



HAL
open science

Méthodes longitudinales pour l'analyse de la qualité de vie relative à la santé en cancérologie

Antoine Barbieri

► **To cite this version:**

Antoine Barbieri. Méthodes longitudinales pour l'analyse de la qualité de vie relative à la santé en cancérologie. Applications [stat.AP]. Université Montpellier, 2016. Français. NNT : 2016MONT026 . tel-02182037

HAL Id: tel-02182037

<https://theses.hal.science/tel-02182037>

Submitted on 12 Jul 2019

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

Délivré par l'Université de Montpellier

Préparée au sein de l'école doctorale
Information Structure et Système

Et de l'unité de recherche
Institut Montpellierain Alexander Grothendieck
Unité de Biométrie de l'Institut régional du Cancer Montpellier

Spécialité: Biostatistique

Présentée par **Antoine BARBIERI**

Méthodes longitudinales pour
l'analyse de la Qualité de Vie
Relative à la Santé en
Cancérologie

Soutenue le 27 Juin 2016 devant le jury composé de

Julien JACQUES	Professeur	Université Lyon 2	Rapporteur
Christian LAVERGNE	Professeur	Université Paul-Valéry Montpellier 3	Directeur
Jean-Michel MARIN	Professeur	Université de Montpellier	Examineur
Caroline MOLLEVI	Biostatisticienne, HDR	Institut régional du Cancer Montpellier	Co-directeur
Bruno PEREIRA	Biostatisticien	CHU Clermont-Ferrant	Examineur
Cécile PROUST-LIMA	CR Inserm, HDR	Université de Bordeaux	Rapporteur



Remerciements

Comme les philosophes utilisent leur plume pour communiquer leurs pensées, comme les chanteurs saluent leur public à travers leur micro, je prends mon ordinateur pour rédiger de sincères remerciements.

Je dédie ces premières lignes à mes deux encadrants de thèse, Caroline et Christian, qui m'ont fait confiance pour ce projet. Ma reconnaissance et mon amitié envers eux ont grandi au fil de ces 3 années à leur coté. Ils forment un duo de professionnels avertis et experts complémentaires et m'ont proposé un cadre de travail plus que confortable. Appartenir au monde universitaire tout en travaillant au sein d'une unité de Biométrie dans une clinique a été une chance pour moi de mêler recherche et application. Je leur exprime au travers de ces quelques lignes toute ma gratitude pour m'avoir laissé une certaine liberté dans l'orientation de ce travail, qui m'a permis d'étudier et d'appréhender de nombreux modèles statistiques jusqu'alors méconnus pour moi. Je tiens particulièrement à souligner la présence dont a fait preuve Caroline, et la remercie vivement pour sa rigueur et notamment pour ses relectures attentives de mes diverses rédactions. Quant à Christian, il a toujours répondu présent quand j'en avais besoin. D'un point de vue plus personnel, nous avons réussi mutuellement à nous apprivoiser au fil de ces années, et aujourd'hui j'éprouve un attachement certain pour eux. Quant à leur marque d'affection, je me souviendrai toujours du clin d'œil discret de Christian, ou encore du souci quotidien de Caroline concernant mon bien être au travail ou de son fameux « *Tu me fais rire ...* ».

Je tiens également à remercier Cécile Proust-Lima et Julien Jacques qui m'ont fait l'honneur de rapporter cette thèse. Leur rapport montre l'implication, la rigueur dont ils ont fait preuve dans la lecture de ce travail, et je leur en suis très reconnaissant. Toujours concernant le jury, je remercie Bruno Pereira d'avoir accepté sans hésitation et avec un enthousiasme manifeste de faire partie de ce jury. Comme dernier membre de ce jury, je tiens chaleureusement à remercier Jean-Michel Marin. Ses conseils, son soutien, sa franchise et sa présence tout au long de mes années de formation me touchent énormément et m'ont beaucoup appris. Enfin, j'adresse une pensée particulière aux deux membres réguliers de mon comité de suivi de thèse : Catherine Trottier et Carine Bellera. Catherine, ta rigueur, tes remarques et le temps que tu m'as accordé m'ont été très bénéfiques. Carine, je te remercie pour tes conseils avisés.

Durant ces trois années, j'ai partagé la vie de deux instituts : l'IMAG et l'ICM. Cela représente énormément de personnes que je ne pourrai citer individuellement. Parmi tous les membres de l'IMAG, mes premiers mots seront pour les doctorants avec qui j'ai partagé de très bons moments Angéline, Christian, Christophe, Coralie, Étienne, Guillaume, Gothier, Jean, Ju, Julien, Jocelyn, Joubine, Micka, Mike, Myriam, Paul, Pierre, Quentin, Théo... J'ai également une pensée particulière pour tous les membres de l'équipe EPS¹. Parmi ces permanents, je remercie Sophie et Berna-

1. Rien à voir avec le sport en collège/lycée !

dette qui m'ont facilité la vie. Je pense naturellement à Ali, Jean-Michel et Mathieu avec qui je garde d'excellentes relations depuis leur enseignement ou encadrement au cours de mes années de master, ils assurent toujours leur rôle de conseiller et répondent présents à chacune de mes sollicitations.

J'en viens maintenant aux personnes de l'ICM, mon camp de base professionnel. Chaleureusement accueilli dans cette unité de Biométrie, je remercie tous les membres (la saisie, les data et les statisticiens) qui m'ont soutenu, supporté et qui ont même résisté à mes blagues quotidiennes en tout genre. Un remerciement particulier pour mes collègues de bureau (Caroline, Lise, Simon et Yassine) qui étaient en première ligne, mais également Sophie, responsable de l'unité, qui m'ont grandement facilité la vie durant ces trois années. Je quitte l'ICM avec bien plus que des collègues, avec des amis. Je pense aux plus récents, Marguito et Coco, mais surtout à Miss Luis (la pâtissière), Ju (le tennisman) ou Kiki (le serial pêcheur/motard) avec qui j'ai partagé de nombreuses activités extra-professionnelles et qui ont le plus subi de blagues et surtout les avants premières de mes communications orales !

Je tiens également à souligner la présence particulière de jeunes chercheurs avec qui j'ai eu la chance de collaborer et d'avancer professionnellement. Myriam, merci de m'avoir initié aux modèles à équations structurelles. Jean, merci de m'avoir aidé à prendre du recul sur la première partie de ma thèse. Amélie, merci de m'avoir servi d'exemple dans l'acharnement au travail et pour tous les échanges, discussions diverses autour de notre sujet commun. Parmi ces jeunes chercheurs, il m'est impossible de ne pas dire deux mots sur Marc et Pierre, précurseurs de l'alliance plaisir-travail ! Suite à nos apports mutuels nombreux et fructueux au sein de nos recherches, ils sont devenus aujourd'hui de très bons amis, amitiés scellées au détour de parties de coinches et de sorties mythiques. Donc merci à eux qui ont été présents dès le début de cette aventure !

J'en viens maintenant à mes proches, que je ne peux omettre dans ces remerciements ! Tout à commencer le 10 Avril 1989, date à laquelle j'ai vu les trois personnes les plus importantes de ma vie : ma mère, ma sœur et mon père. Ces trois personnes sont toujours là, en silence à quelques kilomètres mais plus proches que n'importe qui, ou presque. Mon père est présent dans sa discrétion. Pour ma sœur, je suis SON frère et l'amour fraternel qu'elle me porte quotidiennement me réchauffe tout le temps. Et puis ma mère, sûrement ma plus grande inspiration dans la construction de ma personnalité, et dans l'accomplissement de la personne que je suis et en devenir. Elle tient en partie de mon grand-père, un petit homme par la taille mais grand homme par ses valeurs et ses idées ! Leur présence et leur soutien à toute épreuve m'ont permis d'accomplir des projets qui m'étaient inespérés. Bien entendu, je ne me suis pas construit seulement grâce à ces trois géants de ma vie. En effet, je tiens à souligner la participation d'un ensemble de personnes formidables que j'ai eu l'honneur de rencontrer jusqu'à présent. Je profite de ces quelques lignes libres d'expression pour compléter cette faible reconnaissance écrite. Parmi ma famille, je compte également des personnes qui ne partagent ni mon nom, ni mon sang. Je commencerai par ma marraine, une personne fascinante, pleine de vie, de réflexion et de création. Mes amis (très) proches, que ce soit Fabien (mon frérot), Matou (mon aventurier), Seb (mon

éternel colloque), ou encore Malouch, Mel, Jo, Rico, Royou (mon rédacteur scientifique préféré) et certains handballeurs qui ont fait bien plus que de me supporter. Il m'est difficile de remercier personnellement ces personnes car sinon l'aspect scientifique de cette thèse paraîtrait trop petit. J'espère qu'ils s'y reconnaîtront à travers ces quelques mots, car eux aussi ont contribué à me construire. J'ai essayé de m'inspirer des qualités de chacun qui ont retenu mon attention et admiration. Je finirai ces remerciements avec Ambre qui partage ma vie depuis maintenant de nombreuses années. Je te remercie tout particulièrement pour ton soutien, tes conseils avisés et également pour me supporter au quotidien et continuer à mes côtés. « *Même si je ne sais pas où va mon chemin, je sais que je marche mieux quand ma main tient la sienne.* »

Pour finir, j'exprime toute ma gratitude à l'Institut de Recherche en Santé Publique et au Site de Recherche Intégrée sur le Cancer (SIRIC) de Montpellier pour leur soutien financier qui ont permis la réalisation de ce travail.

Table des matières

1	Introduction	1
1.1	Contexte	2
1.2	La qualité de vie relative à la santé (HRQoL)	2
1.2.1	Place du critère de HRQoL dans les essais cliniques	2
1.2.2	Définition de la HRQoL	4
1.3	Mesure de la HRQoL	5
1.3.1	Groupe EORTC	7
1.3.2	Groupe FACT	8
1.3.3	Approche classique dite du « vrai score »	10
1.4	État de l’art des approches statistiques pour l’analyse de la HRQoL dans les essais cliniques en cancérologie	12
1.4.1	Modèles linéaires mixtes	12
1.4.2	Modèle de survie : temps jusqu’à la détérioration du score	14
1.5	Problématiques de l’analyse longitudinale de la HRQoL	16
1.5.1	Les données manquantes	16
1.5.2	Le phénomène de « Response Shift »	17
1.5.3	Limites du score de HRQoL	18
1.5.4	Multiplicité des tests	19
1.6	Objectifs	20
I	La théorie de réponse aux items pour l’analyse longitudinale de la HRQoL	23
2	Présentation et application sur données réelles	25
2.1	Introduction	26
2.2	Présentation des modèles IRT	28
2.2.1	Modèles IRT pour données dichotomiques	28
2.2.2	Modèles IRT pour données polytomiques	31
2.2.3	Variantes et paramétrisation	34
2.3	Application	37
2.3.1	Résultats et estimations des modèles	38
2.3.2	Sensibilité des estimations du modèle LPCM suivant les observations	41
2.4	Discussion	41
2.5	Article : <i>Applying the longitudinal model from item response theory to assess the health-related quality of life in the PRODIGE 4/ACCORD11 randomized trial</i>	42

3	Sélection conceptuelle de modèle(s) IRT	57
3.1	Introduction	58
3.2	GLMM pour données catégorielles	59
3.2.1	Modèle cumulatif	62
3.2.2	Modèle adjacent	63
3.2.3	Modèle séquentiel	64
3.3	Sélection conceptuelle	65
3.4	Discussion	68
3.5	Article : <i>Selection of item response theory models for the longitudinal analysis of health-related quality of life in cancer clinical trials</i>	69
 II Méthodes complémentaires pour la prise en compte de nouvelles structures dans l'étude de la HRQoL		91
4	Étude globale et longitudinale par modèles à équation structurelle et à effets aléatoires	93
4.1	Introduction	94
4.2	Approche par modèles à équation structurelle	96
4.2.1	Estimation du modèle via l'algorithme EM	99
4.2.2	Analyse de sensibilité par simulations	107
4.3	Application	114
4.3.1	Résultats	115
4.3.2	Compléments sur les données réelles	116
4.4	Discussion	117
4.5	Article : <i>EM algorithm estimation of a structural equation model for the longitudinal study of the quality of life</i>	119
5	Étude longitudinale mono-dimensionnelle par un mélange de modèles mixtes	141
5.1	Introduction	142
5.2	Modèles de mélange	143
5.2.1	Approche IRT	146
5.2.2	Approche CTT pour données gaussiennes	147
5.2.3	Approche CTT pour données ordinales	148
5.3	Application	149
5.3.1	L'étude « Moral »	149
5.3.2	Classification	152
5.3.3	Étude de la classification <i>a posteriori</i>	159
5.4	Discussion	169
5.5	Compléments	172
5.5.1	Estimation par l'algorithme EM pour un mélange de LMM	172
5.5.2	Résultats additionnels de l'étude « Moral »	177

III Conclusions et perspectives	181
Annexes	197
A Liste des abréviations	199
B Questionnaire EORTC QLQ-C30	201
C Productions scientifiques	205

Table des figures

1.1	Schéma représentant les trois phases cliniques en cancérologie en fonction du patient (évolution de sa maladie) et du critère étudié.	3
1.2	Schéma représentant la construction d'un questionnaire psychométrique. Les concepts non directement mesurables sont dans des cercles. Les observations découlent des items caractérisés par des rectangles.	6
1.3	Schéma représentant le questionnaire EORTC QLQ-C30 avec les dimensions associées. Les dimensions de gauche sont les dimensions fonctionnelles et les dimensions de droite sont les dimensions symptomatiques.	8
1.4	Dimension du bien-être physique relative au questionnaire FACT-G (version 4).	10
1.5	Dimension physique du questionnaire EORTC QLQ-C30. Les cercles bleus et verts présentent les réponses fictives de deux individus. Les deux scores obtenus sont identiques et sont égaux à 11.	19
2.1	Continuum lié au concept de vulnérabilité. Les réponses aux questions sont déterminées par le niveau de l'individu sur ce continuum. Les différents items et réponses observées concernant la dimension de vulnérabilité sont visibles en Table 2.1	27
2.2	Schéma représentant le lien entre la variable latente et les réponses observées $(Y^{(j)})_{j=1,\dots,J}$ des J items. La variable latente explique les réponses aux différents items, et les différentes réponses reflètent la variable latente.	27
2.3	Courbes caractéristiques aux items (<i>ICC, item characteristic curves</i>) pour modèles IRT associés à des données dichotomiques. La Figure 2.3a montre les ICC des trois items pour le modèle (2.1), avec $\alpha_j = 1$ quel que soit $j = 1, 2, 3$ et les paramètres de difficulté prenant les valeurs -2, 0 et 2. La Figure 2.3a montre les ICC des trois items pour le modèle (2.2), avec $\delta_j = 0$ quel que soit $j = 1, 2, 3$ et les paramètres de discrimination prenant les valeurs $\frac{1}{3}, 1$ et 3.	29
2.4	Courbes caractéristiques aux items pour le modèle de Rasch, le modèle 1PL et le modèle 2PL. Dans chaque cas, il y a deux items ($j = 1, 2$) ayant pour paramètre de difficulté $\delta_1 = -1$ et $\delta_2 = 1$. Pour le modèle de Rasch (2.4a), $\alpha_1 = \alpha_2 = 1$; pour le modèle 1PL (2.4b), $\alpha_1 = \alpha_2 = 0.5$; pour le modèle 2PL (2.4c), $\alpha_1 = 0.5$ et $\alpha_2 = 2$	30

Table des figures

2.5	Liens entre les paramètres de difficulté et les fonctions de probabilité cumulées pour le GRM. Les paramètres de difficulté sont ordonnés avec $\delta_1 = -1$, $\delta_2 = 0$ et $\delta_3 = 1$	32
2.6	Fonctions des catégories de réponse pour le PCM considérant un item à quatre catégories de réponse ($m = 0, \dots, 3$). 2.6a présente le cas où les paramètres de difficulté sont ordonnés avec $\delta_1 = -2$, $\delta_2 = 0$ et $\delta_3 = 3$. 2.6b présente le cas où les paramètres de difficulté ne sont pas ordonnés : $\delta_1 = -2$, $\delta_2 = 3$ et $\delta_3 = 0$	33
2.7	Distributions des paramètres de difficultés estimés (N=100 estimations) dans l'étude de validation croisée. La croix située à droite du Boxplot représente l'estimation obtenue par le modèle lorsque toutes les observations ont été prises en compte.	40
2.8	Distributions des paramètres β_2 et β_3 estimés (N=100 estimations) dans l'étude de validation croisée. La croix située à droite du Boxplot représente l'estimation obtenue par le modèle lorsque toutes les observations ont été prises en compte. Les pourcentages représentent le nombre de fois (sur N=100) où les paramètres sont significatifs. . . .	41
3.1	Schéma représentant la famille des GLMM (en noir) dans laquelle est comprise la famille des GLMM pour données catégorielles (dans le carré : en bleu pour données nominales et en orange pour données ordinales). Les modèles IRT (en rouge) sont un sous-ensemble de modèles GLMM pour données catégorielles qui se distinguent par une paramétrisation spécifique permettant la considération de plusieurs échelles catégorielles.	59
3.2	Visualisation graphique des résultats du modèle cumulatif pour la dimension de douleur sur l'essai Pordige4/Accord11. Lien entre la variable latente θ et la variable réponse ordinaire Y est rendu possible grâce aux paramètres de difficulté (ceux de l'item 9) : $\delta_1 = -2.2$, $\delta_2 = 0.95$ et $\delta_3 = 2.73$	66
3.3	Liaison entre la CdF et le paramètre de discrimination. La Figure 3.3a montre les différentes pentes des F considérées et la Figure 3.3b montre les différentes pentes suivant les différentes valeurs du paramètre discrimination de la CdF logistique.	68
4.1	Décomposition intuitive de la HRQoL suivant la construction du questionnaire générique EORTC QLQ-C30 et du questionnaire spécifique du cancer du sein EORTC QLQ-BR23.	97
4.2	Schéma représentant le SEM avec les différents liens considérés. . . .	98
4.3	Sensibilité du modèle (4.1) en fonction de la taille de l'échantillon n . Les valeurs utilisées pour simuler les données sont : $(q_r = 12)_{r=1,2}$, $c_1 = -c_2 = 4$, $d = 80$, $\sigma_y = \sigma_r = 10$, $a_r = (4, 4, \dots, 4)$, $D^1 = (d_{11}, \dots, d_{1q_1})$ où $(d_{1s} = 70)_{s=1, \dots, q_1}$ et $D^2 = (d_{21}, \dots, d_{2q_2})$ où $(d_{2s} = 30)_{s=1, \dots, q_2}$	108

4.4	Sensibilité du modèle (4.1) en fonction du nombre d'observations (q_r) par bloc r ($r = 1, 2$). Les valeurs utilisées pour simuler les données sont : $n = 150$, $c_1 = -c_2 = 4$, $d = 80$, $\sigma_y = \sigma_r = 10$, $a_r = (4, 4, \dots, 4)$, $D^1 = (d_{11}, \dots, d_{1q_1})$ où $(d_{1s} = 70)_{s=1, \dots, q_1}$ et $D^2 = (d_{21}, \dots, d_{2q_2})$ où $(d_{2s} = 30)_{s=1, \dots, q_2}$	109
4.5	Corrélation entre les facteurs simulés pour générer les données et les facteurs reconstruits par le modèle (4.1), suivant la taille d'échantillon (n) et le nombre de variables observées par blocs (q_r). Les valeurs utilisées pour simuler les données sont : $\sigma_y = \sigma_1 = \sigma_2 = 10$, $c_1 = -c_2 = 4$, $d = 80$, $D^1 = (d_{11}, \dots, d_{1q_1})$ où $(d_{1s} = 70 + 2s)_{s=1, \dots, q_1}$, $D^2 = (d_{21}, \dots, d_{2q_2})$ où $(d_{2s} = 20 + 2s)_{s=1, \dots, q_2}$ et $a_r = (4, 4, \dots, 4)$. Pour la Figure 4.5a, $(q_r = 12)_{r=1, 2}$ et pour la Figure 4.5b, $n = 150$	110
4.6	Corrélation entre les facteurs simulés pour générer les données et les facteurs reconstruits par le modèle (4.1). La Figure 4.6a montre l'influence des valeurs des paramètres c_r et σ_y , $\forall r \in \{1, 2\}$, avec $a_r = (4, 4, \dots, 4)$ et $\sigma_r = 10$. La Figure 4.6b montre l'influence des valeurs des paramètres a_r et σ_r , $\forall r \in \{1, 2\}$, avec $c_1 = -c_2 = 4$ et $\sigma_y = 10$. Les valeurs utilisées pour simuler les données sont : $(q_r = 12)_{r=1, 2}$, $n = 150$, $d = 80$, $D^1 = (d_{11}, \dots, d_{1q_1})$ où $(d_{1s} = 70 + 2s)_{s=1, \dots, q_1}$ et $D^2 = (d_{21}, \dots, d_{2q_2})$ où $(d_{2s} = 20 + 2s)_{s=1, \dots, q_2}$	111
4.7	Corrélations entre les facteurs simulés f_r et les facteurs estimés par le SEM avec \tilde{f}_g^r ou sans \tilde{f}^r la considération de la covariable liée au groupe. Les différences testées sont $\delta = \{0, 5, 10, 20, 30\}$ et supposées uniforme sur l'ensemble des observations.	112
4.8	Estimation des paramètres a^r et c^r en considérant une différence uniforme sur l'ensemble des observations de $\delta = \{0, 5, 10, 20, 30\}$ entre deux groupes.	113
4.9	Corrélations entre le SS^r par bloc et les facteurs respectifs estimés avec \tilde{f}_g^r ou sans \tilde{f}^r la considération de la covariable liée au groupe. Les différences testées sont $\delta = 0, 5, 10, 20, 30$ et supposées uniforme sur l'ensemble des observations	114
4.10	Scatterplots des facteurs fonctionnels (\tilde{f}_i^1) et des facteurs symptomatiques (\tilde{f}_i^2) en fonction respectivement des sommes des scores fonctionnels (SS_i^1) et des sommes des scores symptomatiques (SS_i^2). $(\rho_r)_{r=1, 2}$ dénote la corrélation entre les deux variables.	117
5.1	Classification pour le score du statut global de santé (GHS). Classe A (respectivement D) fait référence à la classe dont la HRQoL s'Améliore (respectivement se Détérioré) au cours du temps.	153
5.2	Classification relative à la dimension de fatigue. Classe A (respectivement C et D) fait référence à la classe dont la HRQoL s'Améliore (respectivement reste Constante et se Détérioré) au cours du temps.	154
5.3	Classification relative à la dimension liée à l'insomnie (SL).	155

