir une grande valeur de score augmente également (pourvu, bien sûr, que les coefficients associés soient positifs). C'est généralement la paramétrisation que l'on retrouve dans les logiciels pour estimer un modèle cumulatif.

## Interprétation avec une variable latente

Le modèle cumulatif suppose que les observations  $Y_i(t)$  proviennent de la discrétisation d'une variable latente (i.e. non observée) continue  $\tilde{Y}_i(t)$ . Dans ce modèle, les paramètres de seuils  $\lambda_k$  partitionnent  $\tilde{Y}_i(t)$  en  $K$  catégories ordonnées qui sont observées. La relation qui relie la variable latente continue et la variable réellement observée est la suivante :

$$Y_i(t) = k \quad \text{si} \quad \lambda_{k-1} < \tilde{Y}_i(t) \leq \lambda_k.$$

En supposant que :

$$\tilde{Y}_i(t) = \eta_i(t) + \varepsilon_i(t)$$

avec  $\varepsilon_i(t)$  un terme d'erreur distribué selon une loi logistique ([Agresti, 2010, chap3](#)), le modèle cumulatif peut s'interpréter comme un modèle de régression pour une variable latente continue ([Anderson and Philips, 1981](#); [Masyn et al.,](#)

**Tableau F.1** – Estimation du modèle cumulatif pour la modélisation des données de QdV de la dimension PA.

	Est	SE	$p$
Dimension PA (N = 254)			
$\lambda_1$	-0.492	0.128	< 0.001
$\lambda_2$	0.564	0.128	< 0.001
$\lambda_3$	1.631	0.144	< 0.001
$\lambda_4$	2.353	0.165	< 0.001
$\lambda_5$	3.297	0.205	< 0.001
$\lambda_6$	3.974	0.248	< 0.001
$\beta_1$	-0.031	0.021	0.141
$\beta_2$	-0.029	0.027	0.281

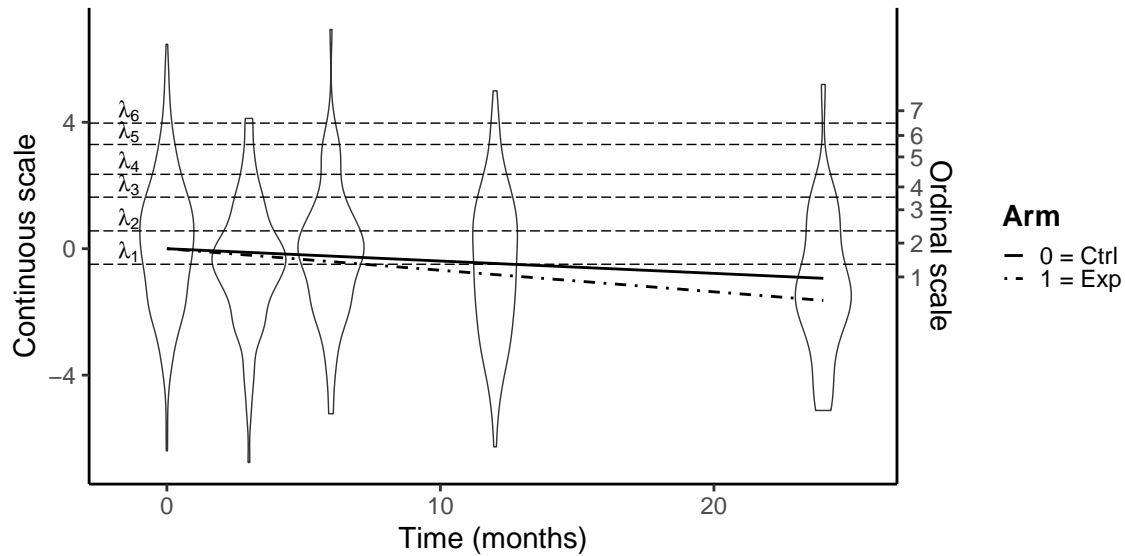
2014).

## Résultat de l'application

Nous avons appliqué le modèle cumulatif pour modéliser les scores de QdV considérés en tant que données ordinales. Nous avons utilisé le package `ordinal` (Christensen, 2015). Les résultats pour la dimension PA sont donnés dans le tableau Tableau F.1.