Table des figures

5.4	Classification relative à la dimension liée à la perte d'appétit (AP). Les figures 5.4a, 5.4b et 5.4c présentent les résultats de l'approche CTT pour données ordinales, et les figures 5.4d, 5.4e et 5.4f ceux de l'approche CTT pour données gaussiennes.	156
5.5	Trajectoires latentes issues de la classifications relatives aux dimen- sions suivantes : score global de Nordin, dimension physique (PF), dimension personnelle (RF) et dimension émotionnelle (EF).	178

Liste des tableaux

1.1	Modules additionnels associés aux différentes maladies. "Additionnel" précise que le questionnaire vient compléter le questionnaire QLQ-C30, et "complémentaire" précise que le questionnaire est indépendant.	9
2.1	Détails de l'exemple fictif concernant la mesure de la vulnérabilité. Les réponses de quatre sujets à trois items dichotomiques sont considérées. La réponse positive « <i>oui</i> » est associée à la valeur 1, et la réponse négative « <i>non</i> » à la valeur 0.	26
2.2	Récapitulatif des différentes paramétrisations IRT associées aux modèles IRT.	34
2.3	Résumé des quatre analyses possibles. Z_{ip} matrice associée au régresseur sur l'individu. D_{jk} matrice associée au régresseur sur l'item. Les variables aléatoires sont définies telles que $\theta \sim \mathcal{N}(0, \sigma^2)$ et $\theta \sim \mathcal{N}(0, \sigma_\varepsilon^2)$	36
2.4	Estimations des effets fixes $(\beta_p)_{p=1,2,3}$ des modèles LMM, LPCM et LGRM pour chaque dimension du QLQ-C30.	39
3.1	Récapitulatif des différentes caractéristiques des GLMM pour données ordinales.	65
3.2	Estimations de -2ℓ associées aux différents modèles adjacents et cumulatifs et aux différentes CdF. Six dimensions de HRQoL sont utilisées : statut global de santé (GHS), fonction physique (PF2), fonction cognitive (CF), fatigue (FA), douleur (PA) et diarrhée (DI). Pour les cases vides, le modèle n'a pas convergé.	67
4.1	Estimations des paramètres $(a^r)_{r=1,2}$ du modèle (voir 4.2a) à chaque visite $t_v = (0, 3, 6, 12, 15, 18, 21, 24)$ (en mois). La covariable liée au groupe de traitement a été prise en compte dans chaque équation. La moyenne des paramètres est celle sur l'ensemble des paramètres pour une même visite.	116
4.2	Estimations des paramètres $(\sigma_m)_{m=y,1,2}$ et $(c_r)_{r=1,2}$ du modèle (4.1) pour chaque visite (en mois). La covariable liée au groupe de traitement a été prise en compte dans chaque bloc de mesure X^1 et X^2 . La moyenne des paramètres est celle sur l'ensemble des visites.	116
4.3	Estimations des paramètres D^1 du modèle (voir 4.2a). La covariable liée au groupe de traitement a été prise en compte dans chaque équation du modèle.	118

5.1	Résumé des différents outils de mesures psycho-sociales associés aux temps de suivi pour lesquels ils ont été utilisés. Les temps de suivi $T_0, T_1, T_2, T_3, T_4, T_5$ correspondent respectivement à la chirurgie, à 1 mois, 4, 7, 10 et 13 mois.	150
5.2	Tableaux de contingence entre les classifications obtenues suivant les deux approches CTT pour $K = 2, 3$ (résultats entre crochets pour $K = 2$). CTT ¹ correspond à l'approche pour données gaussiennes et CTT ² à l'approche pour données ordinales. Lorsque $K = 2$, 82% de la classification est similaire. Lorsque $K = 3$, 84% de la classification est similaire.	157
5.3	Récapitulatif des coefficients qui ont permis de simuler les 500 jeux de données de 150 patients. Ces coefficients correspondent à ceux obtenus pour la classification liée à la dimension de l'insomnie.	158
5.4	Choix du nombre de classes latentes K suivant les critères BIC et ICL et sachant que $K = 3$ dans les $N = 500$ jeux de données simulés. . .	159
5.5	Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score du GHS. Deux classes latentes sont considérées : classe A pour Amélioration de la HRQoL et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.	160
5.6	Résultats des tests d'indépendance entre les variables psychosociales et la classification de la population d'intérêt suivant le score du GHS. Deux classes latentes sont considérées : classe A pour Amélioration de la HRQoL et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.	161
5.7	Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score de fatigue. Trois classes latentes sont considérées : classe A pour Amélioration de la HRQoL, classe C pour HRQoL Constante et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.	162
5.8	Résultats des tests d'indépendance entre les variables psychosociales et la classification de la population d'intérêt suivant le score de fatigue. Trois classes latentes sont considérées : classe A pour Amélioration de la HRQoL, classe C pour HRQoL Constante et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.	163
5.9	Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score lié à la perte d'appétit. Deux classes latentes sont considérées : classe A pour Amélioration de la HRQoL et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.	164

5.10	Résultats des tests d'indépendance entre les variables psychosociales et la classification de la population d'intérêt suivant le score lié à la perte d'appétit. Deux classes latentes sont considérées : classe A pour Amélioration de la HRQoL et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.	165
5.11	Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score d'insomnie. Trois classes latentes sont considérées : classe A pour Amélioration de la HRQoL, classe C pour HRQoL Constante et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages. . . .	166
5.12	Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score d'insomnie. Trois classes latentes sont considérées : classe A pour Amélioration de la HRQoL, classe C pour HRQoL Constante et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages. . . .	167
5.13	Résultats associés au modèle multinomiale logistique concernant le statut global de santé (GHS). OR correspond à l'Odd Ratio, $IC_{95\%}$ son intervalle de confiance et p . la pvalue associée au test de rapport de vraisemblance relatif au paramètre testé.	168
5.14	Résultats associés au modèle de régression logistique multinomiale concernant la fatigue. OR correspond à l'Odd Ratio, $IC_{95\%}$ son intervalle de confiance et p . la pvalue associée au test de rapport de vraisemblance relatif au paramètre testé.	169
5.15	Résultats associés au modèle de régression logistique multinomiale concernant l'insomnie. OR correspond à l'Odd Ratio, $IC_{95\%}$ son intervalle de confiance et p . la pvalue associée au test de rapport de vraisemblance relatif au paramètre testé.	170
5.16	Résultats associés au modèle de régression logistique multinomiale concernant la perte d'appétit. OR correspond à l'Odd Ratio, $IC_{95\%}$ son intervalle de confiance et p . la Pvalue associée au test de rapport de vraisemblance relatif à la variable explicative.	170
5.17	Pvalues des tests d'indépendance entre les variables médicales et sociodémographiques, et les classes latentes obtenues.	177
5.18	Pvalues des tests d'indépendance entre les variables psychosociales (à $t = 1$ pour celles du questionnaire MAC) et les classes latentes obtenues.	178
5.19	Résultats du modèle logistique multinomiale concernant le score global de Nordin.	179
5.20	Estimations du modèle logistique multinomiale concernant la dimension physique (PF).	179
5.21	Estimations du modèle logistique multinomiale concernant la dimension émotionnelle (EF).	179

Liste des tableaux

5.22 Estimations du modèle logistique multinomiale concernant la dimension personnelle (RF). 180

Introduction

Ce chapitre présente de manière générale la qualité de vie relative à la santé (*health-related quality of life*, HRQoL) : l'intérêt de son étude en cancérologie, sa mesure ou encore les différentes méthodes statistiques utilisées actuellement pour analyser ce critère dans les essais cliniques. Les différents axes de recherche liés à la HRQoL sont présentés, ainsi que les objectifs de mes travaux motivés par la proposition de stratégies longitudinales pour l'analyse de ce critère.

Sommaire

1.1	Contexte	2
1.2	La qualité de vie relative à la santé (HRQoL)	2
1.2.1	Place du critère de HRQoL dans les essais cliniques	2
1.2.2	Définition de la HRQoL	4
1.3	Mesure de la HRQoL	5
1.3.1	Groupe EORTC	7
1.3.2	Groupe FACT	8
1.3.3	Approche classique dite du « vrai score »	10
1.4	État de l'art des approches statistiques pour l'analyse de la HRQoL dans les essais cliniques en cancérologie	12
1.4.1	Modèles linéaires mixtes	12
1.4.2	Modèle de survie : temps jusqu'à la détérioration du score	14
1.5	Problématiques de l'analyse longitudinale de la HRQoL	16
1.5.1	Les données manquantes	16
1.5.2	Le phénomène de « Response Shift »	17
1.5.3	Limites du score de HRQoL	18
1.5.4	Multiplicité des tests	19
1.6	Objectifs	20

1.1 Contexte

La recherche en cancérologie a beau être en perpétuel progrès, le cancer est une maladie qui préoccupe les esprits de tous, et pas seulement ceux des professionnels de santé. Par exemple, quasiment tous les Français citent spontanément le cancer parmi les trois maladies les plus graves, loin devant le sida et les maladies cardiovasculaires (Beck et al., 2010). Le cancer est caractérisé par une croissance incontrôlée de cellules, qui peuvent envahir et se propager dans l'organisme. Cette maladie peut engendrer de graves conséquences sur la santé, dont une des principales est le décès.

La recherche associée à la lutte contre le cancer implique de nombreuses disciplines (médecine, biologie, physique, mathématique, économie, psychologie,...) et s'inscrit dans différents thématiques de recherche. La recherche clinique vise à améliorer les traitements et les pratiques médicales pour répondre à des objectifs visant à améliorer la prise en charge des malades et ainsi espérer prolonger leur durée de vie.

La notion de qualité de vie relative à la santé (*health-related quality of life*, HR-QoL) relève principalement de l'aspect médical et de la prise en charge du patient. L'objectif de l'intervention médicale est la suppression des cellules cancéreuses de l'organisme. Cela peut être réalisé de manière chirurgicale ou avec un traitement. Les traitements (chimiothérapie, radiothérapie, ...) sont toxiques pour les cellules cancéreuses mais également pour le reste de l'organisme induisant des effets secondaires notoires tels que la perte des cheveux, des vomissements, de la fatigue, etc... Si les progrès constants de la médecine conduisent à de nouveaux traitements améliorant les chances de prolonger la quantité de vie, il est tout autant nécessaire de s'intéresser à la qualité de ce gain de survie.

1.2 La qualité de vie relative à la santé (HRQoL)

1.2.1 Place du critère de HRQoL dans les essais cliniques

Dans le développement thérapeutique, trois phases cliniques sont nécessaires (voir Figure 1.1). L'essai clinique de phase I correspond à la première évaluation chez l'homme d'une nouvelle molécule. L'objectif principal concerne l'étude des toxicités doses limitantes et plus généralement de la tolérance de la nouvelle molécule afin de déterminer la dose maximale tolérée et la dose recommandée pour les futurs essais de phases 2. En cancérologie, cette phase regroupe une petite cohorte de patients (de 15 à 50) à un stade métastatique et en situation d'échec thérapeutique. L'essai clinique de phase II a pour objectif d'explorer l'efficacité thérapeutique en termes d'activité anti-tumorale, mais également d'étudier la tolérance, la survie ainsi que la HRQoL. Elle concerne une cohorte dont l'effectif est supérieur à celui d'une étude de phase I, mais qui est limité (<100 patients) afin de minimiser le nombre de patients traités par des molécules potentiellement inefficaces. Enfin, l'essai clinique de phase III étudie l'efficacité du médicament en situation réelle (cohorte représentative de la population de patients susceptible de bénéficier du nouveau traitement, de 100 à 1000 patients) en le comparant avec le traitement standard en termes de bénéfice

1.2. La qualité de vie relative à la santé (HRQoL)

clinique. Le nouveau traitement anti-cancéreux en comparaison au traitement standard, doit prolonger la survie globale des patients (*overall survival*, OS) et ne pas détériorer leur HRQoL.

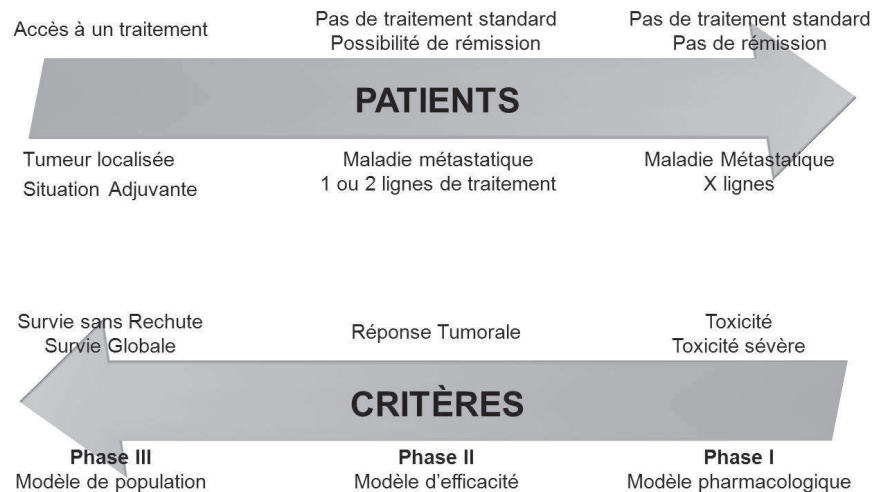


FIGURE 1.1: Schéma représentant les trois phases cliniques en cancérologie en fonction du patient (évolution de sa maladie) et du critère étudié.

Dans la réalisation d'un essai clinique, le choix du critère de jugement principal est essentiel, non seulement pour le calcul du nombre de sujets nécessaires, mais également pour la pertinence clinique des résultats (Fiteni et al., 2015). Dans ce contexte, deux types de critères de jugement se distinguent : les critères centrés sur le patient qui reflètent ce que le patient perçoit d'une fonction ou d'une survie, telle que la OS ou la HRQoL ; et les critères centrés sur la tumeur tels que la survie sans progression.

En cancérologie, l'OS est considérée comme le critère de jugement de référence dans les essais cliniques de phase III pour évaluer le bénéfice clinique d'une nouvelle stratégie thérapeutique (Food and Administration, 2007). En effet, l'OS est un critère clinique objectif, universel et parfaitement défini (intervalle de temps jusqu'au décès toutes causes confondues). À l'heure actuelle, les traitements anticancéreux se multiplient et sont de plus en plus efficaces, allongeant de manière significative la vie des patients. Dans certaines pathologies (notamment le cancer du sein en situation adjuvante¹), les nouveaux traitements ne permettent pas d'espérer des différences significatives sur l'OS pour un nombre de sujets ou une durée d'étude raisonnable. Des critères de jugement alternatifs tels que des critères centrés sur la tumeur, dits de substitution, sont alors de plus en plus préconisés. Ces critères ont l'avantage de pouvoir être évalués plus précocement, et de réduire le nombre de sujets à recruter ainsi que la durée de l'essai (Fiteni et al., 2014).

Pour pouvoir être utilisés en tant que critère de jugement principal, ces critères doivent être soit validés en tant que critère de jugement substitutif de l'OS, soit

1. c'est-à-dire après une chirurgie

s'assurer du bénéfice clinique pour le patient avec l'utilisation conjointe d'un critère centré sur le patient. Dans ce contexte, la HRQoL a souvent été proposée (Anota, 2014). La HRQoL constitue donc un critère de jugement alternatif pertinent et disponible pour s'assurer de l'intérêt du traitement pour le patient et le système de santé (Fiteni et al., 2015). Ce critère entre dans le champ des « Patient-Reported Outcomes » (PROs) (Gotay et al., 1992), i.e. des mesures rapportées par les sujets eux-mêmes, et est considérée comme second critère de jugement principal par l'« American Society of Clinical Oncology » (ASCO) et la « Food and Drug Administration » (FDA) en l'absence d'effet sur l'OS (Beitz et al., 1996). De manière plus spécifique, ce critère de jugement pourrait être considéré comme principal dans les situations gériatriques² et palliatives³ afin de s'assurer en priorité du confort et du bien-être des patients.

Cependant, des freins conceptuels (définition, évaluation subjective par le patient) et méthodologiques (sens clinique d'un changement, données longitudinales, nature multidimensionnelle, données manquantes, etc..) limitent non seulement l'évaluation de la HRQoL, mais également l'utilisation des résultats pour l'aide à la décision. De ce fait, les oncologues sont réticents à changer les pratiques à partir des résultats de HRQoL issus des essais. La complexité des analyses longitudinales de HRQoL et leurs interprétations ne sont pas étrangères à cet « échec » (Bonnetain, 2010).

1.2.2 Définition de la HRQoL

La qualité de vie (QdV) a été définie par l'organisation mondiale de la santé (OMS) en 1948 comme étant « *a state of complete physical, mental and social well-being, and not merely the absence of disease* », soit « un état de bien-être physique, mental et social complet, et pas simplement l'absence de maladie ». En d'autres mots, la QdV est un concept dynamique, multidimensionnel et subjectif incorporant au moins trois dimensions : les fonctionnements physique, psychologique et social. En 1993, l'OMS définit la QdV 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 conceptuel, 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 » (Who, 1993).

Relativement à la santé et dans un contexte médical, la définition de la HRQoL est plus spécifique et intègre des dimensions relatives à l'impact de la maladie et du traitement (aspect multidimensionnel du critère). Ces dimensions particulières sont caractérisées par les conséquences directes du traitement ou de la maladie telles que les symptômes, mais également les conséquences indirectes telles que les difficultés financières. Une seconde caractéristique de la HRQoL est l'aspect subjectif.

2. concernant les personnes âgées ;

3. concernant les malades en phase très avancée, terminale ;

En effet, il existe une forte variabilité inter-individuelle quant à la définition de la HRQoL. Celle-ci dépend entre autre du vécu, de la culture et des attentes de l'individu, mais également des éléments environnementaux. Afin de montrer la complexité de cette notion, les dimensions qui définissent la HRQoL sont eux-mêmes plus ou moins subjectifs. La dernière caractéristique de la HRQoL est l'aspect dynamique, principalement dû à la subjectivité de ce critère. Entre deux instants relativement espacés (en termes de semaine), la variabilité intra-individuelle pour la définition de la HRQoL peut être grande et est conditionnée par l'évolution personnelle de celui qui la juge.

1.3 Mesure de la HRQoL

La HRQoL n'est pas directement mesurable et, malgré sa complexité, il a donc été nécessaire de construire des outils efficaces pour l'évaluer⁴. La mesure de la HRQoL est actuellement réalisée par le biais d'auto-questionnaires, dont leur standardisation est un pré-requis indispensable au développement et à l'optimisation de méthodes statistiques qu'elles soient descriptives ou explicatives.

La mesure de la HRQoL s'effectue par l'intermédiaire d'auto-questionnaire(s) administré(s) aux patients. Pour répondre aux caractéristiques du concept mesuré, ces questionnaires sont multi-dimensionnels et chaque dimension est indirectement mesurée par un ensemble d'items (voir Figure 1.2). En plus de prendre en compte la nature multidimensionnelle de ce concept, ces outils considèrent son aspect subjectif. En effet, la définition du concept de HRQoL comme celles de ses dimensions (fonctionnement psychologique par exemple) est sujet à une forte hétérogénéité inter-individuelle. Ainsi, les items sont construits de telle manière à ce qu'ils soient compréhensibles de tous, utilisant une sémantique simple et claire limitant au maximum la variabilité d'interprétation (Vanier et al., 2015).

Pour la HRQoL, les items sont des questions où les réponses peuvent être dichotomiques (binaires) ou polytomiques (au moins trois modalités de réponse). Dans le cas polytomique, l'item est construit sur le modèle d'une échelle de Likert. Cette dernière évalue les attitudes et comportements en proposant des choix de réponses allant d'un extrême à l'autre (par exemple, « Pas du tout probable » à « Extrêmement probable »). À l'inverse d'une simple question ayant deux modalités de réponse (« Oui » ou « Non »), l'échelle de Likert permet de dégager différents degrés d'opinion. Cette échelle propose un certain nombre de catégories de réponse ordonnées (généralement compris entre 3 et 7). Dans le cas d'un nombre impair de choix de réponse, celle située au milieu est dite neutre. En revanche, dans le cas d'un nombre pair, le sujet est contraint à se positionner soit dans le « négatif », soit dans le « positif ». Dans les questionnaires, ces différentes catégories de réponse sont associées à des valeurs (ou scores) permettant de quantifier un score lié à la dimension de HRQoL considérée.

Chacune des dimensions peuvent être évaluées par un ou plusieurs items. Plus la dimension étudiée va être associée à un concept complexe, plus le nombre d'items va être élevé. En revanche, pour des dimensions précises telles que celles associées à

4. tout comme l'intelligence, conceptuellement similaire à la HRQoL ;

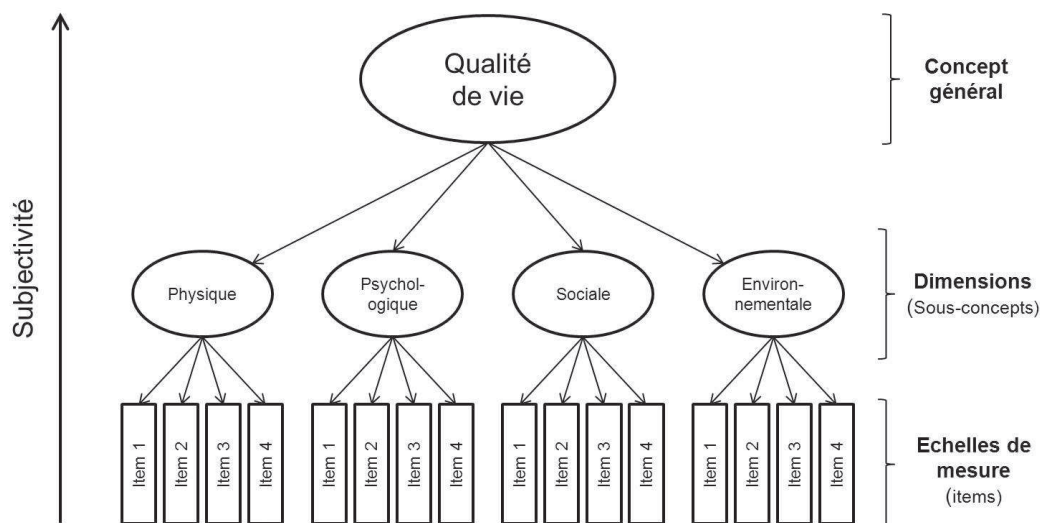


FIGURE 1.2: Schéma représentant la construction d'un questionnaire psychométrique. Les concepts non directement mesurables sont dans des cercles. Les observations découlent des items caractérisés par des rectangles.

un symptôme spécifique (exemple, la diarrhée), un seul item s'avère suffisant pour le mesurer. Pour évaluer la HRQoL, certains questionnaires sont construits sur un mélange de dimensions uni et multi-items (Aaronson et al., 1993).

Un certain nombre de questionnaires de HRQoL a été développé. Parmi les questionnaires génériques, le MOS SF-36 ou le WHOQOL (« World Health Organization Quality Of Life ») ont été conçus afin d'auto-évaluer la HRQoL d'une personne lambda, et ce quel que soit son état de santé. Comme son nom l'indique, le MOS SF-36 est composé de 36 items permettant l'évaluation de huit dimensions de HRQoL (Ware and Sherbourne, 1992). Deux versions du questionnaire WHOQOL, développé par l'OMS, existent. La version longue (non censurée) contient 100 items alors que la version « BREF » n'en contient "que" 26. Ils permettent entre autres d'évaluer les dimensions physique, psychique, les relations sociales ainsi que l'indépendance, la spiritualité et les croyances du patient. Tous les questionnaires développés doivent respecter plusieurs propriétés psychométriques pour être validés et utilisés avec fiabilité (McDowell, 2006; Walters, 2009). Les principales caractéristiques à respecter sont les suivantes :

- Validité culturelle : adaptation à la langue d'utilisation (adaptation socioculturelle et pas une simple traduction) ;
- Validité pratique : taux de participation, taux de complétion⁵, temps de remplissage ;
- Validité structurelle : items d'une sous-échelle liés à une seule variable latente et items fortement liés à cette sous-échelle ;
- Fiabilité et reproductibilité.

⁵. pourcentage de questionnaires rendus

Dans ce manuscrit, nous concentrons notre attention sur des questionnaires spécifiques au cancer. Parmi eux, deux familles de questionnaires se distinguent : les questionnaires de l'EORTC (« European Organization of Research and Treatment of Cancer ») et du groupe FACT (« Functional Assessment of Cancer Therapy ») respectivement utilisés en Europe et aux États Unis.

1.3.1 Groupe EORTC

En Europe, la référence en cancérologie est le questionnaire EORTC QLQ-C30 développé par le groupe EORTC (Aronson et al., 1993). L'EORTC QLQ-C30 est un questionnaire (auto-rapporté) conçu pour évaluer la HRQoL des patients atteints de cancer (questionnaire visible en Annexe B). Historiquement, le développement des questionnaires EORTC a commencé en 1980, puis s'est concrétisé en 1993 suite à la collaboration de plusieurs experts dans le domaine de la HRQoL (Aronson et al., 1993). La première version finalisée (1993) a été modifiée afin d'augmenter la fiabilité et la validité de l'outil. Aujourd'hui, l'EORTC *Quality of Life Group* recommande la version 3.0 du questionnaire. Celui-ci a été vérifié et validé psychométriquement dans 81 langues différentes afin de pouvoir être utilisé dans de nombreuses études et ainsi permettre la comparaison des résultats. Il a été conçu pour être spécifique à des données de cancérologie, abordable et simple à compléter pour des patients issus de différentes cultures. Le temps moyen de remplissage du questionnaire est légèrement supérieur à 10 minutes.

Ce questionnaire est fréquemment utilisé en oncologie dans le cadre de recherches cliniques. Il est en général proposé dans des essais de phase II ou III pour tout type de cancer. L'EORTC QLQ-C30 est composé de trente items regroupés en quinze dimensions incluant les dimensions fonctionnelles, les dimensions symptomatiques et le statut global de santé (voir Figure 1.3). Les dimensions fonctionnelles englobent les fonctions suivantes : physique, personnelle (concernant les activités quotidiennes, professionnelles ou de loisirs), émotionnelle, cognitive (concentration et mémoire) et sociale. Elles sont composées de deux items (dimensions émotionnelle, sociale et cognitive), quatre items (dimension émotionnelle) et cinq items (dimension physique). Les symptômes classiques incluent des questions sur la fatigue (trois items), les nausées et vomissements ou encore la douleur (deux items). Les symptômes additionnels uni-item sont plus précis et couramment répertoriés chez des patients traités pour un cancer. Il s'agit des symptômes dus à la maladie et aux effets secondaires fréquents des thérapies anticancéreuses, comme par exemple la dyspnée (difficulté respiratoire, essoufflement), la perte d'appétit, l'insomnie, la constipation, les diarrhées. Enfin, le QLQ-C30 comprend également une dimension (à un item) liée aux difficultés financières. Chacun des 28 premiers items est construit sur la base de quatre modalités de réponse (1 « pas du tout » ; 2 « un peu » ; 3 « assez » ; 4 « Beaucoup »), tandis que les deux derniers items relevant du statut global de santé s'évaluent sur une échelle à sept niveaux (1 « très mauvaise » ; 7 « excellente »).

Pour chacune des dimensions, un score de HRQoL (correspondant à la moyenne des réponses exprimée sur une échelle allant de 0 à 100) est calculé par sujet à partir

de ses réponses aux items.

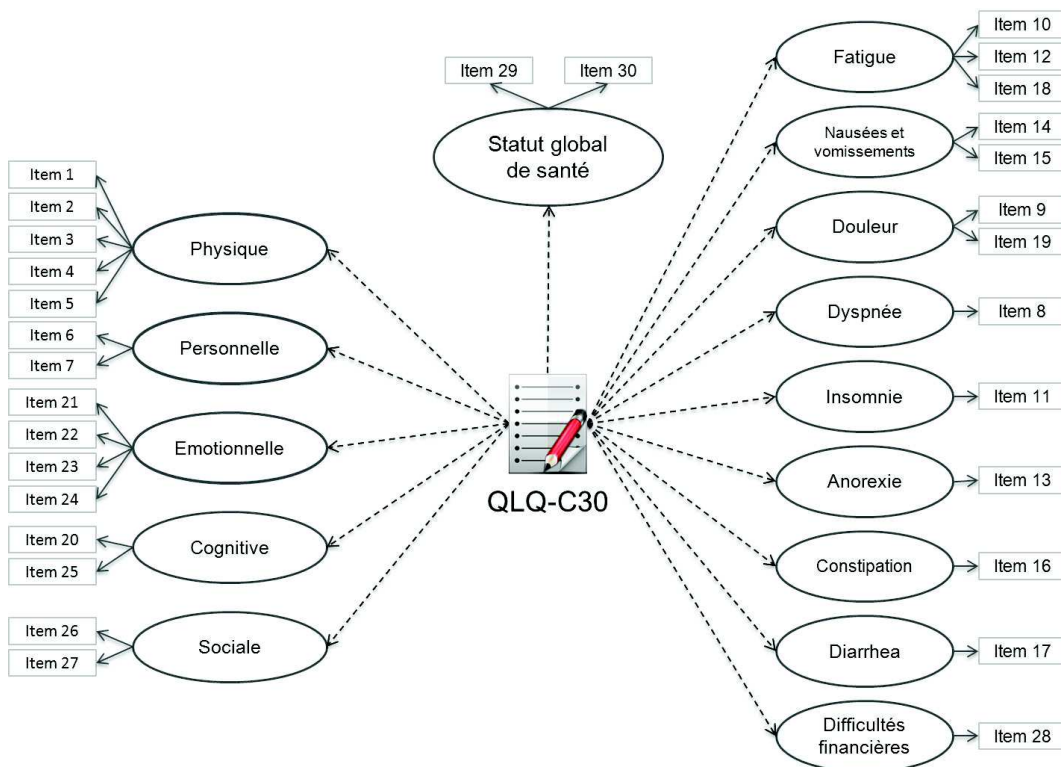


FIGURE 1.3: Schéma représentant le questionnaire EORTC QLQ-C30 avec les dimensions associées. Les dimensions de gauche sont les dimensions fonctionnelles et les dimensions de droite sont les dimensions symptomatiques.

En complément du QLQ-C30, plusieurs modules spécifiques (voir Table 1.1) ont été développés par l'EORTC. Ils permettent une évaluation plus complète de la HRQoL en prenant en considération les spécificités de la localisation cancéreuse du patient. À titre d'exemple, le module additionnel QLQ-BR23 associé au cancer du sein permet l'évaluation de huit dimensions complémentaires de HRQoL : quatre dimensions fonctionnelles (image corporelle, fonctionnement sexuel, plaisir sexuel et perspectives futures) et quatre dimensions symptomatiques (symptômes liés au traitement, symptômes au niveau du bras, symptôme au niveau du sein, inquiétude liée à la perte des cheveux).

1.3.2 Groupe FACT

Le questionnaire général de HRQoL spécifique au cancer, développé et validé par le groupe américain FACT, est le FACT-G (Cella et al., 1993). Ce questionnaire contient 27 items construits sur une échelle de Likert à 5 modalités de réponse codées : 0 « Pas du tout » ; 1 « Un peu » ; 2 « Moyennement » ; 3 « Beaucoup » ; 4 « Énormément ». Quatre dimensions de bien-être sont évaluées par le biais de ces items : le bien-être physique (voir Figure 1.4), le bien-être social et familial, le bien-

TABLE 1.1: Modules additionnels associés aux différentes maladies. "Additionnel" précise que le questionnaire vient compléter le questionnaire QLQ-C30, et "complémentaire" précise que le questionnaire est indépendant.

Localisation cancéreuse	Code questionnaire	Propriété	Nombre d'items
Métastases Osseuses	<i>QLQ – BM22</i>	complémentaire	22
Cerveau	<i>QLQ – BN20</i>	additionnel	20
Carcinome Hépatocellulaire	<i>QLQ – HCC18</i>	additionnel	18
Information	<i>QLQ – INFO25</i>	additionnel	25
Sein	<i>QLQ – BR23</i>	additionnel	23
Poumon	<i>QLQ – LC13</i>	additionnel	13
Cancer de l'utérus	<i>QLQ – CX24</i>	additionnel	24
Cancer colorectal	<i>QLQ – CR29</i>	additionnel	29
Cancer colorectal - Métastase hépatique	<i>QLQ – LMC21</i>	additionnel	21
Endomètre	<i>QLQ – EN24</i>	additionnel	24
Gastrique	<i>QLQ – STO22</i>	additionnel	22
Tête et Cou	<i>QLQ – H&N35</i>	additionnel	35
Myélome multiple	<i>QLQ – MY20</i>	additionnel	20
Tumeurs neuroendocrines et carcinoïdes	<i>QLQ – GINET21</i>	additionnel	21
Œsophage	<i>QLQ – OES18</i>	additionnel	18
Œsophago - gastrique	<i>QLQ – OG25</i>	complémentaire	25
Ovaire	<i>QLQ – OV28</i>	additionnel	28
Prostate	<i>QLQ – PR25</i>	additionnel	25

être émotionnel et le bien-être fonctionnel.

Comme pour les questionnaires de l'EORTC, un score est calculé par dimension. Ces scores varient de 0 à 28 où 0 représente un bas niveau de bien-être et 28 un haut niveau de bien-être. Pour qu'un score puisse être estimé, au moins 50% des items de la dimension doivent être renseignés. Contrairement au questionnaire EORTC QLQ-C30, les items du questionnaire FACT-G sont présentés par dimension⁶ et le nom de chaque dimension évaluée précède les items correspondant. De plus, un score FACT-G global est calculé à partir des scores obtenus pour chaque dimension. Ce score correspond à la somme des scores bruts obtenus pour chaque dimension et il varie de 0 (faible niveau) à 108 (bon niveau de bien-être général). Ce score global peut être calculé à condition que l'ensemble des sous-scores a pu être calculé. Les quatre scores de bien-être ainsi que le score FACT-G global sont généralement standardisés de 0 à 100 dans le but de faciliter la comparaison avec d'autres questionnaires.

À la différence du groupe EORTC, les spécificités des différentes localisations cancéreuses ne sont pas apportées par des modules additionnels, mais par leur inté-

6. les items d'une même dimension sont adjacents

gration directe dans le questionnaire FACT-G. Pour le cancer du sein par exemple, il s'agit du questionnaire FACT-B qui regroupe les 27 items du FACT-G ainsi que 9 items supplémentaires qui permettent d'évaluer une et une seule dimension spécifique associée au cancer du sein (Brady et al., 1997). Pour les questionnaires du groupe FACT, il n'y pas de dimension symptomatique uni-item. Dans le but de considérer une dimension de HRQoL spécifique à un symptôme particulier, le groupe FACT a également développé des questionnaires contenant les items du FACT-G ainsi qu'une dimension additionnelle particulière, telle que la fatigue. Ainsi, le FACT-F contient les 27 items du FACT-G ainsi que 13 items supplémentaires évaluant le niveau de fatigue du patient (Yellen et al., 1997).

Below is a list of statements that other people with your illness have said are important. **Please circle or mark one number per line to indicate your response as it applies to the past 7 days.**

<u>PHYSICAL WELL-BEING</u>		Not at all	A little bit	Some- what	Quite a bit	Very much
GP1	I have a lack of energy	0	1	2	3	4
GP2	I have nausea	0	1	2	3	4
GP3	Because of my physical condition, I have trouble meeting the needs of my family	0	1	2	3	4
GP4	I have pain	0	1	2	3	4
GP5	I am bothered by side effects of treatment	0	1	2	3	4
GP6	I feel ill	0	1	2	3	4
GP7	I am forced to spend time in bed.....	0	1	2	3	4

FIGURE 1.4: Dimension du bien-être physique relative au questionnaire FACT-G (version 4).

1.3.3 Approche classique dite du « vrai score »

L'analyse de la HRQoL en cancérologie est fondée sur l'étude des scores de HRQoL (notés S) associés aux dimensions étudiées. Cette approche dite du vrai score (*Classical Test Theory*, CTT) suggère que le score est une bonne représentation de la variable latente d'intérêt (représentant la HRQoL du patient), notée θ . Pour un individu i ($i = 1 \dots, n$), la relation s'écrit :

$$S_i = \theta_i + \varepsilon_i \tag{1.1}$$

avec θ_i la valeur spécifique de la variable latente pour l'individu i et ε_i un terme d'erreur associé.

La CTT est l'approche habituellement utilisée en cancérologie pour l'étude de la HRQoL, les analyses étant réalisées séparément pour chaque dimension du question-

naire. Elle réfère à la notion de « score-somme » définit par :

$$S = \sum_{j=1}^J (Y^{(j)} - 1), \quad (1.2)$$

où $Y^{(j)} \in \{1, \dots, M_j + 1\}$ est la réponse à l’item j ($j = 1, \dots, J$) (Boeck and Wilson, 2004; Gorter et al., 2015). L’équation 1.2 est utilisé pour générer les scores des dimensions du FACT-G (Cella et al., 1993), engendrant un score allant de 0 à $\sum_{j=1}^J M_j$. L’inconvénient du « score-somme » est dans la comparabilité des scores entre questionnaires, voir même entre dimensions. Il est alors fréquent de considérer des scores standardisés et exprimés sur une échelle de 0 à 100. Dans les essais cliniques européens, le questionnaire EORTC-QLQC30 est dans la plupart du temps distribué aux patients et les scores de HRQoL sont alors calculés suivant les recommandations associées à ce questionnaire (Fayers et al., 2001). Chaque score est lié à une dimension de HRQoL à travers les regroupements d’items définis pour les questionnaires utilisés. L’EORTC propose des formules précises (une procédure de scoring) pour le calcul de ces scores (Fayers et al., 2001). Les scores sont tout simplement calculés comme la moyenne des réponses aux items, standardisée de manière à être exprimée sur une échelle de 0 à 100. Si on note J le nombre d’items à M catégories de réponse pour la dimension considérée et (y_1, \dots, y_J) les réponses respectives. Le score S est égal à :

$$S = \frac{\left(\frac{1}{J} \sum_{j=1}^J y_j\right) - 1}{M - 1} \times 100 \quad (1.3)$$

pour les échelles symptomatiques, et

$$S = \left[1 - \frac{\left(\frac{1}{J} \sum_{j=1}^J y_j\right) - 1}{M - 1} \right] \times 100 \quad (1.4)$$

pour les échelles fonctionnelles. Dans le cas des dimensions fonctionnelles, l’échelle est inversée dans le but d’avoir un score élevé qui reflète une bonne capacité fonctionnelle (physique, relationnelle, ...) et donc un bon niveau de HRQoL. À l’inverse, dans le cas des échelles symptomatiques, un score élevé est l’expression de forts symptômes (fatigue, diarrhée, perte d’appétit, ...), et donc d’un mauvais niveau de HRQoL. En présence de données manquantes, l’EORTC propose de considérer le score comme valide tant qu’au minimum la moitié des réponses aux items est connue. Le score devient la moyenne standardisée sur les réponses aux items non manquantes (Fayers et al., 2001).

Le score de HRQoL présente des avantages pour l’analyse statistique. En effet, l’analyse s’effectuant par dimension de HRQoL, la variable réponse est alors quantitative et univariée.

1.4 État de l'art des approches statistiques pour l'analyse de la HRQoL dans les essais cliniques en cancérologie

Dans les essais cliniques, l'étude de la HRQoL a pour objectif de vérifier l'impact des traitements à l'étude (expérimental versus standard) sur ce critère. Les questionnaires de HRQoL sont donc transmis aux patients tout au long de leur prise en charge. L'administration de ces questionnaires aux patients à l'inclusion, pendant le traitement et au cours du suivi génère une répétition de mesures pour un même patient au cours du temps. Les données d'un même sujet récoltées aux différents temps de mesure sont donc corrélées. Il est alors indispensable d'avoir recours à des analyses longitudinales pour étudier l'évolution de la HRQoL et la comparer entre les traitements à l'étude.

Parmi les approches qui ne sont plus utilisées, nous pouvons rappeler celle de l'aire sous la courbe (*Area Under the Curve, AUC*), où "la courbe" décrit l'évolution du score au cours du temps. Cette méthode peut s'appliquer sur les données complètes et incomplètes, et renvoie une valeur unique par individu. La comparaison entre deux groupes s'effectue par la comparaison en moyenne des AUC pour chacun des groupes. Cependant, cette approche consiste en une comparaison globale sur les temps de suivi mais n'apporte pas d'information quant à l'évolution de la HRQoL. Actuellement, le type d'approche longitudinale proposé pour l'étude de la HRQoL dans les essais cliniques converge vers l'utilisation de modèles mixtes et de modèles de survie qui sont particulièrement adaptés aux données comme aux objectifs d'analyse.

1.4.1 Modèles linéaires mixtes

Une écriture synthétique d'un LMM à K effets aléatoires est de la forme :

$$Y = \underbrace{X\beta}_{\text{Fixe}} + \underbrace{U\xi}_{\text{Aléatoire}} + \varepsilon \quad (1.5)$$

où

- Y est le vecteur aléatoire réponse de taille N ;
- β est le vecteur des paramètres inconnus des effets fixes, de taille p , et X sa matrice d'incidence supposée connue et de dimension $(N \times p)$;
- ξ est le vecteur de l'ensemble des effets aléatoires, décomposé par les K d'effets aléatoires sous la forme $\xi = (\xi'_1, \dots, \xi'_K)$. Chaque vecteur ξ'_k est de dimension q_k et on note $q = \sum_{k=1}^K q_k$ la dimension de ξ . Les effets aléatoires sont supposés de distribution gaussienne centrée, $(\xi_k)_{k=1, \dots, K} \sim \mathcal{N}_{q_k}(0, \sigma_k^2 A_k)$ avec A_k une matrice supposée connue de dimension $q_k \times q_k$. De plus, les effets aléatoires sont indépendants deux à deux, donc $\xi \sim \mathcal{N}_q(0, \Sigma)$ où $\Sigma = \text{diag}\{\sigma_k^2 A_k\}$;
- U la matrice d'incidence connue associée aux effets aléatoires, de dimension $(N \times q)$. Elle est formée des matrices d'incidence sous-jacente U_k de dimension $(N \times q_k)$ telle que $U = \begin{bmatrix} U_1 & \dots & U_K \end{bmatrix}$;
- ε est le vecteur aléatoire des erreurs de tailles N tel que $\varepsilon \sim (0, \sigma_\varepsilon^2 V)$ avec V une matrice connue, et ε et $(\xi_k)_{k=1, \dots, K}$ sont indépendants.