Les estimations du modèle cumulatif montrent qu'il n'y a pas d'effet du temps ( $\hat{\beta}_1 = 0.031, p = 0.141$ ). Il n'y a pas non plus d'effet d'interaction ( $\hat{\beta}_2 = -0.029, p = 0.281$ ), ainsi les probabilités cumulées évoluent de la même façon entre les deux bras de traitement. Les paramètres  $\hat{\lambda}_k$  ( $k = 1, \dots, 6$ ) sont bien ordonnés. Ils sont significatifs, mais ne doivent pas être interprétés en tant que tel puisqu'ils sont par construction différents de 0.

La Figure F.1 représente la distribution de la variable latente prédite à partir du modèle cumulatif pour le score PA. Les traits horizontaux correspondent aux paramètres  $\hat{\lambda}_k$  ( $k = 1, \dots, 6$ ) et permettent de faire le lien entre les valeurs de la variable latente et les catégories de scores qui sont modélisées. Cette représentation illustre bien l'idée que le score serait une quantité continue pour laquelle seulement quelques valeurs possibles sont observées (les catégories de scores sur



**Figure F.1** – Représentation de la variable latente à partir du modèle cumulatif pour le score PA avec 7 catégories de réponse où les traits horizontaux correspondent aux paramètres de seuils  $\lambda_k$  ( $k = 1, \dots, 6$ ) qui partitionnent l'échelle continue (à gauche) en 7 catégories ordonnées (à droite).

l'échelle ordinale). Cette représentation a l'avantage d'être plus facilement interprétable pour comprendre les effets des variables explicatives sur les valeurs du score.





# Références bibliographiques

---

- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., Haes, J. C. J. M. d., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Roze, P. B., Schraub, S., Sneeuw, K., Sullivan, M., and Takeda, F. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology. *JNCI: Journal of the National Cancer Institute*, 85(5):365–376. 5
- Administration, F. a. D. and others (2019). Clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry. 3
- Agresti, A. (2010). *Analysis of ordinal categorical data*, volume 656. John Wiley & Sons. 101, 103, 105, 144
- Alsefri, M., Sudell, M., García-Fiñana, M., and Kolamunnage-Dona, R. (2020). Bayesian joint modelling of longitudinal and time to event data: a methodological review. *BMC Medical Research Methodology*, 20(1):94. 102
- Anderson, J. A. and Philips, P. R. (1981). Regression, Discrimination and Measurement Models for Ordered Categorical Variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 30(1):22–31. Publisher: [Wiley, Royal Statistical Society]. 110, 144
- Andrinopoulou, E.-R., Eilers, P. H. C., Takkenberg, J. J. M., and Rizopoulos, D. (2018). Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. *Biometrics*, 74(2):685–693. Publisher: John Wiley & Sons, Ltd. 98
- Andrinopoulou, E.-R. and Rizopoulos, D. (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different associa-

## Références bibliographiques

---

- tion structures. *Statistics in Medicine*, 35(26):4813–4823. Publisher: John Wiley & Sons, Ltd. 48, 102
- Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J. M., and Lesaffre, E. (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine*, 33(18):3167–3178. 102, 109
- Armero, C., Forné, C., Rué, M., Forte, A., Perpiñán, H., Gómez, G., and Baré, M. (2016). Bayesian joint ordinal and survival modeling for breast cancer risk assessment. *Statistics in Medicine*, 35(28):5267–5282. Publisher: John Wiley & Sons, Ltd. 102, 111, 118
- Armstrong, B. G. and Sloan, M. (1989). Ordinal regression models for epidemiologic data. *American Journal of Epidemiology*, 129(1):191–204. tex.publisher: Oxford University Press. 105
- Barbieri, A., Anota, A., Conroy, T., Gourgou-Bourgade, S., Juzyna, B., Bonnetain, F., Lavergne, C., and Bascoul-Mollevi, C. (2016). Applying the Longitudinal Model from Item Response Theory to Assess Health-Related Quality of Life in the PRODIGE 4/ ACCORD 11 Randomized Trial. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 36(5):615–628. 117
- Barbieri, A., Peyhardi, J., Conroy, T., Gourgou, S., Lavergne, C., and Mollevi, C. (2017). Item response models for the longitudinal analysis of health-related quality of life in cancer clinical trials. *BMC medical research methodology*, 17(1):148. 111, 118
- Bascoul-Mollevi, C., Barbieri, A., Bourgier, C., Conroy, T., Chauffert, B., Hebbar, M., Jacot, W., Juzyna, B., De Forges, H., Gourgou, S., Bonnetain, F., Touraine, C., and Anota, A. (2020). Longitudinal analysis of health-related quality of life in cancer clinical trials: methods and interpretation of results. *Quality of Life Research*. 118
- Bascoul-Mollevi, C., Gourgou, S., Galais, M.-P., Raoul, J.-L., Bouché, O., Douillard, J.-Y., Adenis, A., Etienne, P.-L., Juzyna, B., Bedenne, L., and Conroy, T. (2017). Health-related quality of life results from the PRODIGE 5/ACCORD 17 randomised trial of FOLFOX versus fluorouracil–cisplatin regimen in oesophageal cancer. *European Journal of Cancer*, 84:239–249. publisher: Elsevier. 13