1.4. État de l'art des approches statistiques pour l'analyse de la HRQoL dans les essais cliniques en cancérologie

Les modèles linéaires mixtes (*linear mixed model*, LMM) constituent un outil puissant et flexible pour analyser des données groupées telles que les données longitudinales ou les mesures répétées sur les individus. Ils permettent d'introduire deux niveaux d'analyse des données : un niveau global (fini) commun à l'ensemble des individus par les effets fixes et un niveau spécifique (infini) traduit par les effets aléatoires. Les paramètres d'effets fixes sont communs à l'ensemble des individus alors que les effets aléatoires varient d'un individu à l'autre. L'ajout d'effets aléatoires dans la modélisation constitue un moyen élaboré d'étudier la variabilité des données en précisant les diverses sources de variation. Cela permet de séparer la variance totale en deux parties : la variation due aux effets aléatoires et celle due à l'erreur.

Le LMM est actuellement le modèle le plus utilisé pour l'analyse des scores de HRQoL en cancérologie (Ferrandina et al., 2014; Pan et al., 2012; Schaake et al., 2013). Cette méthode permet de prendre en compte l'aspect longitudinal des données, l'influence des covariables sur la variable réponse et la dépendance des données d'un même individu. Soit S_{iv} la variable réponse (score de HRQoL) concernant l'individu i ($i = 1, \dots, n$) au temps $\{t_{iv}\}_{v=1, \dots, n_i}$, où n_i est le nombre de mesures répétées pour l'individu i . De manière générale pour les analyses longitudinales, le LMM est de la forme suivante :

$$\begin{cases} S_{iv} &= \theta_i(t_{iv}) + \varepsilon_{iv}, \quad \varepsilon_{iv} \sim \mathcal{N}(0, \sigma_\varepsilon^2) \\ \theta_i(t_{iv}) &= x_i(t_{iv})' \beta + u_i(t_{iv})' \xi_i, \quad \xi_i \sim \mathcal{N}(0, \Sigma) \end{cases}, \quad (1.6)$$

où β est le vecteur des effets fixes, ξ_i celui des effets aléatoires associés spécifiques à l'individu i et ε_{iv} est le terme d'erreur. Cette modélisation permet d'analyser directement les changements de la variable réponse S au cours du temps, puis de décomposer la variabilité des données. L'approche intuitive est de considérer une tendance moyenne pour l'évolution au cours du temps (partie fixe) et une tendance spécifique pour chaque sujet au cours du temps qui diverge autour de la tendance moyenne (partie aléatoire) (Rizopoulos, 2012; Agresti, 2010). En général, pour les analyses longitudinales, deux effets aléatoires sont considérés : ξ_{0i} et ξ_{1i} respectivement les effets aléatoires liés à l'ordonnée à l'origine et à la pente. Par exemple, la modélisation longitudinale classique du score dans le but de comparer l'évolution de la HRQoL au cours du temps entre deux groupes est :

$$\begin{cases} S_{iv} &= \theta_i(t_{iv}) + \varepsilon_{iv} \\ \theta_i(t_{iv}) &= \beta_0 + \xi_{0i} + \beta_1 g_i + t_{iv} (\beta_2 + \beta_3 g_i + \xi_{1i}) \end{cases}, \quad (1.7)$$

avec :

- β_0 est l'intercept lié au groupe de référence $g_i = 0$;
- $\beta_0 + \beta_1$ est l'intercept lié au groupe concurrent $g_i = 1$;
- β_2 pente du groupe de référence ;
- $\beta_2 + \beta_3$ pente du groupe concurrent ;
- $\xi_i = (\xi_{0i}, \xi_{1i})'$ est le vecteur des effets aléatoires d'espérance nulle et de matrice de covariance Σ .

Ces modèles sont avantageux pour estimer et analyser l'influence de certaines variables explicatives dans l'étude de données répétées. La partie fixe du modèle (1.7) permet de distinguer les évolutions moyennes des deux groupes. Concernant la partie

aléatoire, ξ_0 permet de considérer que les sujets n'ont pas le même niveau de HRQoL à $t = 0$; ξ_1 permet de prendre en compte le fait que la HRQoL n'évolue pas de la même manière entre les sujets.

Les LMM sont très flexibles suivant les situations. En effet, différentes couches de dépendance peuvent être considérées dans les données de manière emboîtée. Par exemple, si les patients ne sont pas traités dans le même centre et que cette variabilité veut être prise en compte, le modèle 1.7 s'écrit alors :

$$\begin{cases} S_{it} = \theta_i(t) + \varepsilon_{it} \\ \theta_i(t) = \gamma_i(t) + \xi_c \\ \gamma_i(t) = \beta_0 + \xi_{0i} + \beta_1 g_i + t(\beta_2 + \beta_3 g_i + \xi_{1i}) \end{cases}, \quad (1.8)$$

avec $\xi_c \sim \mathcal{N}(0, \sigma_c^2)$ un effet aléatoire a priori indépendant des autres, lié au centre dans lequel est traité le patient, modélisant la variabilité des mesures entre les centres. Pour les modélisations (1.7) et (1.8), la variable latente qui représente le concept mesuré dans l'équation (1.1) ne référence plus un effet aléatoire, mais à une combinaison linéaire d'effets fixes et aléatoires.

Dans le modèle classique (1.7), la contrainte linéaire de l'évolution de la HRQoL peut sembler stricte dans la mesure où la HRQoL est dynamique et peut être sujette à des fluctuations spontanées ou temporaires, dues à la succession d'état de santé des patients. Toujours concernant les LMM, une solution plus souple est de considérer une modélisation polynomiale ou linéaire par morceaux, relative aux modèles à courbes de croissance. La première présente une limite quant à l'interprétation clinique des termes d'ordre élevé. En revanche, la seconde a l'avantage de pouvoir distinguer des périodes où l'évolution de la HRQoL est susceptible de changer (pendant le traitement versus le suivi post-traitement).

Dans ce contexte, une hypothèse fondamentale des LMM est la normalité de la variable réponse. Or, le score de HRQoL étant calculé sur la base d'une moyenne de données ordinales (maximum 5 items pour le questionnaire QLQ-C30), la modélisation du score de HRQoL par un LMM ne semble pas adéquate. Concernant les données de QdV, les LMM ne sont pas adaptés d'un point de vue méthodologique, et d'autant plus avec un possible effet plafond ou plancher des données dû à la propriété bornée du score.

Depuis quelques années, une alternative aux modèles mixtes pour l'analyse longitudinale de la HRQoL en oncologie, consiste à utiliser des modèles de survie. Les cliniciens plébiscitent ces méthodes qui leur sont beaucoup plus familières.

1.4.2 Modèle de survie : temps jusqu'à la détérioration du score

L'analyse des données de survie s'intéresse à l'apparition d'un événement au cours du temps (décès, récurrence, etc . . .). Dans le contexte de la HRQoL, il est nécessaire de définir l'événement susceptible d'être la cause de l'effet du traitement sur la HRQoL. Par convention, l'événement considéré est une détérioration de la HRQoL (diminution des fonctions et de la santé globale, ou augmentation des symptômes). Un travail approfondi de l'approche temps jusqu'à détérioration d'un score de HRQoL a été mené dans les essais de phase III par [Bonnetain et al. \(2010\)](#). Cette approche consiste

1.4. État de l’art des approches statistiques pour l’analyse de la HRQoL dans les essais cliniques en cancérologie

à étudier le temps jusqu’à détérioration (TJD) d’un ou plusieurs scores de HRQoL. L’utilisation de ce type de modèle suggère que la détérioration du score reflète la détérioration de l’état de santé du patient.

Plusieurs définitions du TJD ont été proposées. Elles dépendent du score de référence (score à l’inclusion, score antérieur le plus élevé ou score antérieur à la mesure considérée), des règles de censures intégrant ou non le décès et/ou l’absence de mesure à l’inclusion et/ou l’absence de suivi comme évènement, et de la différence minimale cliniquement importante (DMCI). La DMCI est définie comme la plus petite différence dans un score de HRQoL perçue comme ayant un sens clinique pour le patient (Jaeschke et al., 1989). Dans l’objectif de capter la signification subjective d’un changement auprès de patients, Osoba et al. (1998) ont montré qu’une modification de 5 à 10 points pour un score QLQ-C30 donné correspond au plus petit changement ayant un sens clinique. Une évolution de 10 à 20 points équivaut à un changement modéré et plus de 20 points à un changement élevé. Le choix de la DMCI est déterminant dans la définition du TJD. En pratique, une différence de 5 à 10 points est considérée comme cliniquement significative pour les questionnaires de l’EORTC (Bonnetain et al., 2010; Hamidou et al., 2011b; Gourgou-Bourgade et al., 2012), et de 5 à 7 points pour le score global des questionnaires du groupe FACT (Wu et al., 2011).

La définition la plus intuitive concerne le TJD. Il s’agit de l’intervalle de temps entre la date d’inclusion (ou date de randomisation) et l’apparition d’une première détérioration significative du score de HRQoL par rapport au score à l’inclusion (Hamidou et al., 2011b). En situation adjuvante, c’est la détérioration temporaire qui est utilisée. À l’inverse, en situation palliative ou avancée, il paraît plus pertinent d’étudier une détérioration définitive (temps jusqu’à détérioration définitive, TJDD) (Anota et al., 2013). Le TJDD d’un score de HRQoL a été défini comme l’intervalle de temps entre la date de randomisation et l’apparition d’une première détérioration significative du score de HRQoL par rapport au score à l’inclusion, sans amélioration ultérieure significative par rapport au score à l’inclusion (Bonnetain et al., 2010).

Cette approche tend à être la plus utilisée en oncologie pour l’analyse longitudinale de la HRQoL dans les essais cliniques. En effet, elle séduit de nombreux cliniciens puisqu’elle propose des résultats compréhensibles et interprétables de par la familiarité de la méthodologie employée. Elle présente de nombreux avantages. Sa grande flexibilité due aux différents scores de référence possibles ou encore à l’ajout du décès comme évènement lui permet de s’adapter à de nombreuses situations thérapeutiques. Cependant, les dates de remplissage des questionnaires sont prévues à l’avance et sont effectuées à intervalles de temps réguliers et cela peut influencer l’estimation de Kaplan-Meier. Enfin, la méthode du TJD pourrait également rencontrer des problèmes liés aux risques compétitifs, que ce soit entre les scores des différentes dimensions, ou entre les scores de HRQoL et le décès.

1.5 Problématiques de l'analyse longitudinale de la HR-QoL

L'étude de la HRQoL est assez ancienne. Cependant, son intérêt pour l'aide à la décision est de plus en plus marqué. La recherche statistique autour de ce critère suit différentes directions associées aux différentes problématiques rencontrées. Cette section présente les principaux axes de recherche et freins méthodologiques rencontrés pour l'analyse longitudinale de la HRQoL. Un premier axe de recherche est associé à la problématique des données manquantes pour les analyses longitudinales comme transversales.

1.5.1 Les données manquantes

Une des premières problématiques dans l'analyse longitudinale de la HRQoL est l'occurrence des données manquantes. Les données manquantes découlent de situations diverses. Dans les essais cliniques en oncologie, la HRQoL est mesurée depuis l'inclusion du patient jusqu'à sa sortie d'étude. Cette dernière peut correspondre à la date de la fin du suivi, à une progression de la tumeur ou encore au décès du patient. Suite à cette sortie d'étude, aucune mesure de HRQoL n'est attendue. Ces données manquantes sont dites prévues. Ensuite, parmi les données manquantes non prévues, nous distinguons celles associées à l'échelle de l'item et celles associées au questionnaire. Différents taux de remplissage sont alors décrits à chaque temps de mesure : le taux de questionnaires manquants ou totalement vides, le taux de questionnaires complétés ou partiellement complétés (par rapport à l'ensemble des patients ou par rapport à ceux encore présents dans l'étude), et le taux d'items manquants par individu et par questionnaire. Une dernière caractéristique est le fait que les données manquantes soient monotones ou intermittentes. Si la donnée manquante est observée à un ou plusieurs temps de mesures donnés et qu'elle est non manquante à un temps ultérieur, alors elle sera considérée comme intermittente. Au contraire, si elle n'est plus observée à partir d'un certain temps de mesure donné, alors la donnée manquante sera considérée comme monotone.

La problématique des données manquantes résulte des données qui ne sont pas manquantes par hasard et portent de l'information quant au niveau de HRQoL du patient. Dans ce sens, [Little and Rubin \(2002\)](#) ont proposé de classifier les données manquantes suivant trois types de données : manquantes complètement aléatoirement (*missing completely at random*, MCAR), manquantes au hasard (*missing at random*, MAR) et manquantes non aléatoirement (*missing not at random*, MNAR). Dans le meilleur des cas, les données manquantes sont le simple fruit du hasard et indépendantes de covariables ou de l'état de santé du patient : elles sont donc non-informatives et considérées MCAR. Par exemple, les patients peuvent simplement oublier de répondre à un item, négliger plusieurs questions ou encore oublier de répondre à un questionnaire à un temps donné. En revanche, il peut également arriver que les données manquantes ne soient malheureusement pas manquantes de manière aléatoire et soient donc informatives. Dans ce cas, nous distinguons les données manquantes MAR et celles MNAR. Les données manquantes sont dites MAR si elles

dépendent de variables explicatives (données recueillies à l'inclusion). Par exemple, une personne qui ne répond pas à la question concernant la perte des cheveux peut s'expliquer si cette dernière a une calvitie⁷. Enfin, la dernière famille de données manquantes est celle des données manquantes MNAR. Elles sont considérées comme manquantes non aléatoirement car elles dépendent du niveau de HRQoL non observé du patient. Par exemple, si le patient ne peut remplir un questionnaire à un temps de mesure fixé en raison d'une hospitalisation, d'une fatigue trop importante ou plus généralement d'une détérioration de son état de santé, les données manquantes sont alors dépendantes du niveau de HRQoL et sont considérées comme MNAR.

Les données manquantes de type MAR et MCAR n'interfèrent pas dans l'analyse de la HRQoL car elles ne sont pas informatives sur l'état de santé ou encore le niveau de HRQoL. Cependant, elles peuvent engendrer une perte de puissance dans l'analyse, due au fait qu'il y ait moins d'observations que attendues. Dans la situation où la donnée manquante est informative, il semble alors pertinent de prendre en considération cette information dans l'évaluation de la HRQoL. En pratique, il peut s'avérer compliqué de déterminer la cause de la donnée manquante cas par cas. Outre le décès, il serait nécessaire d'en déterminer la cause selon l'information disponible sur le patient pour s'assurer du lien (ou pas) avec le niveau de HRQoL. Plusieurs méthodes ont été proposées pour prendre en compte les données manquantes de type MAR, telles que l'imputation multiple ou encore l'utilisation d'algorithme EM (Dempster et al., 1977). Concernant les données manquantes de type MNAR, l'utilisation de valeurs arbitraires (« Worst case ») peut être envisagée (Anota, 2014). Cependant, cette méthode est controversée.

1.5.2 Le phénomène de « Response Shift »

Un phénomène souvent observé dans les données rapportées par les patients au cours du temps est le phénomène de « Response Shift » (RS) (Howard et al., 1979). Le phénomène de RS est induit par un événement qui amène un individu à modifier sa manière de répondre à un questionnaire au cours du temps. Dans le contexte des essais cliniques, la notion de RS traduit le processus d'adaptation du patient autour de sa maladie (acceptation de la maladie elle-même, adaptation à la toxicité des traitements) et peut se révéler comme un biais potentiel dans l'analyse de l'effet différentiel du traitement sur la HRQoL. En effet, la HRQoL étant auto-évaluée par le patient, elle peut donc être influencée par des interactions, des phénomènes psychologiques ou des facteurs environnementaux. Les priorités du patient peuvent changer, ou bien celui-ci peut revoir sa propre échelle de valeurs au cours du temps. Schwartz and Sprangers (1999) définissent la RS comme un ensemble de trois composantes correspondant à un changement :

- dans les références internes du patient (« recalibration »), i.e. il ne répond pas de la même manière aux items au cours du temps ;
- dans l'importance des items : « reprioritization ») i.e. les patients accordent plus d'importance pour une dimension ou un item qu'auparavant pour évaluer leur HRQoL ;

7. Exemple tiré par les cheveux ;

- dans la redéfinition du concept de la HRQoL i.e. le concept de HRQoL est réévalué (« reconceptualization »).

Lors de chaque mesure, le changement observé ne correspond alors pas nécessairement à un effet temps ou à un effet traitement si la RS s'opère. Il est à présent démontré que le phénomène de RS interfère dans l'évaluation de la HRQoL (Hamidou et al., 2011a), mais ce concept est finalement peu ou pas intégré dans les stratégies d'analyse longitudinale proposées. Sa prise en compte permettrait ainsi d'évaluer l'impact réel de la maladie ou du traitement au cours du temps mais nécessite d'être planifiée dès la conception de l'essai.

Deux écoles s'opposent pour la considération de la RS dans l'analyse de la HRQoL. La première est celle qui veut capter cet effet afin d'évaluer l'impact réel de la maladie ou du traitement au cours du temps. La seconde est celle qui considère que la RS est indissociable du réel effet recherché, car c'est un mécanisme sujet-dépendant. L'analyse de la HRQoL est alors réalisée pour évaluer les changements de la perception de HRQoL des patients au cours du temps. L'accent est mis ici sur le ressenti du patient et comment celui-ci évolue. Dans ce travail, nous nous plaçons comme partie intégrante de la seconde école. Nous ne prendrons pas en compte le processus RS dans les modélisations que nous présentons. Par abus de langage, nous mentionnons l'analyse de la HRQoL à la place de celle de la perception de HRQoL qui est vraiment analysée et recherchée.

1.5.3 Limites du score de HRQoL

Malgré le fait que le score de HRQoL présente certains avantages pratiques (interprétation intuitive, outils validés et utilisés en routine), des limites méthodologiques et conceptuelles existent. Les réponses aux items sur lesquelles le score est construit proviennent d'échelles de Likert avec des réponses qualitatives (ordinales). Considérer ces réponses comme des réponses quantitatives revient à attribuer des valeurs aux réponses de manière arbitraire. Pour les questionnaires de l'EORTC, chaque modalité adjacente est séparée d'une unité (Figure 1.5a) ce qui renvoie à une même différence entre les états adjacents de l'échelle ($a = b = c$). Or, d'un point de vue conceptuel, il est difficile d'affirmer que la difficulté de passer de "Assez" à "Beaucoup" est la même que celle de passer de "Un peu" à "Assez" (Bradburn et al., 1979). Travailler sur le score de HRQoL est sujet à discussion (Gorter et al., 2015). En effet, des patients peuvent avoir le même score sans pour autant avoir choisi les mêmes réponses aux items. La Figure 1.5b illustre les arguments avancés où deux patients ont obtenus le même score de 11 points associé à leur capacité fonctionnelle physique. Le fait de considérer un niveau identique de HRQoL pour ces deux patients revient implicitement à considérer un poids similaire pour chaque item dans le regroupement.

Pour finir, analyser une dimension de HRQoL caractérisée par un seul item à travers les approches présentées paraît absurde. En effet, dans ce cas de figure, cela revient à considérer quatre valeurs pour le score ($S = 0, 33.33, 66.67, 100$) associées aux quatre modalités de réponse. Concernant le TJD, le DMCI est au minimum

1.5. Problématiques de l'analyse longitudinale de la HRQoL

	Pas du tout	Un peu	Assez	Beaucoup
2. Avez-vous des difficultés à faire une <u>longue</u> promenade?	1	2	3	4
		+a	+b	+c
		1 →	2 →	3 → 4
(a) Échelle de Likert				
	Pas du tout	Un peu	Assez	Beaucoup
1. Avez-vous des difficultés à faire certains efforts physiques pénibles comme porter un sac à provision chargé ou une valise?	1	2	3	4
2. Avez-vous des difficultés à faire une <u>longue</u> promenade?	1	2	3	4
3. Avez-vous des difficultés à faire un <u>petit</u> tour dehors?	1	2	3	4
4. Etes-vous obligée de rester au lit ou dans un fauteuil pendant la journée?	1	2	3	4
5. Avez-vous besoin d'aide pour manger, vous habiller, faire votre toilette ou aller aux toilettes?	1	2	3	4
(b) Dimension physique				

FIGURE 1.5: Dimension physique du questionnaire EORTC QLQ-C30. Les cercles bleus et verts présentent les réponses fictives de deux individus. Les deux scores obtenus sont identiques et sont égaux à 11.

équivalent à l'écart entre deux modalités. L'approche par LMM est encore moins appropriée du fait du non-respect de l'hypothèse de normalité. Dans ce cas, la modélisation appropriée est alors celle d'un modèle linéaire généralisé mixte (*generalized linear mixed model*, GLMM) pour données ordinales qui prend en compte directement les données brutes (réponses à l'item). Les données issues des questionnaires ont fait l'objet de nombreux travaux. Ils ont donné lieu aux modèles issus de la théorie de réponse aux items (*item response theory*, IRT) qui prennent en compte la nature des données brutes.

1.5.4 Multiplicité des tests

Pour l'étude d'un traitement dans un essai clinique, l'analyse exhaustive de la HRQoL pose le problème de la multiplicité des tests. En effet, l'influence des variables explicatives, telle que le traitement, est testée pour chacune des dimensions de HRQoL et de manière indépendante. Sans ajustement du risque de première espèce, le risque global α_G de conclure à tort une différence sur les minimums $k = 15$ dimensions est : $\alpha_G = 1 - (1 - \alpha)^k \approx 0.54$ pour $\alpha = 5\%$.

Plusieurs solutions s'offrent alors aux statisticiens : ajuster le seuil α pour contrôler le risque global ou/et faire un choix sur les dimensions précises à étudier en priorité en accord avec le clinicien. Dans un essai clinique randomisé, la connaissance des cliniciens sur les traitements étudiés peut permettre de focaliser ou non l'analyse de ce critère sur une ou plusieurs dimensions particulières de HRQoL. Par exemple, dans l'essai clinique Prodigé4/ACCORD11, le Folfirinox est connu pour accentuer le symptôme de diarrhée. Connaître la (les) toxicité(s) et les effets secondaires va alors

permettre de cibler une (des) dimension(s) particulière(s) de HRQoL pour l'analyse, en plus du statut global de santé (*global health statut*, GHS) qui est toujours analysé. Dans ce cas, un nombre plus restreint d'analyses indépendantes est effectué. En revanche, si aucune connaissance ou spécificité des produits testés n'est connue, l'analyse s'effectue alors de manière globale : sur chaque dimension de la HRQoL et de manière indépendante. Très récemment, un design associant la HRQoL à l'OS comme co-critère principal a été proposé dans lequel l'analyse de la HRQoL est basée sur trois dimensions de HRQoL réduisant ainsi la dépense de α_G de manière raisonnable. Le choix de ces dimensions s'est dirigé naturellement sur le GHS qui représente la HRQoL globale, la fonction physique qui comprend le plus d'items et la fatigue qui reste le symptôme le plus général et le plus représenté en nombre d'items.

Aujourd'hui, la prise en compte de la HRQoL dans le design d'un essai clinique pour comparer deux bras de traitement réside dans le choix de quelques dimensions de HRQoL. Une alternative au fait de choisir des dimensions spécifiques est la considération de scores plus globaux. Dans la littérature, des scores globaux ont été proposés pour remplacer le statut global de santé (Nordin et al., 2001; Phillips et al., 2015). Ils sont basés sur la somme de tous les scores tout comme le score global du FACT-G. Nous pouvons alors très bien imaginer un score global fonctionnel et un score global symptomatique permettant de restreindre l'analyse au statut fonctionnel et au symptomatique des patients.

1.6 Objectifs

Les travaux présentés dans ce manuscrit s'inscrivent dans un axe de recherche concernant l'analyse longitudinale de la HRQoL en cancérologie. Nous nous sommes particulièrement concentrés sur l'étude de différents types de modèles de régression pour proposer les stratégies d'analyse les plus adaptées aux situations rencontrées dans les essais thérapeutiques ou dans des études de la HRQoL en cancérologie. Les travaux réalisés ont été guidés par trois objectifs :

1. Modéliser directement les données brutes (réponses multi-items) pour palier aux limites du score de HRQoL ;
2. Prendre en compte l'aspect multidimensionnel afin d'étudier globalement la HRQoL et réduire le nombre d'analyses ;
3. Distinguer différentes sous-populations homogènes suivant la trajectoire de HRQoL, et définir des profils de patient associés afin d'aider à une meilleure prise en charge.

Nous proposons dans la première partie (chapitres 2 et 3) de se concentrer sur la modélisation statistique à utiliser pour l'analyse des dimensions de HRQoL dans les essais cliniques, et plus particulièrement pour prendre en compte directement les données brutes. Les modèles IRT font référence aux modèles développés pour l'analyse directe des données issues de questionnaires psychométriques. Ils ont vu le jour à la moitié du vingtième siècle et tendent de plus en plus à se démocratiser. Tout comme l'approche CTT, l'IRT ne réfère pas à une théorie au sens propre du terme,

mais plutôt à une approche conceptuelle permettant d'estimer une variable latente à partir des mesures issues des questionnaires. L'intérêt pour ces modèles est grandissant aux vues des inconvénients du score soulignés dans la partie précédente. Ils sont particulièrement utilisés pour l'étude de la RS (Anota et al., 2014b; Guilleux et al., 2015) et ont fait l'objet de nombreuses études de simulation telles que : comparaison avec les modèles CTT (Blanchin et al., 2011; Bock et al., 2013), étude de la puissance (Blanchin et al., 2013; Anota et al., 2014a) et recommandation du nombre de sujets nécessaires dans la comparaison de groupes (Sébille et al., 2014; Hardouin et al., 2015). En revanche, ils ne sont que rarement utilisés en pratique pour l'analyse longitudinale sur des données réelles (Liu and Hedeker, 2006) et en particulier dans les essais cliniques (Douglas, 1999; Glas et al., 2009).

Le chapitre 2 traite de la modélisation IRT et permet notamment d'introduire les notions et notations qui seront utilisées par la suite. La section 2.2 décrit les différents modèles de référence pour l'étude de données dichotomiques et polytomiques. Ces modèles sont d'abord présentés pour des analyses transversales avec la paramétrisation la plus connue. Ensuite, plusieurs variantes sont détaillées pour en venir à celle utilisée pour l'analyse longitudinale de la HRQoL. Enfin, la section 2.3 présente une application des modèles IRT pour l'analyse des données de HRQoL dans l'essai clinique Prodigé 4/ACCORD 11 qui a comparé les traitements Gemcitabine versus Folfirinox dans le cancer du pancréas métastatique. L'analyse de ces données via le LMM est également proposée.

Le chapitre 3 propose une réflexion sur le choix du modèle IRT le plus adapté et à recommander en pratique. Pour cela, nous replaçons les modèles IRT dans le contexte des GLMM. En effet, les modèles IRT sont des GLMM pour données catégorielles qui se caractérisent par une paramétrisation du prédicteur linéaire qui leur est spécifique. Parmi les GLMM pour données ordinales, il existe trois familles de modèles de régression : la famille des modèles adjacents, la famille des modèles cumulatifs et la famille des modèles séquentiels. Ces trois familles font écho aux trois différentes modélisations IRT pour données polytomiques ordinales, respectivement le *partial credit model* (PCM), le *graded response model* (GRM) et le *sequential Rasch model* (SRM). Une fois les modèles présentés, nous discutons dans la section 3.3 de leur pertinence pour l'analyse longitudinale de la HRQoL en cancérologie sur la base d'arguments pratiques et méthodologiques. Le choix parmi ces modèles sera appuyé par les résultats présentés dans l'article relatif à ce chapitre (visible en section 3.5).

Nous proposons dans la seconde partie (chapitre 4 et 5) de s'éloigner du contexte conventionnel des essais cliniques en cancérologie. Dans le chapitre 4, nous proposons une approche globale prenant en compte l'aspect multidimensionnel et longitudinale du critère de HRQoL. Afin de considérer l'intégralité de l'information recueillie par les questionnaires, nous avons développé un modèle à équation structurelle (*structural equation modeling*, SEM) qui prend en compte les relations conceptuelles entre les dimensions de HRQoL du questionnaire EORTC QLQ-C30. De manière transversale, ce modèle permet de résumer l'information apportée par les dimensions fonctionnelles et les dimensions symptomatiques au travers de deux facteurs qui expliquent le GHS. Pour cela, trois équations de régression inter-connectées par l'intermédiaire

des facteurs sont nécessaires. Dans chacune d'elle, l'ajout de variables explicatives est rendu possible. L'estimation des paramètres mais également celle des facteurs sujet-spécifiques du modèle sont réalisées avec l'algorithme EM. La section 4.2.2 présente les résultats de simulation afin de valider le modèle développé, puis d'étudier l'information portée par le facteur par rapport à celle portée par la simple somme des scores. L'aspect longitudinal dans l'approche proposée est réalisé par l'utilisation de LMM permettant d'étudier l'évolution des facteurs ou du statut global de santé. Une conséquence et un avantage de cette approche, est la possibilité de réduire le nombre d'analyses longitudinales pour répondre au problème de la multiplicité des tests décrit en section 1.5.4. Une application sur des données cliniques de cette approche est présentée dans l'article associé à ce travail et visible en section 4.5.

Enfin, dans le chapitre 5, nous nous sommes intéressés à une complexité supplémentaire des données de HRQoL en cancérologie : la présence d'une hétérogénéité non observée au sein de la population. Concernant l'analyse longitudinale, la population d'intérêt peut être composée des sous-populations homogènes qui se distinguent par les trajectoires de HRQoL, mais qui s'expliquent par des profils particuliers non observés. Identifier les profils de patients associés aux différentes évolutions de la HRQoL pourrait ainsi aider à l'amélioration de la prise en charge des malades. Un premier objectif de ce chapitre va être de déterminer les classes latentes (sous-populations homogènes) caractérisées par des trajectoires différentes de HRQoL. Pour cela, nous utiliserons des mélanges de modèles mixtes présentés dans la section 5.2. Une fois la classification obtenue, le dernier objectif de cette approche sera de décrire l'appartenance aux classes latentes suivant des profils de patients associés. Ce travail a été motivé par une étude psychosociale réalisée sur des patientes atteintes de cancer du sein. L'analyse de ces données réalisée en fonction de ces approches sera présentée en section 5.3.

Première partie

La théorie de réponse aux items pour l'analyse longitudinale de la HRQoL

Présentation et application sur données réelles

Dans ce chapitre, l'objectif est d'analyser directement les données brutes des questionnaires en utilisant des modèles issus de la théorie de réponse à l'item (*Item Response Theory*, IRT). Une première partie est consacrée à une présentation générale des modèles IRT proposés dans la littérature. Une fois les notions et les notations relatives à l'IRT introduites, ces modèles seront appliqués pour l'analyse longitudinale de la HRQoL dans un essai clinique.

Sommaire

2.1	Introduction	26
2.2	Présentation des modèles IRT	28
2.2.1	Modèles IRT pour données dichotomiques	28
2.2.2	Modèles IRT pour données polytomiques	31
2.2.3	Variantes et paramétrisation	34
2.3	Application	37
2.3.1	Résultats et estimations des modèles	38
2.3.2	Sensibilité des estimations du modèle LPCM suivant les observations	41
2.4	Discussion	41
2.5	Article : <i>Applying the longitudinal model from item response theory to assess the health-related quality of life in the PRODIGE 4/ACCORD11 randomized trial</i>	42

2.1 Introduction

À ce jour, les analyses longitudinales de la HRQoL dans un essai clinique en oncologie réalisées avec une modélisation statistique directement sur les données brutes, sont peu fréquentes (Douglas, 1999; Glas et al., 2009). La théorie de l’IRT repose sur l’idée de l’échelle de *Guttman* caractérisée par un continuum représentant le concept à évaluer, non directement mesurable mais accessible par un ensemble d’items. Les deux extrémités du continuum sont les deux extrêmes du concept, et les individus et les items sont placés tous deux sur cette même échelle. La position des items sur cet axe est caractérisée par des seuils qui quantifient la difficulté de répondre positivement à l’item. La position des individus découle quant à elle des différentes réponses observées, conditionnellement à la position des seuils relatifs aux items. Cette échelle n’a de sens que si les items et individus figurent tous deux sur le continuum. Enfin, la capacité de l’individu est déterminée par sa position en comparaison avec celles des items et des autres individus.

D’un point de vue déterministe, la position des individus sur le continuum peut être directement déduite à partir des réponses observées et suivant les seuils liés aux items. À l’inverse, les réponses aux items peuvent être directement déduites suivant le niveau latent de l’individu (position sur le continuum) et les seuils. Dans le cas de réponses binaires, la réponse négative 0 (respectivement positive 1) est observée si le niveau est inférieur (respectivement supérieur) au seuil associé à l’item considéré. La Table 2.1 présente les observations relatives à la dimension de vulnérabilité (exemple complètement fictif). Ces réponses permettent de classer les individus suivant leur vulnérabilité, i.e. du moins au plus vulnérable : *Hulk* (très faible vulnérabilité), *Ironman* (faible vulnérabilité), *Captain America* (moyenne vulnérabilité) et la personne *lambda* (haute vulnérabilité).

TABLE 2.1: Détails de l’exemple fictif concernant la mesure de la vulnérabilité. Les réponses de quatre sujets à trois items dichotomiques sont considérées. La réponse positive « oui » est associée à la valeur 1, et la réponse négative « non » à la valeur 0.

Êtes-vous vulnérable...	<i>Hulk</i>	<i>Ironman</i>	<i>Captain America</i>	Personne <i>Lambda</i>
1. ... à l’épreuve des bombes ?	0	1	1	1
2. ... à l’épreuve des balles ?	0	0	1	1
3. ... à l’épreuve des coups ?	0	0	0	1

La Figure 2.1 illustre le continuum sous-jacent avec la position des individus et celle des seuils relatifs aux items. Les seuils sont ordonnés entre eux suivant leur niveau de difficulté à y répondre positivement, et les individus sont positionnés en fonction de leurs réponses.

Pour assouplir ce point de vue déterministe, une modélisation probabiliste a été proposée par Rasch (Rasch, 1960). Ses travaux ont été les précurseurs des modèles IRT étudiés aujourd’hui. L’approche IRT se réfère à un ensemble de modèles proba-

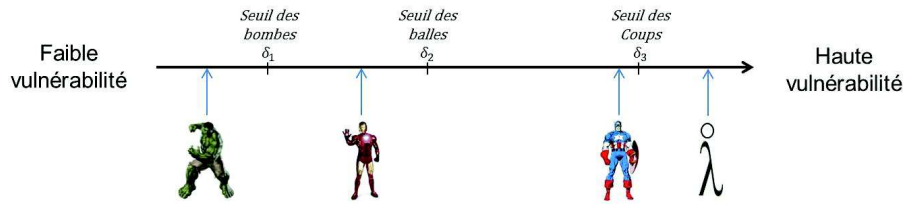


FIGURE 2.1: Continuum lié au concept de vulnérabilité. Les réponses aux questions sont déterminées par le niveau de l'individu sur ce continuum. Les différents items et réponses observées concernant la dimension de vulnérabilité sont visibles en Table 2.1

bilistes qui permet de faire le lien entre les réponses aux items observées et le concept sous-jacent non directement mesurable qu'elles sont supposées refléter. Ce concept d'intérêt est caractérisé par une variable latente, appelée également dans la littérature trait latent. Dans l'IRT, elle représente généralement un concept hypothétique (trait, domaine, capacité) qui est supposé exister mais ne peut pas être mesuré directement comme un unique élément observable (Fox, 2010). Ce concept est indirectement mesuré au travers de plusieurs items et peut représenter n'importe quelle construction mesurable, comme la santé mentale, la fatigue, le fonctionnement émotionnelle, ou plus largement la HRQoL ou l'intelligence (Fayers and Machin, 2007). Plus précisément, les modèles IRT sont des équations mathématiques décrivant l'association entre le niveau de chaque individu (patient) quantifié par la variable latente, et la probabilité de répondre une modalité particulière à un item. Cette variable latente modélise la dépendance entre les réponses multi-items provenant d'un même individu (voir Figure 2.2). Conditionnellement à celle-ci, chacune de ces réponses sont considérées comme indépendantes (Verhagen and Fox, 2013). La valeur de la variable latente pour un individu est supposée être le seul facteur influençant les probabilités de réponse aux items.

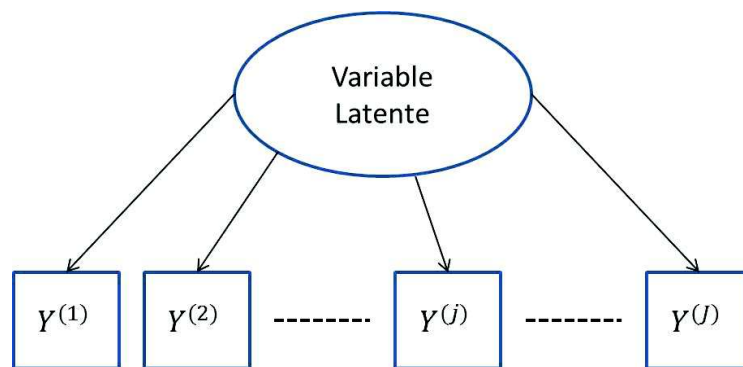


FIGURE 2.2: Schéma représentant le lien entre la variable latente et les réponses observées $(Y^{(j)})_{j=1,\dots,J}$ des J items. La variable latente explique les réponses aux différents items, et les différentes réponses reflètent la variable latente.

Les modèles IRT sont également utilisés pour valider un questionnaire en cours de développement (Edelen and Reeve, 2007), tester la validité d'un questionnaire déjà existant et déjà validé par la théorie classique (Jafari et al., 2012), ou encore

réduire le nombre d'items d'un questionnaire (Petersen et al., 2006). Aujourd'hui, nous rencontrons également ces modèles pour l'analyse de données issues de questionnaire (Edelen and Reeve, 2007; Hardouin et al., 2015), et ils suscitent de plus en plus d'intérêts pour des analyses longitudinales (Gorter et al., 2015).

Deux familles de modèles se distinguent dans l'IRT : les modèles de la famille de Rasch et les modèles de la famille de Lord. Les modèles de la famille de Rasch considèrent que ce sont les données qui doivent s'adapter au modèle alors que pour les modèles de la famille de Lord, ce sont les modèles qui doivent s'adapter aux données (Anota, 2014). En d'autres termes, ces modèles comprennent des paramètres d'items permettant de distinguer les diverses réponses observées supposées refléter un même concept général non directement mesurable. Les modèles de la famille de Lord considèrent d'avantage de paramètre d'item, induisant une plus grande flexibilité suivant les données observées. Les modèles de la famille de Rasch imposent plus de contraintes sur ces paramètres, leur permettant de vérifier des hypothèses qui leur sont propres.

2.2 Présentation des modèles IRT

2.2.1 Modèles IRT pour données dichotomiques

Les modèles IRT pour items dichotomiques correspondent aux modèles probabilistes dont la variable réponse Y est binaire, codée 0 versus 1 respectivement pour une réponse dite négative ou positive. Parmi ces modèles IRT pour données dichotomiques (binaires), il en existe plusieurs qui se distinguent par différentes paramétrisations (déclinaisons du prédicteur linéaire). Étant donné un questionnaire composé de J items (questions), la probabilité que l'individu i ($i = 1, \dots, N$) réponde positivement ($y_i^{(j)} = 1$) à l'item j ($j = 1, \dots, J$) dépend du niveau de la variable latente θ_i de l'individu i et des paramètres d'item. Le plus connu, notamment pour son appellation, est le modèle de Rasch (1960) :

$$\Pr \left(Y_i^{(j)} = 1 | \theta_i \right) = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)}, \quad (2.1)$$

où $\theta_i \sim \mathcal{N}(0, \sigma^2)$. Il modélise la probabilité que l'individu i a de répondre positivement (réponse 1) à l'item j suivant sa valeur de la variable latente θ_i et le paramètre d'item δ_j . Une extension du modèle (2.1) appelée le *Two-Parameter logistic (2PL) model* (Lord, 1980) est plus complète. Le modèle 2PL est défini tel que :

$$\Pr \left(Y_i^{(j)} = 1 | \theta_i \right) = \frac{\exp\{\alpha_j (\theta_i - \delta_j)\}}{1 + \exp\{\alpha_j (\theta_i - \delta_j)\}}, \quad (2.2)$$

où (δ_j, α_j) les deux types de paramètre d'item, avec $\alpha_j > 0$ est le paramètre d'inclinaison (de pente) de l'item j ($j = 1, \dots, J$) et δ_j le paramètre de position (ou seuil) de l'item j . Afin que le modèle (2.2) soit identifiable, il est nécessaire de fixer un paramètre. Il est courant de fixer la variance de la variable latente (par exemple, $\sigma^2 = 1$), mais un des J paramètres de discrimination peut également être fixé.

La notion de difficulté de l’item est portée par le paramètre δ_j . Il est également appelé paramètre de difficulté de l’item j . Il correspond à la valeur de la variable latente pour laquelle l’individu i a la même chance de répondre positivement ($y_i^{(j)} = 1$) que négativement ($y_i^{(j)} = 0$), soit $\Pr(Y_i^{(j)} = 1|\theta_i) = 0.5$ (voir Figure 2.3a). Dans le contexte d’un questionnaire composé de J items, le paramètre δ_j indique la difficulté de répondre positivement à l’item j par rapport aux autres items sachant la valeur du trait latent. La Figure 2.3a permet de mettre en évidence cette notion. Sachant un θ fixé, il est plus difficile de répondre positivement à l’item 2 (en vert) qu’à l’item 1 (en rouge), et plus difficile de répondre positivement à l’item 3 (en bleu) qu’aux deux autres items. Autrement dit, il est attendu plus de réponse 1 pour les items dits faciles et plus de réponse 0 pour les items associés à un fort paramètre de difficulté.

Le second paramètre d’item α est le paramètre dit de discrimination. Il représente le degré pour lequel les réponses à l’item varient en fonction de la variable latente θ . Pour un ensemble d’item, α_j indique le pouvoir discriminant de l’item j , où plus le paramètre est élevé, plus l’item permet de discriminer deux individus ayant une valeur du trait latent proche (voir Figure 2.3b).

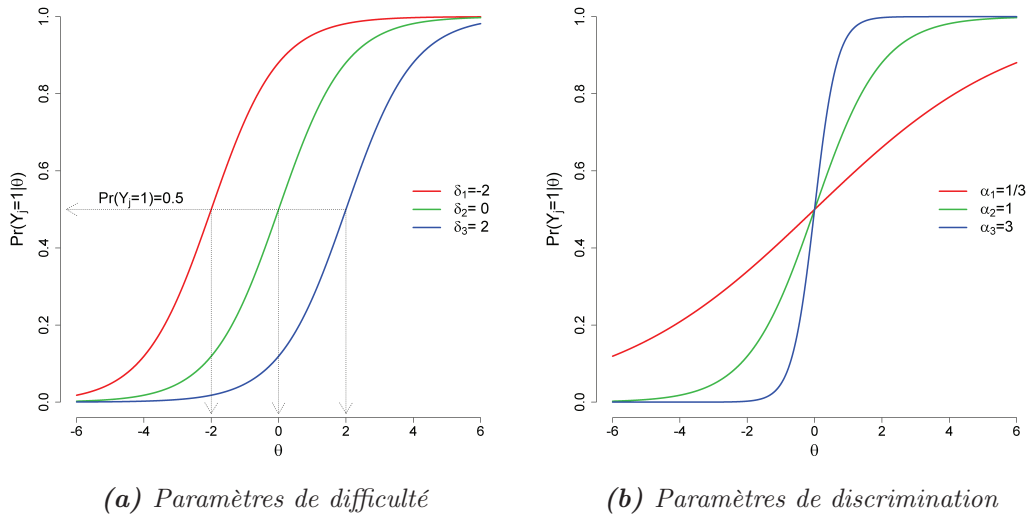


FIGURE 2.3: Courbes caractéristiques aux items (ICC, item characteristic curves) pour modèles IRT associés à des données dichotomiques. La Figure 2.3a montre les ICC des trois items pour le modèle (2.1), avec $\alpha_j = 1$ quel que soit $j = 1, 2, 3$ et les paramètres de difficulté prenant les valeurs $-2, 0$ et 2 . La Figure 2.3a montre les ICC des trois items pour le modèle (2.2), avec $\delta_j = 0$ quel que soit $j = 1, 2, 3$ et les paramètres de discrimination prenant les valeurs $\frac{1}{3}, 1$ et 3 .