- Bender, R. and Benner, A. (2000). Calculating Ordinal Regression Models in SAS and S-Plus. *Biometrical Journal*, 42(6):677–699. 103
- Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., Mosconi, P., Osoba, D., Simes, J., and Hürny, C. (1998). Missing quality of life data in cancer clinical trials: serious problems and challenges. *Statistics in Medicine*, 17(5-7):517–532. 12
- Berridge, D. M. and Whitehead, J. (1991). Analysis of failure time data with ordinal categories of response. *Statistics in Medicine*, 10(11):1703–1710. Publisher: John Wiley & Sons, Ltd. 105
- Blanchin, M., Hardouin, J.-B., Le Neel, T., Kubis, G., and Sébille, V. (2011). Analysis of longitudinal patient-reported outcomes with informative and non-informative dropout: comparison of CTT and Rasch-based methods. *International Journal of Applied Mathematics & Statistics [Internet]*, 24:I–11. 101
- Blazeby, J., Conroy, T., Hammerlid, E., Fayers, P., Sezer, O., Koller, M., Arraras, J., Bottomley, A., Vickery, C., Etienne, P., and Alderson, D. (2003). Clinical and psychometric validation of an EORTC questionnaire module, the EORTC QLQ-OES18, to assess quality of life in patients with oesophageal cancer. *European Journal of Cancer*, 39(10):1384–1394. 6
- Bonnetain, F., Borg, C., Adams, R., Ajani, J., Benson, A., Bleiberg, H., Chibaudel, B., Diaz-Rubio, E., Douillard, J., Fuchs, C., Giantonio, B., Goldberg, R., Heinemann, V., Koopman, M., Labianca, R., Larsen, A., Maughan, T., Mitchell, E., Peeters, M., Punt, C., Schmoll, H., Tournigand, C., and de Gramont, A. (2017). How health-related quality of life assessment should be used in advanced colorectal cancer clinical trials. *Annals of Oncology*, 28(9):2077–2085. Publisher: Elsevier. 3
- Bottomley, A., Pe, M., Sloan, J., Basch, E., Bonnetain, F., Calvert, M., Campbell, A., Cleeland, C., Cocks, K., Collette, L., Dueck, A. C., Devlin, N., Flechtner, H.-H., Gotay, C., Greimel, E., Griebisch, I., Groenvold, M., Hamel, J.-F., King, M., Kluetz, P. G., Koller, M., Malone, D. C., Martinelli, F., Mitchell, S. A., Moinpour, C. M., Musoro, J. Z., O'Connor, D., Oliver, K., Piau-Louis, E., Piccart, M., Pimentel, F. L., Quinten, C., Reijneveld, J. C., Schürmann, C., Smith, A. W., Soltys, K. M., Sridhara, R., Taphoorn, M. J. B., Velikova, G., and Coens, C. (2018). Moving forward toward standardizing analysis of quality of life data in randomized cancer



## Références bibliographiques

---

- clinical trials. *Clinical Trials*, 15(6):624–630. tex.ids: Bottomley2018 publisher: SAGE Publications. 119
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25. tex.publisher: Taylor & Francis Group. 99
- Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73. tex.publisher: Wiley Online Library. 86
- Brown, H. and Prescott, R. (2014). *Applied mixed models in medicine*. John Wiley & Sons. 8, 100
- Calvert, M., Blazeby, J., Altman, D. G., Revicki, D. A., Moher, D., Brundage, M. D., and CONSORT PRO Group, f. t. (2013). Reporting of Patient-Reported Outcomes in Randomized Trials: The CONSORT PRO Extension. *JAMA*, 309(8):814–822. 12, 26, 119
- Calvert, M., Kyte, D., Mercieca-Bebber, R., Slade, A., Chan, A.-W., King, M. T., and and the SPIRIT-PRO Group (2018). Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *JAMA*, 319(5):483–494. 119
- Chakraborty, A. and Das, K. (2010). Inferences for joint modelling of repeated ordinal scores and time to event data. *Computational and mathematical methods in medicine*, 11(3):281–295. Place: United States. 111, 118
- Christensen, R. H. B. (2015). Analysis of ordinal data with cumulative link models—estimation with the R-package ordinal. *R-package version*, 28. 145
- Coens, C., Pe, M., Dueck, A. C., Sloan, J., Basch, E., Calvert, M., Campbell, A., Cleeland, C., Cocks, K., Collette, L., Devlin, N., Dorme, L., Flechtner, H.-H., Gotay, C., Griebisch, I., Groenvold, M., King, M., Kluetz, P. G., Koller, M., Malone, D. C., Martinelli, F., Mitchell, S. A., Musoro, J. Z., O'Connor, D., Oliver, K., Piault-Louis, E., Piccart, M., Quinten, C., Reijneveld, J. C., Schürmann, C., Smith, A. W., Soltys, K. M., Taphoorn, M. J. B., Velikova, G., and Bottomley, A. (2020). International

- standards for the analysis of quality-of-life and patient-reported outcome end-points in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. *The Lancet Oncology*, 21(2):e83–e96. Publisher: Elsevier. 119
- Conroy, T., Galais, M.-P., Raoul, J.-L., Bouché, O., Gourgou-Bourgade, S., Douillard, J.-Y., Etienne, P.-L., Boige, V., Martel-Lafay, I., Michel, P., Llacer-Moscardo, C., François, E., Créhange, G., Abdelghani, M. B., Juzyna, B., Bedenne, L., and Adenis, A. (2014). Definitive chemoradiotherapy with FOLFOX versus fluorouracil and cisplatin in patients with oesophageal cancer (PRODIGE5/ACCORD17): final results of a randomised, phase 2/3 trial. *The Lancet Oncology*, 15(3):305–314. publisher: Elsevier. 13
- Council, N. R. (2010). *The prevention and treatment of missing data in clinical trials*. The National Academies Press, Washington, DC. tex.ids: national2010prevention. 26
- Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265. tex.ids: cox1968general tex.publisher: Wiley Online Library publisher: Wiley Online Library. 93, 95
- Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M. G. (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical journal. Biometrische Zeitschrift*, 52(1):111–125. Place: Germany. 49
- Cuer, B., Mollevi, C., Anota, A., Charton, E., Juzyna, B., Conroy, T., and Touraine, C. (2020). Handling informative dropout in longitudinal analysis of health-related quality of life: application of three approaches to data from the esophageal cancer clinical trial PRODIGE 5/ACCORD 17. *BMC Medical Research Methodology*, 20(1):223. 28
- Curran, D., Molenberghs, G., Aaronson, N. K., Fossa, S. D., and Sylvester, R. J. (2002). Analysing longitudinal continuous quality of life data with dropout. *Statistical Methods in Medical Research*, 11(1):5–23. publisher: SAGE Publications Ltd STM. 47
- de Bock, E., Hardouin, J.-B., Blanchin, M., Le Neel, T., Kubis, G., Bonnaud-Antignac, A., Dantan, E., and Seville, V. (2016). Rasch-family models are