Le modèle (2.1) correspond au modèle (2.2) avec la contrainte $\alpha_j = 1$ ($j = 1, \dots, J$). Un modèle intermédiaire, le *One-parameter model (1PL)*, est également proposé, dans lequel le paramètre d’item α serait le même quel que soit l’item considéré, mais non fixé au préalable. La Figure 2.4 montre les différentes contraintes des paramètres d’item pour ces modèles par rapport à la variation de la variable latente.

Une extension du modèle (2.2) comprenant un nouveau paramètre d’item (noté

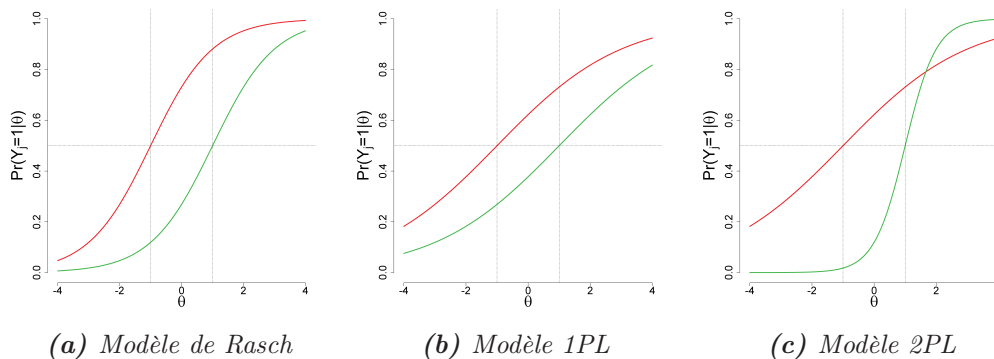


FIGURE 2.4: Courbes caractéristiques aux items pour le modèle de Rasch, le modèle 1PL et le modèle 2PL. Dans chaque cas, il y a deux items ($j = 1, 2$) ayant pour paramètre de difficulté $\delta_1 = -1$ et $\delta_2 = 1$. Pour le modèle de Rasch (2.4a), $\alpha_1 = \alpha_2 = 1$; pour le modèle 1PL (2.4b), $\alpha_1 = \alpha_2 = 0.5$; pour le modèle 2PL (2.4c), $\alpha_1 = 0.5$ et $\alpha_2 = 2$.

γ), appelé paramètre de pseudo-chance, permet d'inclure le potentiel à deviner une bonne réponse (Titman et al., 2013). Ce modèle appelé dans la littérature le *Three-Parameter logistic (3PL) model* (De Ayala, 2009) est le suivant :

$$\Pr\left(Y_i^{(j)} = 1|\theta_i\right) = \gamma_j + (1 - \gamma_j) \frac{\exp\{\alpha_j(\theta_i - \delta_j)\}}{1 + \exp\{\alpha_j(\theta_i - \delta_j)\}}. \quad (2.3)$$

Ce modèle (2.3) permet de prendre en compte la possibilité qu'a un individu de donner au hasard une bonne réponse à une question à choix multiples. Le paramètre $\gamma_j \in [0, 1]$ représente la plus faible probabilité qu'a un individu de donner une réponse correcte à l'item j . Pour les données de HRQoL, il n'y a pas de bonne ou de mauvaise réponse car ce sont des questions concernant le ressenti des patients. Le modèle (2.3) n'est donc pas adapté pour cette analyse.

Hypothèses de l'IRT

Ces modèles réfèrent à une famille de modèles mathématiques et trois hypothèses sous-jacentes doivent être respectées. La première implique la dimensionnalité de l'espace latent. L'IRT suppose que la probabilité de répondre positivement pour un individu découle de son positionnement sur les k traits latents (capacités, statuts) considérés. Géométriquement, sa position sur chacun de ces traits latents peut être conceptualisée comme un point dans l'espace à k -dimension. Cependant, pour beaucoup d'applications de modèle IRT, il est supposé que l'espace du trait latent est unidimensionnel. La première hypothèse est donc l'unidimensionnalité de la variable latente : l'ensemble des items considérés mesure une et une seule variable (capacité, connaissance, habilité, trait personnel). Pour la HRQoL, l'analyse par approche IRT s'effectue par dimension de HRQoL où pour chacune d'entre elles, une variable latente est considérée. Si l'analyse doit être réalisée sur un ensemble d'items caractérisant plusieurs dimensions de HRQoL, alors la variable latente sera caractéristique de l'ensemble des dimensions considérées.

La seconde hypothèse est l'indépendance locale. Conditionnellement à θ , les réponses aux différents items sont indépendantes. Pour un individu donné, les réponses aux différents items sont indépendantes les unes des autres telles que :

$$\Pr(Y_i = y_i | \theta_i, \delta_j) = \prod_{j=1}^J \Pr(Y_i^{(j)} = y_i^{(j)} | \theta_i, \delta_j),$$

où $Y_i = Y_{i,1:J}$ est le vecteur aléatoire réponse de l'individu i .

La troisième et dernière hypothèse est la monotonie : les probabilités de répondre positivement doivent être non décroissantes en θ , c'est-à-dire qu'elles vérifient :

$$\frac{d \Pr(Y_i^{(j)} = y_i^{(j)} | \theta_i, \delta_j)}{d\theta_i} \geq 0.$$

2.2.2 Modèles IRT pour données polytomiques

Les modèles pour données dichotomiques permettent une bonne introduction à l'approche IRT. Cependant, une grande majorité des données issues de questionnaires psychométriques sont catégorielles ordinales (polytomiques). Il existe plusieurs candidats pour la généralisation au cas polytomique des modèles IRT pour données binaires. Samejima (1969), Masters (1982), et Tutz (1990) discutent de modèles IRT ordinaux plus généraux, en proposant respectivement le *Graded Response Model (GRM)*, le *Partial Credit Model (PCM)* et le *Sequential Rasch Model (SRM)*.

Le *Graded Response Model*

Historiquement, le GRM est le premier modèle IRT proposé dans la littérature concernant les données polytomiques (Samejima, 1969). Conceptuellement, il est le modèle le plus simple à aborder concernant les différentes généralisations du modèle de Rasch. Il modélise la probabilité de répondre au moins à une certaine catégorie $m \in \{0, 1, \dots, M_j\}$ sachant la capacité θ_i du sujet i . En utilisant une paramétrisation similaire au modèle 2PL, le GRM est défini par :

$$\pi_{im}^{*(j)} = \Pr(Y_i^{(j)} \geq m | \theta_i) = \frac{\exp\{\alpha_j(\theta_i - \delta_{jm})\}}{1 + \exp\{\alpha_j(\theta_i - \delta_{jm})\}}, \quad (2.4)$$

où les M_j valeurs seuils $(\delta_{jm})_{m=1, \dots, M_j}$ (paramètres de difficulté) vérifient la contrainte $\delta_{j1} \leq \delta_{j2} \leq \dots \leq \delta_{jM_j}$. De manière identique au modèle 2PL, un paramètre de discrimination α_j est pris en compte pour chaque item j ($j = 1, \dots, J$). La probabilité qu'un individu i réponde la modalité m à l'item j est déduite de l'équation (2.4), tel que :

$$\pi_{im}^{(j)} = \Pr(Y_i^{(j)} = m | \theta_i) = \Pr(Y_i^{(j)} \geq m | \theta_i) - \Pr(Y_i^{(j)} \geq m + 1 | \theta_i). \quad (2.5)$$

La Figure 2.5 montre les relations entre les paramètres de difficulté et les différentes probabilités de réponse. Ils se déduisent à partir des fonctions des probabilités cumulées. La partie inférieure de la Figure 2.5 présente les différentes proportions des catégories observées. Nous remarquons la dépendance des différents paramètres de difficulté et la nécessité qu'ils soient croissants par ordre des catégories.

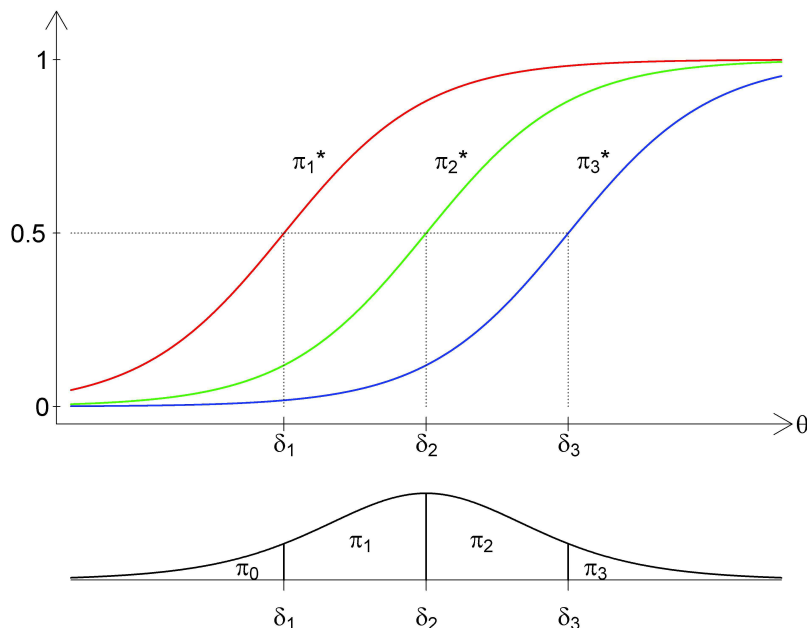


FIGURE 2.5: Liens entre les paramètres de difficulté et les fonctions de probabilité cumulées pour le GRM. Les paramètres de difficulté sont ordonnés avec $\delta_1 = -1$, $\delta_2 = 0$ et $\delta_3 = 1$.

Le *Partial Credit Model*

Bien que le PCM soit conceptuellement moins facile à appréhender que le GRM, ce modèle appartient à la famille des modèles Rasch. Il correspond à la généralisation du modèle (2.1) pour données polytomiques (Masters, 1982). Au lieu de mettre en opposition la probabilité de répondre positivement versus la probabilité de répondre négativement, le PCM met en opposition chaque paire de modalité adjacente :

$$\Pr\left(Y_i^{(j)} = m | \theta_i, Y_i^{(j)} \in \{m-1, m\}\right) = \frac{\exp(\theta_i - \delta_{jm})}{1 + \exp(\theta_i - \delta_{jm})}, \quad m = 1, \dots, M_j. \quad (2.6)$$

Il en advient l'écriture du modèle la plus utilisée :

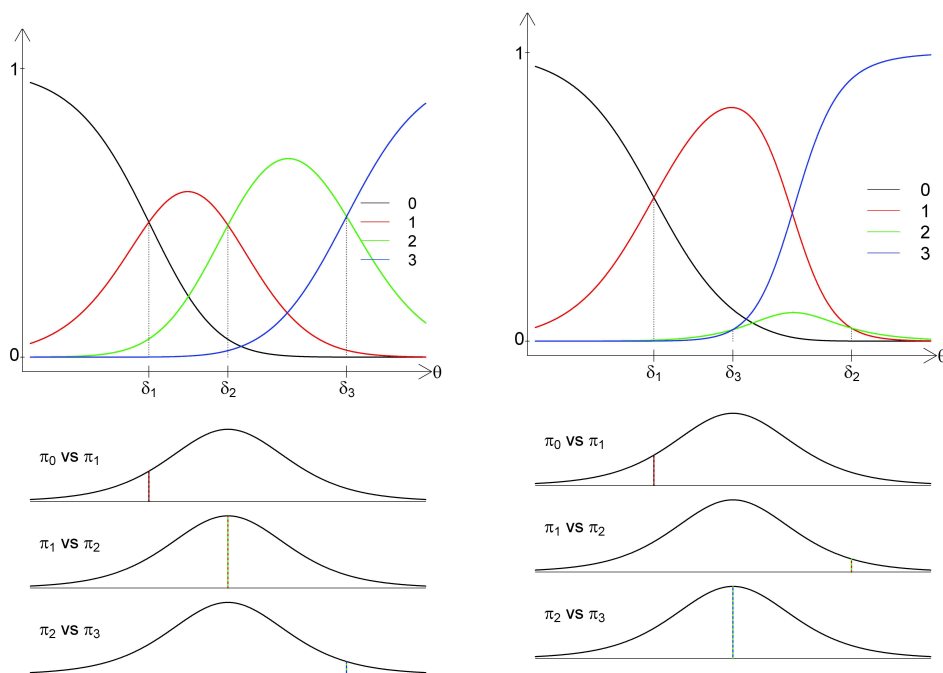
$$\Pr\left(Y_i^{(j)} = m | \theta_i\right) = \frac{\exp\left\{\sum_{k=1}^m (\theta_i - \delta_{jk})\right\}}{1 + \sum_{h=1}^{M_j} \exp\left\{\sum_{k=1}^h (\theta_i - \delta_{jk})\right\}}, \quad (2.7)$$

pour $m = 1, \dots, M_j$, et

$$\Pr\left(Y_i^{(j)} = 0 | \theta_i\right) = \frac{1}{1 + \sum_{h=1}^{M_j} \exp\left\{\sum_{k=1}^h (\theta_i - \delta_{jk})\right\}}.$$

Pour cette modélisation, il n'a pas de contrainte d'ordre sur les paramètres de difficulté contrairement au GRM. Le seuil δ_{jm} correspond à la valeur de la variable latente pour laquelle il y a la même probabilité de répondre la catégorie m que la

catégorie $m + 1$ à l'item j (voir Figure 2.6). De plus, le paramètre de discrimination est pris égal à 1 pour chaque item dans le modèle (2.7). Muraki (1992) propose de considérer ce paramètre de discrimination dans le *Generalized Partial Credit Model (GPCM)* au même titre que pour le GRM. Les parties inférieures de la Figure 2.6 permettent de visualiser le fait que le PCM modélise la probabilité de répondre une certaine catégorie par rapport à sa catégorie adjacente. Les aires sous les courbes représentent les proportions de réponse pour chaque paire considérée.



(a) Paramètres de difficulté ordonnés

(b) Paramètres de difficulté non ordonnés

FIGURE 2.6: Fonctions des catégories de réponse pour le PCM considérant un item à quatre catégories de réponse ($m = 0, \dots, 3$). 2.6a présente le cas où les paramètres de difficulté sont ordonnés avec $\delta_1 = -2$, $\delta_2 = 0$ et $\delta_3 = 3$. 2.6b présente le cas où les paramètres de difficulté ne sont pas ordonnés : $\delta_1 = -2$, $\delta_2 = 3$ et $\delta_3 = 0$.

Le Sequential Rasch Model

La dernière généralisation au cas polytomique est la moins connue et a été proposée par Tutz (1990). Cette modélisation est motivée par le fait qu'un processus séquentiel dirige la réponse à l'item. En effet, nous distinguons ici deux types conceptuels de données ordinales où l'une provient d'une échelle ordonnée (par exemple une échelle de Likert) et l'autre provenant d'un processus ordonné. Pour le dernier, les catégories sont atteintes successivement. Cela est le cas pour des données de l'accomplissement académique (Brevet, Baccalauréat, Licence, Master), ou encore les stades du développement de l'enfant selon Piaget (Agresti, 2010). L'exemple proposé

par Wright and Masters (1982) illustre bien ce second type de réponse. Considérons l’item : ” $\sqrt{\frac{9.0}{0.3}} - 5 = ?$ ”. Trois niveaux (catégories) de performance sont pris en compte où la réalisation d’un niveau particulier dépend de la réalisation du niveau précédent :

- Niveau 0 : Aucun problème sous-jacent n’est résolu ;
- Niveau 1 : $9.0/0.3 = 30$ est résolu ;
- Niveau 2 : $30 - 5 = 25$ est résolu ;
- Niveau 3 : $\sqrt{25} = 5$ est résolu.

La transition entre un niveau et le suivant est modélisée et correspond à la probabilité de répondre une catégorie supérieure à m sachant que l’étape m a déjà été réussie (Linden and Hambleton, 1997) :

$$\Pr \left(Y_i^{(j)} > m | Y_i^{(j)} \geq m, \theta_i \right) = \frac{\exp(\theta_i - \delta_{j,m+1})}{1 + \exp(\theta_i - \delta_{j,m+1})}, \quad m = 0, \dots, m_j - 1. \quad (2.8)$$

2.2.3 Variantes et paramétrisation

Le nombre de paramètres dans les modèles IRT peut rapidement être conséquent. Il dépend entre autre du nombre d’items considérés ainsi que du nombre de modalités de réponse de chacun de ces items, et bien entendu des contraintes supposées sur les paramètres d’item. Une première variante est déjà proposée dans les modèles (2.1) et (2.7) : contraindre le modèle à un paramètre de discrimination fixe entre les items ou même égale à 1. Le nombre de paramètres d’item peut encore être réduit sous certaines conditions. En effet, il existe plusieurs déclinaisons de chaque modèle. Pour un même "type" de modèle, les différents paramètres d’item peuvent être libres ou contraints (voir Table 2.2) (Bacci et al., 2012). Par exemple, dans le cas où tous les items sont constitués d’un même nombre de catégories de réponse, Andrich (1978) propose de simplifier la paramétrisation du modèle (2.7) en considérant que les paramètres de difficulté entre les différents items sont identiques plus ou moins un certain décalage τ_j propre à l’item j . Ce modèle est connu sous le nom du *Rating Scale Model (RSM)*, mais peut également se généraliser aux différents modèles présentés à condition que les items soient composés d’un même nombre de catégories de réponse.

TABLE 2.2: Récapitulatif des différentes paramétrisations IRT associées aux modèles IRT.

Format de réponse à l’item	Paramètres d’item		Paramétrisation	Modèle IRT
	Discrimination	Difficulté		
Dichotomique	contraints	libres	$\theta_i - \delta_j$	Modèle de Rasch
	libres	libres	$\alpha_j (\theta_i - \delta_j)$	Modèle 2PL
Polytomique	contraints	contraints	$\theta_i - \delta_m - \tau_j$	RSM
	contraints	libres	$\theta_i - \delta_{jm}$	PCM et SRM
	libres	contraints	$\alpha_j (\theta_i - \delta_m - \tau_j)$	Autres modèles
	libres	libres	$\alpha_j (\theta_i - \delta_{jm})$	GPCM et GRM

Considérant le modèle de Rasch (2.1), notons η_{ij} le prédicteur linéaire défini par

$\eta_{ij} = \theta_i - \delta_j$. Cette présentation de η_{ij} est la plus présentée dans l'IRT. Quatre paramétrisations de η_{ij} peuvent être considérées (Boeck and Wilson, 2004) :

1. $\theta_i - \delta_j$
2. $\theta_i + \delta_j^*$ où $\delta_j^* = -\delta_j$
3. $-\theta_i^* - \delta_j^*$ où $\theta_i^* = -\theta_i$
4. $\delta_j - \theta_i^*$

Selon les contextes, le choix d'une des quatre écritures améliore et facilite l'interprétation. En général, les deux paramétrisations utilisant la différence (1 et 4) se prêtent à des métaphores (situations) de comparaison et de concurrence (par exemple, la capacité et la difficulté). Cette notation prend tout son sens pour l'étude de la HR-QoL. En revanche, les deux formulations d'addition (2 et 3) sont adaptées pour la métaphore liée à l'intensification (Boeck and Wilson, 2004). Dans les modèles de régression pour données catégorielles, il est habituel de voir la seconde écriture.

Les modèles IRT classiques présentés dans cette section n'incluent pas de variables explicatives, autres que les paramètres liés aux items ainsi que la variable latente spécifique à l'individu (Liu and Hedeker, 2006). Ces modèles IRT sont présentés dans la littérature suivant la distribution logistique et pour l'analyse *doublement descriptive* (voir Table 2.3). Dans l'IRT, il y a une distinction entre l'approche descriptive et l'approche explicative. Suivant les intentions d'analyse, il est possible d'expliquer la différence entre les individus ou entre les items. Il suffit pour cela de décomposer linéairement la partie du modèle souhaitée (Boeck and Wilson, 2004) (voir Table 2.3). En oncologie, il est nécessaire de pouvoir évaluer la HRQoL de manière longitudinale et étudier l'influence de variables explicatives sur la partie individuelle. Par exemple, dans les essais cliniques utilisant la HRQoL comme critère, la différence de capacité suivant le traitement administré va être étudiée afin de s'assurer ou non d'un bénéfice clinique.

Concernant la paramétrisation liée à l'item, elle permet d'étudier le fonctionnement différentiel des items (*Differential Item Functioning*, DIF) (De Ayala, 2009) ou même le phénomène de RS (Guilleux et al., 2015). Le DIF correspond au fait que les personnes (ou groupe de personne) ne répondent pas de la même manière aux items. En général, ce phénomène est lié à des différences culturelles (exemple : hommes (H) et femmes (F)) et s'explique par une différence de « référence externe ». Les modèles IRT permettent de tester la présence de DIF (Hardouin et al., 2012). Ce changement peut être uniforme ou non. Un changement uniforme est représenté par un décalage moyen suivant une même manière de répondre ($\delta_H = (-1, 0, 1)$ et $\delta_F = (-0, 1, 2)$: un décalage de 1 en moyenne). Pour un même niveau du trait latent entre un homme et une femme, les réponses peuvent différer de par leur « référence interne » qui n'est pas la même. La RS est très proche de la notion de DIF mais la comparaison ne se fait pas entre les personnes mais sur une même personne au cours du temps. La RS est causée par une adaptation (à la maladie par exemple) ou bien un traumatisme (l'annonce de la maladie) : c'est la « référence interne » qui est modifiée par un événement. La présence de DIF ou de RS peut biaiser des conclusions. En effet, une différence de niveau de HRQoL entre deux groupes peut être détectée alors qu'elle est causée par un DIF. Cependant, ce biais est maîtrisé dans les essais

cliniques à l'aide d'une randomisation des patients qui provienne généralement d'une sous-population relativement homogène suivant des critères d'inclusion précis.