## Références bibliographiques

---

- more valuable than score-based approaches for analysing longitudinal patient-reported outcomes with missing data. *Statistical Methods in Medical Research*, 25(5):2067–2087. tex.publisher: SAGE Publications Sage UK: London, England. 101
- Demidenko, E. (2013). *Mixed models: theory and applications with R*. John Wiley & Sons. 100, 101
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22. Publisher: Wiley Online Library. 48
- Diggle, P. and Kenward, M. G. (1994). Informative Drop-Out in Longitudinal Data Analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):49–93. tex.ids: diggle1994informative publisher: [Wiley, Royal Statistical Society]. 47, 130
- Diggle, P. J., editor (2009). *Analysis of longitudinal data*. Number 25 in Oxford statistical science series. Oxford Univ. Press, Oxford, 2. ed., reprinted edition. OCLC: 551951384. 9
- Ding, J. and Wang, J.-L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*, 64(2):546–556. tex.publisher: Wiley Online Library. 86
- Dobson, A. and Henderson, R. (2003). Diagnostics for joint longitudinal and dropout time modeling. *Biometrics*, 59(4):741–751. tex.publisher: Wiley Online Library. 93
- Fagerland, M. W. and Hosmer, D. W. (2016). Tests for goodness of fit in ordinal logistic regression models. *Journal of Statistical Computation and Simulation*, 86(17):3398–3418. 103
- Fairclough, D. L. (2010). *Design and analysis of quality of life studies in clinical trials*. Chapman and Hall/CRC. 8, 51, 86
- Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC. 100, 101

- Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in medicine*, 15(15):1663–1685. Publisher: Wiley Online Library. 43
- Fayers, P., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., and Bottomley, A. (2001). *EORTC QLQ-C30 scoring manual*. European Organisation for Research and Treatment of Cancer, 3rd edition. 7
- Ferrandina, G., Petrillo, M., Mantegna, G., Fuoco, G., Terzano, S., Venditti, L., Marcellusi, A., De Vincenzo, R., and Scambia, G. (2014). Evaluation of quality of life and emotional distress in endometrial cancer patients: A 2-year prospective, longitudinal study. *Gynecologic Oncology*, 133(3):518–525. 10
- Fielding, S., Ogbuagu, A., Sivasubramaniam, S., MacLennan, G., and Ramsay, C. R. (2016). Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Quality of Life Research*, 25(12):2977–2983. 28
- Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2009). *Handbooks of modern statistical methods: Longitudinal data analysis*. New York: Taylor & Francis Group. 17
- Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D. F., and Meulders, M. (2005). Multiple imputation for model checking: completed-data plots with missing and latent data. *Biometrics*, 61(1):74–85. tex.publisher: Wiley Online Library. 95
- Group, W. (1993). Study Protocol for the World Health Organization Project to Develop a Quality of Life Assessment Instrument (WHOQOL). *Quality of Life Research*, 2(2):153–159. 4
- Harrell Jr, F. E. (2013). rms: Regression modeling strategies. R package version 4.0-0. City. 105
- Harrell Jr, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer. 98, 103, 105
- Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480. tex.ids: henderson2000joint tex.publisher: Oxford University Press. 43

## Références bibliographiques

---

- Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, 16(1). 119
- Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2018). joinerML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC Medical Research Methodology*, 18(1). 119
- Hunger, M., Döring, A., and Holle, R. (2012). Longitudinal beta regression models for analyzing health-related quality of life scores over time. *BMC Medical Research Methodology*, 12(1):144. 118
- INCa, R. (2011). La situation du Cancer en France en 2011. 1
- Joly, F., Hilpert, F., Okamoto, A., Stuart, G., Ochiai, K., and Friedlander, M. (2017). Fifth Ovarian Cancer Consensus Conference of the Gynecologic Cancer Inter-Group: Recommendations on incorporating patient-reported outcomes in clinical trials in epithelial ovarian cancer. *European Journal of Cancer*, 78:133–138. 3
- Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons. 43
- Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in medicine*, 17(23):2723–2732. tex.publisher= Wiley Online Library. 47
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974. tex.publisher: JSTOR. 8
- Lawrence Gould, A., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., and Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in medicine*, 34(14):2181–2195. Citation Key Alias: lawrence\_gould\_joint\_2015 tex.publisher: [object Object]. 45
- Li, N., Elashoff, R. M., Li, G., and Saver, J. (2010). Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Statistics in Medicine*, 29(5):546–557. Publisher: John Wiley & Sons, Ltd. 111, 118

- Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83(404):1198–1202. 17
- Lunn, D. J., Wakefield, J., and Racine-Poon, A. (2001). Cumulative logit models for ordinal data: a case study involving allergic rhinitis severity scores. *Statistics in Medicine*, 20(15):2261–2285. Publisher: John Wiley & Sons, Ltd. 103
- Luo, S. (2014). A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statistics in Medicine*, 33(4):580–594. Publisher: John Wiley & Sons, Ltd. 111
- Masyn, K. E., Petras, H., and Liu, W. (2014). Growth Curve Models with Categorical Outcomes. In Bruinsma, G. and Weisburd, D., editors, *Encyclopedia of Criminology and Criminal Justice*, pages 2013–2025. Springer New York, New York, NY. 110, 144
- McCullagh, P. and Nelder, J. (1989). *Generalized linear models, second edition*. Chapman & Hall/CRC monographs on statistics & applied probability. Taylor & Francis. tex.lccn: 99013896. 100, 143
- McCulloch, C. E. and Searle, S. R. (2001). *Generalized, linear, and mixed models*. Wiley series in probability and statistics. Applied probability and statistics section. John Wiley & Sons, New York. 100
- Molenberghs, G. (2014). *Handbook of Missing Data Methodology*. Chapman and Hall/CRC, 1 edition. 17
- Molenberghs, G. and Kenward, M. (2007). *Missing data in clinical studies*, volume 61. John Wiley & Sons. 17, 47, 49, 129
- Molenberghs, G., Michiels, B., Kenward, M. G., and Diggle, P. J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica*, 52(2):153–161. 46
- Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313. 47, 99

## Références bibliographiques

---

- Njagi, E. N., Molenberghs, G., Kenward, M. G., Verbeke, G., and Rizopoulos, D. (2014). A characterization of missingness at random in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data, and sensitivity analysis: A characterization of missingness at random. *Biometrical Journal*, 56(6):1001–1015. 49
- O’Connell, A. (2006). *Logistic Regression Models for Ordinal Response Variables*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America. 103
- Organization, W. H., Organization, W. H., and others (1948). Constitution in basic documents. *Geneva: World Health Organization*. 3
- Pallis, A. G., Ring, A., Fortpied, C., Penninckx, B., Van Nes, M. C., Wedding, U., vonMinckwitz, G., Johnson, C. D., Wyld, L., Timmer-Bonte, A., Bonnetain, F., Repetto, L., Aapro, M., Luciani, A., and Wildiers, H. (2011). EORTC workshop on clinical trial methodology in older individuals with a diagnosis of solid tumors. *Annals of Oncology*, 22(8):1922–1926. 3
- Pan, A.-W., Chen, Y.-L., Chung, L.-I., Wang, J.-D., Chen, T.-J., and Hsiung, P.-C. (2012). A longitudinal study of the predictors of quality of life in patients with major depressive disorder utilizing a linear mixed effect model. *Psychiatry Research*, 198(3):412–419. 10
- Papageorgiou, G., Mauff, K., Tomer, A., and Rizopoulos, D. (2019). An Overview of Joint Modeling of Time-to-Event and Longitudinal Outcomes. *Annual Review of Statistics and Its Application*, 6(1):223–240. 119
- Papageorgiou, G. and Rizopoulos, D. (2020). An alternative characterization of MAR in shared parameter models for incomplete longitudinal data and its utilization for sensitivity analysis. *Statistical Modelling*, page 1471082X20927114. Publisher: SAGE Publications India. 49
- Paul H. C. Eilers and Marx, B. D. (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science*, 11(2):89–102. Publisher: Institute of Mathematical Statistics. 98

- Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., and Schmid, M. (2019). A review of spline function procedures in R. *BMC Medical Research Methodology*, 19(1):46. 86, 98
- Philipps, V., Amieva, H., Andrieu, S., Dufouil, C., Berr, C., Dartigues, J.-F., Jacqmin-Gadda, H., and Proust-Lima, C. (2014). Normalized mini-mental state examination for assessing cognitive change in population-based brain aging studies. *Neuroepidemiology*, 43(1):15–25. Publisher: Karger Publishers. 117
- Pinheiro, J., Bates, D., and others (2000). *Mixed-effects models in s and s-plus*. Springer Science & Business Media. 10
- Proust-Lima, C., Séne, M., Taylor, J. M., and Jacqmin-Gadda, H. (2014). Joint latent class models for longitudinal and time-to-event data: A review. *Statistical methods in medical research*, 23(1):74–90. Publisher: Sage Publications Sage UK: London, England. 43
- Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software, Articles*, 35(9):1–33. tex.ids: rizopoulos2010jm. 47
- Rizopoulos, D. (2012a). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis*, 56(3):491–501. tex.publisher: Elsevier. 48, 86
- Rizopoulos, D. (2012b). *Joint models for longitudinal and time-to-event data: With applications in R*. Chapman and Hall/CRC. 45, 95
- Rizopoulos, D. (2016). The r package JMBayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software, Articles*, 72(7):1–46. tex.ids: rizopoulos2014r. 119
- Rizopoulos, D. (2019). GLMMadaptive: Generalized linear mixed models using adaptive gaussian quadrature. *R package version 0.5-1*. URL: <https://CRAN.R-project.org/package=GLMMadaptive>. tex.ids: rizopoulos2019glmmdaptive. 117
- Rizopoulos, D. and Ghosh, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30(12):1366–1380. 86



## Références bibliographiques

---

- Rizopoulos, D., Verbeke, G., and Lesaffre, E. (2009). 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. 48
- Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2010). Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics*, 66(1):20–29. tex.publisher: Wiley Online Library. 93
- Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63(3):581–592. publisher: [Oxford University Press, Biometrika Trust]. 17
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge university press. 86
- Santos Nobre, J. and da Motta Singer, J. (2007). Residual Analysis for Linear Mixed Models. *Biometrical Journal*, 49(6):863–875. 93
- Sargent, D. J., Wieand, H. S., Haller, D. G., Gray, R., Benedetti, J. K., Buyse, M., Labianca, R., Seitz, J. F., O’Callaghan, C. J., Francini, G., Grothey, A., O’Connell, M., Catalano, P. J., Blanke, C. D., Kerr, D., Green, E., Wolmark, N., Andre, T., Goldberg, R. M., and De Gramont, A. (2005). Disease-Free Survival Versus Overall Survival As a Primary End Point for Adjuvant Colon Cancer Studies: Individual Patient Data From 20,898 Patients on 18 Randomized Trials. *Journal of Clinical Oncology*, 23(34):8664–8670. 3
- Schaake, W., de Groot, M., Krijnen, W. P., Langendijk, J. A., and van den Bergh, A. C. (2013). Quality of life among prostate cancer patients: A prospective longitudinal population-based study. *Radiotherapy and Oncology*, 108(2):299–305. 10
- Stone, C. J. and Koo, C.-Y. (2015). Additive splines in statistics. 98
- Stroup, W. W. (2012). *Generalized linear mixed models: modern concepts, methods and applications*. CRC press. 100, 101
- Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160. tex.publisher: Oxford University Press. 93, 95
- Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., and Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics*, 3(2):245–265. 46, 49

- Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, pages 809–834. Citation Key Alias: tsiatis\_joint\_nodate tex.publisher: [object Object]. 43
- Verbeke, G. and Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. Springer New York. tex.lccn= 00026596. 10, 17, 46, 93
- Verbeke, G., Molenberghs, G., and Beunckens, C. (2008). Formal and informal model selection with incomplete data. *Statistical Science*, pages 201–218. tex.publisher: JSTOR. 93
- Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M. G. (2001). Sensitivity Analysis for Nonrandom Dropout: A Local Influence Approach. *Biometrics*, 57(1):7–14. publisher: [Wiley, International Biometric Society]. 49
- Wu, L., Liu, W., Yi, G. Y., and Huang, Y. (2011). Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues. *Journal of Probability and Statistics*, 2012:640153. Publisher: Hindawi Publishing Corporation. 43





## **Modèles conjoints pour l'analyse longitudinale de la qualité de vie relative à la santé en présence de sorties d'étude informatives en oncologie**