TABLE 2.3: Résumé des quatre analyses possibles. Z_{ip} matrice associée au régresseur sur l'individu. D_{jk} matrice associée au régresseur sur l'item. Les variables aléatoires sont définies telles que $\theta \sim \mathcal{N}(0, \sigma^2)$ et $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.

Type de Modèles	Prédicteur linéaire η_{ij}	
	Part individuelle	Part item
Double descriptif	θ_i	$-\delta_j$
Explicatif sur item	θ_i	$-\sum_{k=1}^K \delta_k D_{jk}$
Explicatif sur individu	$\sum_{p=1}^P \beta_p Z_{ip} + \varepsilon_i$	$-\delta_j$
Double explicatif	$\sum_{p=1}^P \beta_p Z_{ip} + \varepsilon_i$	$-\sum_{k=1}^K \delta_k D_{jk}$

La variable réponse multivariée est utilisée pour accéder à la variable latente qui est l'objet d'étude. En théorie, la variable latente θ représentant la capacité de l'individu, est un facteur. En effet, la variable réponse est multivariée (plusieurs réponses provenant de différents items) et la variable inobservée caractérise la dépendance de ces variables observables. En pratique, la variable réponse est considérée comme univariée où la distinction entre les différentes variables observées s'effectue à travers les paramètres d'items. Ces derniers sont associés aux différentes contraintes supposées : plus les items sont supposés différents, moins le modèle est contraint et plus le nombre de paramètres d'item est important. Dans ce cas de figure, les modèles IRT sont des GLMM pour données ordinales avec une paramétrisation spécifique du prédicteur linéaire, caractérisé par η . La variable latente θ est alors un effet aléatoire où sa réalisation/prédiction pour un individu quantifie le concept mesuré. Les modèles considérés dans la suite sont les modèles IRT étendus au cas *explicatif sur individu* vus dans la Table 2.3. Dans la littérature, les modèles IRT sont étendus à l'analyse longitudinale en décomposant linéairement la variable latente de manière identique à la décomposition du score dans un LMM (Huber et al., 2013) :

$$\begin{cases} \eta &= \theta - \delta \\ \theta &= X\beta + U\xi \end{cases} \quad (2.9)$$

Pour mieux comprendre ces modèles et les démocratiser, des études de comparaison ont été réalisées. Comme toute étude comparative, il est nécessaire d'avoir une référence. Le modèle de référence est basé sur l'étude du score de HRQoL et est généralement le LMM malgré le fait qu'il ne soit pas adapté à ce type de réponse. Ces études ont convergé vers la conclusion que les modèles IRT sont plus adaptés pour l'analyse longitudinale que les LMM (Blanchin et al., 2011; Bock et al., 2013; Gorter et al., 2015). Récemment, Anota et al. (2014a) ont comparé le LMM, la méthode du TJD et un modèle IRT pour données polytomiques (PCM étendu à l'analyse longitudinale, LPCM) lors d'une étude exhaustive de simulation avec plusieurs scénarios visant à détecter une différence de HRQoL entre deux bras de traitement (contexte

d'un essai clinique). Les scénarios se distinguent de par le nombre de visites considérées (5 ou 10), le nombre d'items par dimension (1, 2 ou 4 items à 4 ou 7 modalités de réponse), différents effets entre les deux bras de traitement ou encore de par le nombre et type de données manquantes (MNAR intermittentes ou monotones). Ces simulations ont montré une capacité équivalente entre le LPCM et le LMM à détecter une évolution différente de la HRQoL entre les deux groupes sur données complètes avec 5 temps de mesure. Lorsque 10 temps de mesure ont été considérés, la puissance du modèle IRT était plus faible que celle du LMM. En présence de données manquantes informatives, les puissances de chaque modèle diminuent à l'exception de celle du modèle IRT qui s'avère plus robuste. D'après cette étude, le LPCM semble le modèle le plus adapté pour analyser ce critère, et en particulier pour une dimension uni-item. Concernant la méthode du TJD, elle s'avère la moins puissante et la moins adaptée pour les dimensions à un seul item.

Dans la section suivante, nous proposons une application visant à présenter l'utilisation de modèles IRT pour l'analyse longitudinale de la HRQoL dans un essai clinique.

2.3 Application

Dans cette section, nous présentons les résultats issus de l'approche IRT (et CTT avec le LMM) pour l'analyse de la HRQoL des données de l'essai clinique de phase III Prodigé 4/ACCORD 11. Il a été conçu afin de comparer le Folfirinox, un protocole associant quatre médicaments (oxiplatine, irinotecan, fluorouracile et leucovorine), à la gemcitabine dans le traitement du cancer métastatique du pancréas. Cet essai a recruté 342 patients dans 48 établissements de soins français (centres anticancéreux, hôpitaux publics universitaires et régionaux, cliniques privées) et a permis de mettre en évidence une réponse thérapeutique plus importante, une stabilisation accrue de la maladie et une durée de survie médiane des patients traités de 11.1 mois avec le Folfirinox contre 6.8 avec la Gemcitabine. Ces résultats ont induit une modification des pratiques et donc du standard de traitement dans une pathologie très grave (Conroy et al., 2011). Cependant, le Folfirinox se caractérise par une toxicité plus importante que la Gemcitabine et son utilisation doit être limitée aux patients présentant un bon état général. Une analyse approfondie de la HRQoL des patients dans cet essai a été donc réalisée afin de vérifier la tolérance du traitement par Folfirinox.

Les données de HRQoL sont issues du questionnaire EORTC QLQ-C30 version 3.0 et ont été recueillies sur une période de 10 mois à huit visites ($v = 1, \dots, 8$) : 0, 15 jours, 1, 2, 4, 6, 8 et 10 mois. Gourgou-Bourgade et al. (2012) ont étudié et comparé les scores de HRQoL entre les deux bras de traitement avec la méthode des TJDD. Malgré une toxicité plus élevée du Folfirinox, la perception de la HRQoL des patients est meilleure au cours du temps pour ce bras de traitement que pour celui ayant reçu la Gemcitabine. En effet, une majorité de dimension de HRQoL montre qu'il y a un plus grand nombre de détériorations du score (DMCI de 10 ou 20 points, décès compris ou non) pour les patients ayant reçu la Gemcitabine.

Dans cette section, nous proposons d'effectuer l'analyse de données de HRQoL

en utilisant des modèles IRT. Nous avons considéré les modélisations longitudinales pour données polytomiques associés au PCM et au GRM, respectivement nommés LPCM (*longitudinal partial credit model*) et LGRM (*longitudinal graded response model*). Pour l'analyse longitudinale, la paramétrisation du prédicteur linéaire des deux modèles est :

$$\begin{cases} \eta_{ivm}^{(j)} &= \theta_{iv} - \delta_{jm} \\ \theta_{iv} &= \beta_0 + \xi_{0i} + \beta_1 g_i + t_{iv} (\beta_2 + \beta_3 g_i + \xi_{1i}) \end{cases}, \quad (2.10)$$

avec :

- $m = 1, \dots, M_j$ la modalité de réponse de l'individu i ($i = 1, \dots, n$) à l'item j ($j = 1, \dots, J$);
- δ_{jm} le paramètre de difficulté de l'item j lié aux modalités $m - 1$ et m , avec $\delta_{j1} = 0$ par convention;
- paramètres de discrimination contraint à $\alpha_j = 1, \forall j = 1, \dots, J$;
- t_v est le temps en jour(s) entre la visite d'inclusion ($v = 1$) et la visite v ;
- β_0 est l'intercept lié au groupe ayant reçu la Gemcitabine $g_i = 0$;
- $\beta_0 + \beta_1$ est l'intercept lié au groupe ayant reçu le Folfirinox $g_i = 1$ (β_1 représente la différence entre les deux groupes à la baseline);
- β_2 pente du groupe $g_i = 0$;
- $\beta_2 + \beta_3$ pente du groupe $g_i = 1$;
- $\xi_i = (\xi_{0i}, \xi_{1i})'$ est le vecteur des effets aléatoires d'espérance nulle et de matrice de covariance Σ .

Les résultats de l'approche par LMM sont également présentés à titre comparatif. Le modèle (1.7) est considéré pour cette analyse longitudinale. La même paramétrisation des prédicteurs linéaires pour le LMM et pour les modèles IRT rend comparable les conclusions et interprétations aux vues des estimations obtenues. Cependant, les valeurs ne sont pas comparables entre les paramètres des différentes modélisations.

2.3.1 Résultats et estimations des modèles

L'article relatif à cette section (voir section 2.5) présente les résultats de l'analyse de HRQoL par le LMM et le LPCM. Des représentations graphiques des scores de HRQoL pour chacune des dimensions, mais également celles des résultats du LPCM sont proposées dans cet article. La Table 2.4 retourne les estimations des paramètres $(\beta_p)_{p=2,3}$ qui quantifient l'évolution de la HRQoL pour chacune des dimensions du questionnaire suivant les deux groupes de patients. Elle correspond aux estimations présentées dans l'article en section 2.5, auxquelles les estimations du LGRM ont été rajoutées. Le paramètre β_1 a été considéré seulement pour la dimension liée au symptôme de diarrhée. En effet, pour toutes les autres dimensions de HRQoL analysées, ce paramètre n'est pas sorti significativement différent de 0 suite au test de Wald. Ce résultat était attendu étant donné que l'essai était randomisé et que les patients n'avaient pas encore reçu le traitement à t_0 .

2.3. Application

TABLE 2.4: Estimations des effets fixes (β_p)_{p=1,2,3} des modèles LMM, LPCM et LGRM pour chaque dimension du QLQ-C30.

	LMM			LPCM			LGRM		
	Coef.	(sd)	p.	Coef.	(sd)	p.	Coef.	(sd)	p.
Statut Global de Santé									
β_2	0.0370	(0.0121)	0.002	0.0022	(0.0015)	0.147	0.0032	(0.0023)	0.166
β_3	0.0236	(0.0138)	0.088	0.0030	(0.0018)	0.102	0.0043	(0.0028)	0.128
Fonction Physique									
β_2	0.0118	(0.0135)	0.382	0.0036	(0.0019)	0.063	0.0049	(0.0025)	0.0512
β_3	-0.0002	(0.0161)	0.990	-0.0036	(0.0025)	0.152	-0.0040	(0.0032)	0.212
Fonction Personnelle									
β_2	0.0301	(0.0198)	0.130	0.0001	(0.0021)	0.944	0.0004	(0.0027)	0.892
β_3	0.0159	(0.0233)	0.495	-0.0041	(0.0027)	0.132	-0.0051	(0.0034)	0.131
Fonction Émotionnelle									
β_2	0.0690	(0.0149)	<.001	-0.0098	0.0019	<.001	-0.0110	(0.0023)	<.001
β_3	-0.0004	(0.0176)	0.983	-0.00004	(0.0023)	0.985	-0.0003	(0.0028)	0.992
Fonction Cognitive									
β_2	0.0007	(0.0128)	0.958	0.0001	(0.0015)	0.918	0.0001	(0.0018)	0.9717
β_3	0.0166	(0.0152)	0.276	-0.0031	(0.0018)	0.096	-0.0029	(0.0022)	0.1887
Fonction Sociale									
β_2	0.0131	(0.0176)	0.458	-0.0005	(0.0020)	0.799	-0.0003	(0.0024)	0.8884
β_3	0.0148	(0.0210)	0.481	-0.0031	(0.0025)	0.228	-0.0038	(0.0030)	0.2106
Fatigue									
β_2	-0.0502	(0.0183)	0.006	-0.0025	(0.0023)	0.283	-0.0028	(0.002789145)	0.308
β_3	0.0038	(0.0217)	0.860	-0.0009	(0.0030)	0.753	-0.0011	(0.0035)	0.761
Nausée et vomissements									
β_2	-0.0297	(0.0143)	0.039	-0.0015	(0.0015)	0.337	-0.0017	(0.0020)	0.393
β_3	-0.0046	(0.0159)	0.774	-0.0020	(0.0018)	0.273	-0.0023	(0.0023)	0.336
Douleur									
β_2	-0.0826	(0.0181)	<.001	-0.0089	(0.0021)	<.001	-0.0110	(0.0025)	<.001
β_3	-0.0364	(0.0201)	0.071	-0.0053	(0.0025)	0.033	-0.0062	(0.0030)	0.042
Dyspnée									
β_2	-0.0129	(0.0164)	0.431	-0.0017	(0.0021)	0.433	-0.0020	(0.0024)	0.420
β_3	-0.0235	(0.0191)	0.218	-0.0032	(0.0025)	0.207	-0.0030	(0.0029)	0.295
Insomnie									
β_2	-0.1103	(0.0193)	<.001	-0.0103	(0.0023)	<.001	-0.0118	(0.0026)	<.001
β_3	0.0081	(0.0214)	0.706	0.0011	(0.0023)	0.643	0.0013	(0.0027)	0.627
Perte d'appétit									
β_2	-0.1289	(0.0215)	<.001	-0.0080	(0.0017)	<.001	-0.0118	(0.0024)	<.001
β_3	-0.0149	(0.0241)	0.537	-0.0011	(0.0019)	0.569	-0.0009	(0.0026)	0.747
Constipation									
β_2	-0.1124	(0.0190)	<.001	-0.0084	(0.0019)	<.001	-0.0107	(0.0025)	<.001
β_3	0.0107	(0.0204)	0.599	-0.0001	(0.0020)	0.977	0.0001	(0.0027)	0.974
Diarrhée									
β_1	9.0733	(2.5793)	<.001	0.5167	(0.1749)	0.003	0.7395	(0.2717)	0.007
β_2	0.0143	(0.0185)	0.439	0.0001	(0.0018)	0.965	0.0006	(0.0022)	0.792
β_3	-0.0273	(0.0228)	0.232	-0.0003	(0.0018)	0.852	-0.0007	(0.0025)	0.786
Difficultés Financières									
β_2	-0.0045	(0.0125)	0.717	-0.0115	(0.0074)	0.124	-0.0172	(0.0093)	0.066
β_3	0.0261	(0.0155)	0.093	0.0093	(0.0055)	0.091	0.0099	(0.0068)	0.146

Concernant la dimensions de diarrhée, les différents modèles mixtes (IRT et LMM) semblent en accord sur le fait qu'il existe une différence significative entre les groupes de traitement à la visite d'inclusion. Cependant, en regardant de plus près les données liées à cette dimension, nous nous rendons compte qu'à la visite à l'inclusion le score moyen est similaire entre les deux groupes puis augmente de manière distincte pour le groupe Folfirinox à la seconde, troisième et quatrième visite (pendant les deux premiers mois du traitement) pour redescendre au niveau d'origine.

La Table 2.4 montre des résultats similaires entre les 3 modèles. En effet, pour la majorité des dimensions de HRQoL étudiées, la significativité des effets fixes est identique quel que soit le modèle utilisé. Quelques dimensions présentent cependant des différences (statut global de santé, fatigue, nausée et vomissements et douleur) et ces différences interviennent notamment entre le LMM et les deux modèles IRT. De plus, pour les dimensions fonctionnelles, les signes des coefficients significatifs sont opposés entre le LMM et les modèles IRT, ceci étant la conséquence de l'inversion de l'échelle des modalités de réponse dans le calcul des scores fonctionnels. Concernant les deux modèles IRT, aucune différence d'interprétation n'est remarquée.

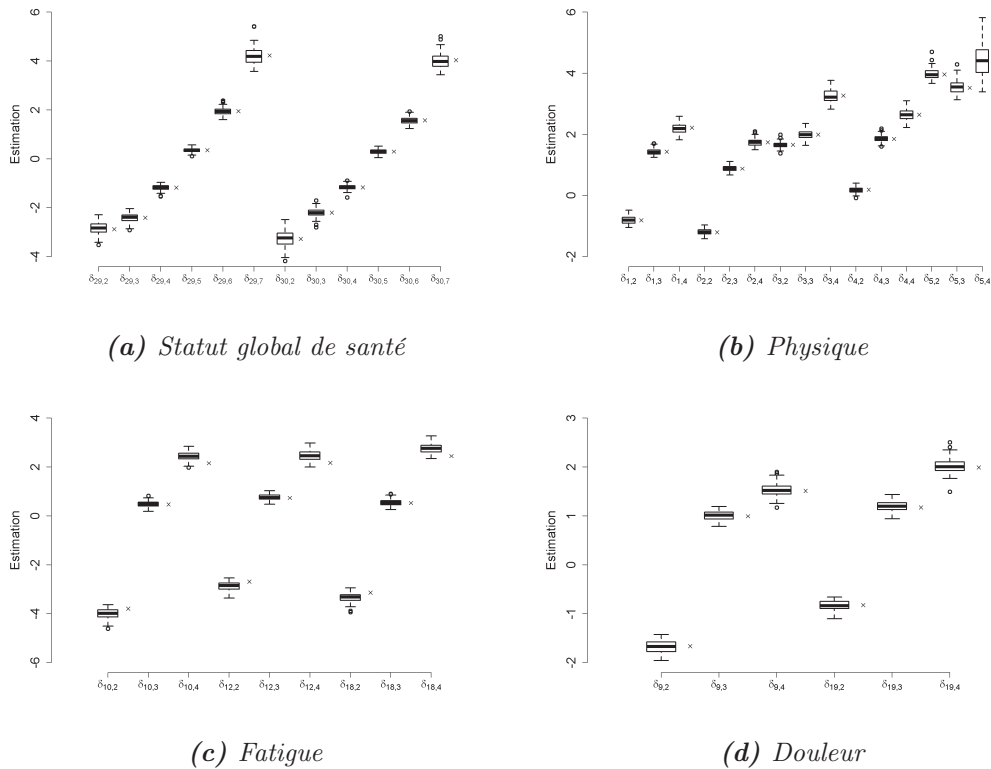


FIGURE 2.7: Distributions des paramètres de difficultés estimés ($N=100$ estimations) dans l'étude de validation croisée. La croix située à droite du Boxplot représente l'estimation obtenue par le modèle lorsque toutes les observations ont été prises en compte.

2.3.2 Sensibilité des estimations du modèle LPCM suivant les observations

L'utilisation de ces modèles IRT pour l'analyse de la HRQoL dans un essai clinique en cancérologie est une nouveauté. Afin de s'assurer de la robustesse des modèles IRT (validation interne), nous avons étudié la sensibilité des estimations du LPCM (voir article en section 2.5) au moyen d'une validation croisée. Pour cela, les paramètres du LPCM des dimensions GHS, fonction physique, fatigue et douleur ont été estimés (cent fois) sur deux tiers de l'échantillon pris aléatoirement.

L'estimation des paramètres du LPCM est stable suivant les différents sous-échantillons considérés. La Figure 2.7 montre que les estimations des paramètres de difficulté sont similaires à l'estimation obtenue sur la totalité de l'échantillon. Ce constat est le même pour les estimations des effets fixes $(\beta_p)_{p=2,3}$ (voir Figure 2.8). De plus, les conclusions concernant l'évolution de la HRQoL au cours du temps (significativité des paramètres β_2 et β_3) sont sensiblement les mêmes.

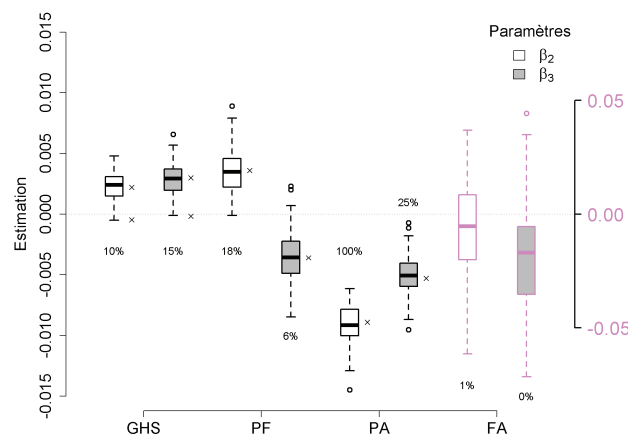


FIGURE 2.8: Distributions des paramètres β_2 et β_3 estimés ($N=100$ estimations) dans l'étude de validation croisée. La croix située à droite du Boxplot représente l'estimation obtenue par le modèle lorsque toutes les observations ont été prises en compte. Les pourcentages représentent le nombre de fois (sur $N=100$) où les paramètres sont significatifs.

2.4 Discussion

Les modèles IRT présentent une méthodologie appropriée pour l'analyse des données de HRQoL. La paramétrisation du prédicteur linéaire permet un ajustement plus ou moins flexible suivant les contraintes considérées. Cependant, les conclusions qui découlent de ces modèles sont sensiblement identiques à celles du LMM. Ces résultats confortent les conclusions obtenues par (Anota et al., 2014a) où les puissances du LMM et du LPCM étaient équivalentes pour détecter une différence entre deux groupes.

En revanche, ces résultats sont différents de ceux obtenus avec les modèles TJD. Pour l'essai Prodigé 4/ACCORD 11, les modèles IRT et LMM montrent une évolution de la HRQoL similaire entre les deux groupes de traitement mis à part pour la dimension de douleur. L'approche par TJD démontre une meilleure HRQoL (moins de détériorations du score au cours du temps) dans le bras Folfirinox. Les modèles mixtes (LMM et IRT) et les modèles TJD sont finalement des modèles pertinents pour l'analyse de la HRQoL dans les essais cliniques car ils mettent en évidence des résultats différents mais complémentaires. Les modèles de TJD (ou TJDD) permettent de démontrer les différences en termes de détérioration du score HRQoL survenue au cours du temps. Les modèles mixtes comparent quant à eux les évolutions du niveau HRQoL entre les groupes.