**Résumé :** La qualité de vie relative à la santé (QdV) est de plus en plus utilisée comme critère de jugement dans les essais cliniques en oncologie. Elle est évaluée par le biais d'auto-questionnaires transmis aux patients à différentes visites au cours de leur prise en charge. Cependant, ces questionnaires sont souvent en partie incomplets, voire complètement manquants ; en particulier, chez les patients en situation palliative où les questionnaires non remplis pourraient être le résultat d'une progression de la maladie ou d'un décès. Les modèles linéaires mixtes (LMMs) sont habituellement utilisés pour l'analyse longitudinale de la QdV, mais peuvent donner des estimations biaisées en présence de données manquantes informatives, i.e. si le mécanisme de données manquantes est lié aux données manquantes de QdV. L'objectif de ce travail a été d'étudier différentes stratégies pour analyser les données longitudinales de QdV tout en tenant compte des sorties d'étude informatives. Dans l'ensemble du manuscrit, les modèles ont été illustrés sur les données de l'essai clinique PRODIGE 5/ACCORD 17, qui a inclus 267 patients atteints d'un cancer de l'œsophage avancé. Tout d'abord, trois modèles – le modèle de sélection (SM), le modèle de mélange de profils (PMM) et le modèle à effets partagés (SPM) – ont été étudiés. Le SPM (ou modèle conjoint), composé d'un sous-modèle pour les trajectoires de QdV (LMM) associé à un sous-modèle pour le temps jusqu'à sortie d'étude (modèle de survie), semble être l'approche la plus intéressante aussi bien d'un point de vue théorique que pratique. Ensuite, nous avons étudié l'intérêt de considérer un sous-modèle à risques concurrents pour le temps jusqu'à sortie d'étude, afin d'analyser l'évolution de la QdV en présence de deux causes de sortie d'étude. La comparaison d'un modèle conjoint standard par rapport à un modèle conjoint à risques concurrents a été réalisée en considérant plusieurs scénarios au travers d'une étude de simulations. Nous avons montré que le modèle conjoint à risques concurrents pouvait être utile pour faire la distinction entre des sorties d'étude informatives et non informatives. Enfin, nous avons envisagé d'autres modélisations du sous-modèle pour les trajectoires des scores de QdV plus adaptées aux spécificités associées aux données de score de QdV. Nous avons utilisé des splines pour modéliser des trajectoires non linéaires de QdV, et des modèles linéaires mixtes généralisés (GLMMs), pour considérer le score de QdV en tant que variable ordinale.

**Mots clés :** Modèle conjoint ; Analyse longitudinale ; Qualité de vie relative à la santé ; Oncologie ; Essais cliniques.

### **Joint models for the longitudinal analysis of the health-related quality of life in presence of informative dropouts in oncology**

**Abstract:** The use of health-related quality of life (HRQoL) as an endpoint in cancer clinical trials is growing rapidly. HRQoL is assessed at different visit times throughout the care process by self-administered questionnaires. However, these questionnaires are frequently incomplete (partially or entirely), and, among patients in palliative care, this could unfortunately be the consequence of dropouts related to disease progression or death. Linear mixed models (LMMs) are generally used to analyze longitudinal data of HRQoL, but they are likely to produce biased estimates in the presence of informative missing data, i.e., if missingness is correlated with the missing HRQoL outcome. The objective of this work was to study modeling alternatives to analyze longitudinal data of HRQoL while accounting for informative dropouts. Throughout the manuscript, all the models were illustrated on the clinical trial PRODIGE 5/ACCORD 17 including 267 patients with advanced esophageal cancer. First, three statistical models – the selection model (SM), the pattern-mixture model (PMM), and the shared-parameters model (SPM) – were investigated. The SPM, or joint model, which is composed of a sub-model for the HRQoL trajectory (LMM) linked with a sub-model for the time-to-dropout (survival model), appeared to be the most interesting from both theoretical and practical viewpoints. Then, we studied the interest of using a competing risks sub-model for the time-to-dropout to analyze HRQoL longitudinal data in the presence of two causes of dropout. The impact of using a standard instead of a competing risks joint model in various situations was investigated in a simulation study. We showed that the competing risks joint model could be useful to distinguish between informative and non-informative dropouts. Finally, we explored how the sub-model for the HRQoL score trajectory could be more adapted to the specificities of the HRQoL score data. We used splines to fit nonlinear trajectories and consider generalized linear mixed models (GLMMs) to treat the HRQoL score as an ordinal variable.

**Keywords:** Joint model; Longitudinal analysis; Health-related quality of life; Oncology; Clinical trials.