Dans l'IRT, plusieurs modèles sont envisageables pour l'analyse des données polytomiques. Nous avons choisi d'appliquer le modèle PCM et GRM sur les données de l'essai clinique Prodigé 4/ACCORD 11 car leur utilisation est la plus fréquente (Hardouin et al., 2012, 2015; Liu and Hedeker, 2006). En effet, la modélisation proposée par (Tutz, 1990) est peu utilisée en pratique. Dans la littérature, peu d'arguments sont finalement mis en avant quant à l'utilisation d'un modèle IRT plutôt qu'un autre. Certains auteurs utilisant le modèle GRM mettent en avant l'argument que c'est le modèle le plus intuitif et le plus ancien. D'autres auteurs favorisent le PCM car il appartient aux modèles de la famille de Rasch.

Les modèles IRT sont des GLMM pour données ordinales avec une paramétrisation spécifique du prédicteur linéaire. Cette paramétrisation particulière permet de prendre en compte les différentes sources (items) dont proviennent les réponses. Les GLMM pour données catégorielles fait l'objet de nombreuses recherches qui ont permis de dégager des propriétés ainsi que de mettre en évidence leurs équivalences et leurs différences. Dans le chapitre suivant, nous remplaçons les modèles IRT dans le contexte des GLMM pour données ordinales dans le but de choisir quel modèle IRT est le plus approprié pour l'analyse de la HRQoL en cancérologie.

2.5 Article : *Applying the longitudinal model from item response theory to assess the health-related quality of life in the PRODIGE 4/ACCORD11 randomized trial*

Applying the Longitudinal Model from Item Response Theory to Assess the Health-Related Quality of Life in the PRODIGE 4/ACCORD 11 Randomized Trial

Antoine Barbieri, MSc, Amélie Anota, PhD, Thierry Conroy, MD, PhD,
Sophie Gourgou-Bourgade, MSc, Beata Juzyna, MSc, Franck Bonnetain, PhD,
Christian Lavergne, PhD, Caroline Bascoul-Mollevis, PhD

Introduction. A new longitudinal statistical approach was compared to the classical methods currently used to analyze health-related quality-of-life (HRQoL) data. The comparison was made using data in patients with metastatic pancreatic cancer. **Methods.** Three hundred forty-two patients from the PRODIGE4/ACCORD 11 study were randomly assigned to FOLFIRINOX versus gemcitabine regimens. HRQoL was evaluated using the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30. The classical analysis uses a linear mixed model (LMM), considering an HRQoL score as a good representation of the true value of the HRQoL, following EORTC recommendations. In contrast, built on the item response theory (IRT), our approach considered HRQoL as a latent variable directly estimated from the raw data. For polytomous items, we extended the partial credit model to a longitudinal analysis (longitudinal partial credit model [LPCM]), thereby modeling the latent trait

as a function of time and other covariates. **Results.** Both models gave the same conclusions on 11 of 15 HRQoL dimensions. HRQoL evolution was similar between the 2 treatment arms, except for the symptoms of pain. Indeed, regarding the LPCM, pain perception was significantly less important in the FOLFIRINOX arm than in the gemcitabine arm. For most of the scales, HRQoL changes over time, and no difference was found between treatments in terms of HRQoL. **Discussion.** The use of LMM to study the HRQoL score does not seem appropriate. It is an easy-to-use model, but the basic statistical assumptions do not check. Our IRT model may be more complex but shows the same qualities and gives similar results. It has the additional advantage of being more precise and suitable because of its direct use of raw data. **Key words:** health-related quality of life; EORTC QLQ-C30; longitudinal analysis; generalized linear mixed model; item response theory. (*Med Decis Making XXXX;XX:xx-xx*)

Health-related quality of life (HRQoL) is becoming an increasingly important issue in oncology clinical trials. A major feature of this concept is that the HRQoL is not directly observable but rather shaped by underlying notions. Indeed, the HRQoL is multidimensional, incorporating at least physical, psychological, and social domains as well as symptoms because of treatments and disease progression. Statistical research regarding HRQoL continues to be necessary to improve patients' care and to better

evaluate the impact of treatments on their everyday life.¹⁻³ HRQoL is a subjective and dynamic concept, and its longitudinal analysis remains nonhomogeneous across clinical trials, which makes it difficult to compare results. An appropriated methodology for patient-reported outcomes with clinically meaningful results remains mandatory for clinicians. Indeed, the evaluation of HRQoL as secondary endpoint in cancer clinical trials is now recommended by US administration guidelines,⁴ while it could be also used as primary since it is a clinical endpoint. Assessment of HRQoL is a key issue to evaluate the impact of cancer and its treatments on patients, in particular those in a palliative situation.⁵

The open-label multicenter randomized phase II/III clinical trial PRODIGE (Partenariat de Recherche

© The Author(s) 2015
Reprints and permission:
<http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/0272989X15621883

en Oncologie Digestive) 4/ACCORD (Actions Concertées dans les Cancers Colo-Rectaux et Digestifs) 11 showed the benefits provided by the FOLFIRINOX protocol over the former standard therapy, gemcitabine, as first-line treatment in metastatic pancreatic cancer.⁶ Overall survival was found to be significantly improved in the arm receiving FOLFIRINOX without the patients' HRQoL being affected.

We have previously reported an in-depth HRQoL analysis.⁷ Contrary to other studies in oncology, the HRQoL assessment did not rely on linear mixed modeling^{8,9}: our analysis used instead a time-to-event approach^{10–12} based on time until definitive deterioration (TUDD) for a 10- and 20-point minimal clinically important difference.¹³ In current studies, time to deterioration (TTD) and TUDD approaches are increasingly used for the longitudinal analysis of HRQoL in oncology clinical trials.^{10,14} These statistical methods have the advantage of being meaningful and easily understandable for clinicians. The use of scoring procedures is common practice because statistical methods for quantitative variables are more powerful and easier to implement and interpret than those conventionally applied to qualitative variables.¹⁵ With ordinal data from a Likert scale, an integer is associated with each category using a scoring procedure (e.g., 1 = *not at all*, 2 = *a little*, 3 = *quite a bit*, 4 = *very much*). However, the gap separating each adjacent response category is not necessarily the same, and it is unknown. The scoring procedure is based on an arbitrary assumption where these gaps are equivalent, unlike in item response theory (IRT) models. This assumption can cause biased results even if it has been shown through simulation studies that the

minimum number to have an acceptable bias with quantitative methods is 5 categories.¹⁵ In particular for HRQoL dimensions with 1 item with 4 categories of responses where the HRQoL score has the same features as the ordinal variable (always 4 response values expressed on a different scale), using the LMM and time-to-event models, which are based on the score study, does not seem suitable. Moreover, Anota and others² showed that the time-to-event models were less powerful (statistical power associated with the type II error rate) than the regression mixed models in highlighting a significant difference between the 2 treatment arms over time. The purpose of the present study was to propose an alternative longitudinal statistical approach based on IRT and mixed-effects regression models.^{16–18} This strategy of longitudinal analysis for HRQoL data is illustrated in the PRODIGE 4/ACCORD 11 trial. We compared results from usual linear mixed models (LMMs) for repeated measures used to analyze HRQoL data to results obtained from the proposed model. The 2 methodologies are then discussed, pointing out their drawbacks and advantages.

PATIENTS AND METHODS

Trial Design

Three hundred forty-two patients with measurable metastatic pancreatic adenocarcinomas were randomly assigned to FOLFIRINOX (oxaliplatin, 85 mg/m²; irinotecan, 180 mg/m²; leucovorin, 400 mg/m²; fluorouracil, 400 mg/m² by intravenous bolus followed by 2400 mg/m² by continuous intravenous infusion) versus gemcitabine (1000 mg/m² weekly for 7 of 8 wk and weekly for 3 of 4 wk) regimens. Demographic and baseline disease characteristics were similar in the FOLFIRINOX and gemcitabine groups. Median follow-up was 26.6 mo (95% confidence interval [CI], 20.5 to 44.9). Median overall survival was 11.1 mo (95% CI, 9.0 to 13.1) and 6.8 mo (95% CI, 5.5 to 7.6), respectively, in the FOLFIRINOX and gemcitabine groups ($P < 0.001$). Significant clinical and demographic variables for overall survival were treatment arm, age >65 y, low serum albumin, synchronous metastases, and liver metastases. Detailed inclusion and exclusion criteria, study design, protocol, treatment, HRQoL measurements, and compliance are documented elsewhere.^{6,7}

HRQoL Assessment

HRQoL was assessed with the European Organization for Research and Treatment of Cancer (EORTC)

Received 6 January 2015 from Biometrics Unit-CTD INCa, Institut régional du Cancer Montpellier (ICM) - Val d'Aurelle, Montpellier, France (AB, SG-B, CB-M); Institut Montpelliérain Alexander Grothendieck, University of Montpellier, Montpellier, France (AB, CL); National Quality of Life in Oncology Platform, France (AA, TC, FB); Methodological and Quality of Life Unit in Oncology (EA 3181), Besançon, France (AA, FB); Institut de Cancérologie de Lorraine, Nancy, France (TC); Unicancer R&D, Paris, France (BJ); and Université Paul-Valéry Montpellier 3, Montpellier, France (CL). This study was supported by a grant from the French Public Health Research Institute (<http://www.iresp.net>) under the 2012 call for projects as part of the 2009–2013 Cancer Plan. Revision accepted for publication 16 November 2015.

Supplementary material for this article is available on the *Medical Decision Making* Web site at <http://mdm.sagepub.com/supplemental>.

Address correspondence to Antoine Barbieri, ICM, Biometrics Unit-CTD INCa, 208 Avenue des Apothicaires, 34298 Montpellier, France; telephone: +33467612450; e-mail: Antoine.Barbieri@icm.unicancer.fr.

quality of life questionnaire core (QLQ-C30 version 3.0).¹⁹ This questionnaire contains 30 items summarized into 15 scales or dimensions: 5 functional scales, 9 symptomatic scales, and the global health status/QoL scale (GHS/QoL). These dimensions have been validated as follows: 5 items regarding physical functioning; 4 items assessing emotional functioning; 3 items about fatigue symptoms; 2 items for role, cognitive and social functioning, nausea and vomiting, pain and global health status; and finally 1 item for the following symptoms: dyspnea, insomnia, appetite loss, diarrhea, constipation, and financial difficulties. The items 1 to 28 are coded into 4 ordered categories according to the Likert scale: from 1 to 4, *not at all, a little, quite a bit, and very much*; the items 29 and 30 corresponding to GHS/QoL are scored from 1 (*very poor*) to 7 (*excellent*). According to the EORTC recommendations,^{20,21} the HRQoL analysis was based on a separate statistical analysis for each HRQoL dimension. For this purpose, as done elsewhere, a score was associated with each scale through item responses. Each score was calculated as the average of the responses to these items and expressed on a scale of 0 to 100. High functional and GHS/QoL scores reflected good functional capacities and a good level of HRQoL. Conversely, high symptomatic scores represented strong symptoms and pointed out difficulties (fatigue, nausea and vomiting, etc.).

The questionnaire was filled out by the patients themselves and collected at different follow-up times defined in the protocol: baseline, day 15, day 30, and months 2, 4, 6, 8, and 10. HRQoL compliance was detailed in a previous paper.⁷ The different measuring times reflect the uneven longitudinal aspect of the analysis of HRQoL. However, these repeated measurements over time introduced a correlation within the results measured for each patient. The raw data structure was described by 3 characteristics: ordinal categorical data, multiple-item responses, and repeated measures.

Classical Test Theory and IRT Approaches

Two approaches are often used to analyze data from questionnaires: the classical test theory (CTT) and the item response theory (IRT).^{22,23} HRQoL is generally considered as a latent variable (unobservable continuous variable), that is, a hidden variable not directly measurable. In this study, it was considered as a random variable associated with the combination of fixed and random effects. The difference between the 2 theories lies in the relationship they postulate between item responses and the latent trait.

Also called the *true score theory*,²² CTT is the most commonly used method in HRQoL studies. The HRQoL concept is quantified by a score directly linked to the different item categories chosen, which is meant to be close to the “real” HRQoL level. For longitudinal analysis, the approach commonly used to study HRQoL is the CTT through an LMM.²³

The IRT refers to a set of probabilistic models that link the item responses reported (raw data) with the latent variable (HRQoL concept). These models take into account a general individual parameter called the “capacity” (or ability) parameter and item parameters named “difficulty” parameters. The capacity parameter represents the HRQoL level of the patient, and the difficulty parameters characterize the different thresholds associated with the different item response categories. IRT models rely on 3 main assumptions: the capacity parameter is unique for all items associated with a HRQoL scale (unidimensionality); conditionally to the HRQoL level of the patient, the item responses are independent; and an increase of the HRQoL level induces a higher probability to produce an answer corresponding to a high category response for a patient. IRT models are widely used in the process of development and validation of questionnaires^{18,24} but are not much exploited for the longitudinal analysis of HRQoL in clinical trials.²⁵

Statistical Analysis

Mixed models are powerful and flexible tools for the analysis of longitudinal data and allow for the consideration of the different characteristics structuring the data.^{26,27} The HRQoL data collected can be structured at 3 important levels: the patients, the measuring times, and the items for each dimension. From a general point of view, these models of regression allow the assessment of different factors that influence a common behavior for every individual (fixed effects), such as time or treatment. Also, there is an individual factor due to repeated measurements. However, the variability of the data for each patient was taken into account. The individual effect was thus considered through the random effect. These models enabled the separation of the variability due to an error of measurement and to random effects in order to take into account the specific correlation between the different observations for each patient. Thus, an individual random effect and an individual time random effect can be fitted, assuming for each patient a specific HRQoL level at baseline and a specific evolution of HRQoL close to the general trend

captured by the fixed part of model. These models are often used, but their complexity can be seen as a barrier for clinicians.

LMM describes the change of HRQoL, for a given dimension, over time, combining fixed effects (at a global level) and random effects (at the individual level). It considers the standardized score (S) built from item responses as a true representation of the patients' HRQoL. Focusing the analysis on the score makes it possible to reduce the dimensionality of the data arising from the multiple HRQoL dimensions over several items. The model for the longitudinal analysis of HRQoL score used in our study was the following:

$$S_{it} = \alpha + u_{i0}^{CTT} + x_i \beta_1^{CTT} + [\beta_2^{CTT} + x_i \beta_3^{CTT} + u_{it}^{CTT}]f(t) + \varepsilon_{it} \tag{1}$$

where S_{it} denotes the HRQoL score for the i^{th} subject at time t and $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ represents the error term. Here, x_i denotes an indicator variable for treatment arm: $x_i = 1$ if patient i belongs to the experimental group and $x_i = 0$ if patient i belongs to the standard group. The time function $f(t)$ was defined as the time (in days) from the baseline visit to visit time t . We sought to assess the influence of β parameters: the group effect at baseline ($t = 0$) was represented by β_1^{CTT} and the HRQoL perception had a linear trend over time with slope β_2^{CTT} for the standard group and slope $\beta_2^{CTT} + \beta_3^{CTT}$ for the experiment group.²⁸ The random effects u_{i0}^{CTT} and u_{it}^{CTT} were normally distributed with zero mean and unstructured covariance matrix V^{CTT} .

The second model is the partial credit model (PCM)¹⁷: a generalized LMM commonly used in IRT. The PCM was extended in order to analyze the HRQoL change over time considering variables of influence (e.g., fixed effect of time) and taking into account the serial dependence of data from a single patient through some random effects.²⁹ Unlike LMM, PCM takes into account all the characteristics of the raw data, such as their ordinal categorical structure and their multiple responses. The PCM is a popular IRT model used for polytomous data (item modalities ordered from 1 to m_j). This model belongs to the Rasch-family models,³⁰ but it can be used only for transversal analysis. The main feature of this model lies in its ability to use the raw data directly without transforming the qualitative response variable into a pseudo-continuous scale, according to the EORTC QLQ-C30 scoring manual.²¹ IRT assumes

that the HRQoL is not directly measurable, but responses to items and a latent trait (θ) are linked. Indeed, a (not necessarily linear) link function relates the items responses ($Y_{it}^{(j)}$) with the HRQoL through an individual parameter (person ability θ_{it}) and through item parameters (item difficulty δ). The longitudinal partial credit model (LPCM) is an adjacent-categories logit model for ordered data²³ and considers each adjacent pair of response categories. Its general expression is

$$P_{jc}(\theta_{it}) = \Pr(Y_{it}^{(j)} = c | \theta_{it}, \delta) = \frac{\exp\{\sum_{k=1}^c (\theta_{it} - \delta_{jk})\}}{\sum_{h=1}^{m_j} \exp\{\sum_{k=1}^h (\theta_{it} - \delta_{jk})\}} \tag{2}$$

where $P_{jc}(\theta_{it})$ is the probability that an individual i attributes the category c (from 1 to m_j) to item j at visit t . Regarding item parameters, LPCM includes $(m_j - 1)$ difficulty parameters $(\delta_{jc})_{c=2, \dots, m_j}$ for each item j at m_j possible response categories. Note that for each item j with m_j categories, difficulty parameters are such as $\delta_{j2} < \dots < \delta_{jm_j}$, with δ_{j1} set to 0 to ensure the model identifiability. These δ_{jc} parameters reflected the latent variable value at which one had the same probability to answer the response category c than $c - 1$ for the item j . These parameters provided the possibility of seeing a hierarchy between the different items of one HRQoL scale, according to the difficulty for each patient to answer a question. More intuitively, the parameters marked out regions of "single most probable response" for all categories.³⁰ Concerning the general interpretation of this model, an increase in the latent variable (θ_{it}) induces a higher probability ($P_{jc}(\theta_{it})$) to observe high categories of responses.

Under equation 2, the LPCM can be seen as a generalized LMM with a multinomial logit link. To take into account the variability due to repeated measurements and to treat time and treatment arm as factors, the latent trait (θ_{it}) was decomposed linearly as in the CTT approach in equation 1, as follows:

$$\theta_{it} = u_{i0}^{IRT} + x_i \beta_1^{IRT} + [\beta_2^{IRT} + x_i \beta_3^{IRT} + u_{it}^{IRT}]f(t) \tag{3}$$

The latent trait θ_{it} represents the level of HRQoL for patient n at time t .

Analyses were performed for both models and separately for each HRQoL scale, according to the item grouping proposed by EORTC.²¹ Indeed, the latent variable (θ_{it}) was shared for every item at 1 dimension. The difficulty parameters allowed for taking into account responses from different items,

Table 1 Results of QLQ-C30 Functional Scales and GHS/QoL

	LMM			LPCM		
	Coefficient	SE	<i>P</i>	Coefficient	SE	<i>P</i>
GHS/QoL						
β ₂	0.0370	0.0121	0.002	0.0022	0.0015	0.147
β ₃	0.0236	0.0138	0.088	0.0030	0.0018	0.102
Physical functioning						
β ₂	0.0118	0.0135	0.382	0.0036	0.0019	0.063
β ₃	-0.0002	0.0161	0.990	-0.0036	0.0025	0.152
Role functioning						
β ₂	0.0301	0.0198	0.130	0.0001	0.0021	0.944
β ₃	0.0159	0.0233	0.495	-0.0041	0.0027	0.132
Emotional functioning						
β ₂	0.0690	0.0149	<0.001	-0.0098	0.0019	<0.001
β ₃	-0.0004	0.0176	0.983	-0.0001	0.0023	0.985
Cognitive functioning						
β ₂	0.0007	0.0128	0.958	0.0001	0.0015	0.918
β ₃	0.0166	0.0152	0.276	-0.0031	0.0018	0.096
Social functioning						
β ₂	0.0131	0.0176	0.458	-0.0005	0.0020	0.799
β ₃	0.0148	0.0210	0.481	-0.0031	0.0025	0.228

Note: GHS = global health status; QoL = quality of life; LMM = linear mixed model; LPCM = longitudinal partial credit model; β₂ = slope (evolution) of the health-related quality-of-life perception for the gemcitabine group; β₃ = time-treatment interaction; β₂ + β₃ = slope (evolution) of health-related quality-of-life perception for the FOLFIRINOX group; SE = standard error; *P* = *P* value.

concerning the common individual part for the HRQoL dimension. The covariate effect was directly interpretable on the score for LMM and on the latent variable for LPCM from equation 3.

For LMM and LPCM, several fixed effects were tested, such as treatment, time, and their interaction. In a randomized clinical trial, there should be no difference at baseline (β₁ = 0), and one would expect the slopes to possibly differ only between groups (β₃). Indeed, the questionnaire was completed at first visit (baseline) before the randomization and usually involved no difference at baseline. We tested, using the Wald test, that there was no difference of HRQoL perception at baseline with β₁ not being significantly different from 0 (Supplementary Tables 1 and 2). In either approach, the parameter interpretation is the same.

Software

The data analyses of the study were generated using SAS software (version 9.3). The mixed and nlmixed SAS procedures were used to implement, respectively, the LMM and the LPCM. The restricted maximum likelihood method (REML) was used for the estimation of the fixed and random effect parameters. An adaptive Gauss-Hermite algorithm allowed

the calculation of the REML, and the Newton-Raphson algorithm was used for the maximization of the model likelihood.

RESULTS

HRQoL Data Characteristics

Figures 1 and 2 present the description of the variable *S* over time for both treatment groups for the functioning dimensions (Figure 1) and the symptomatic dimensions (Figure 2). This variable was not continuous, and its distribution was not symmetric, reflecting the categorical nature of the data. The categorical feature of the data was still present when we used the HRQoL score *S*. Indeed, the fewer items for each dimension, the more clearly the qualitative aspect of the associated variable emerged. This aspect was less obvious when the dimension comprised several items with a higher number of response categories. The way data are represented in Figures 1 and 2 made it possible to see the response frequency for each treatment arm, the value of *S*, and the visit. For single-item dimensions (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties), there were only 4 possible values for the score, considering

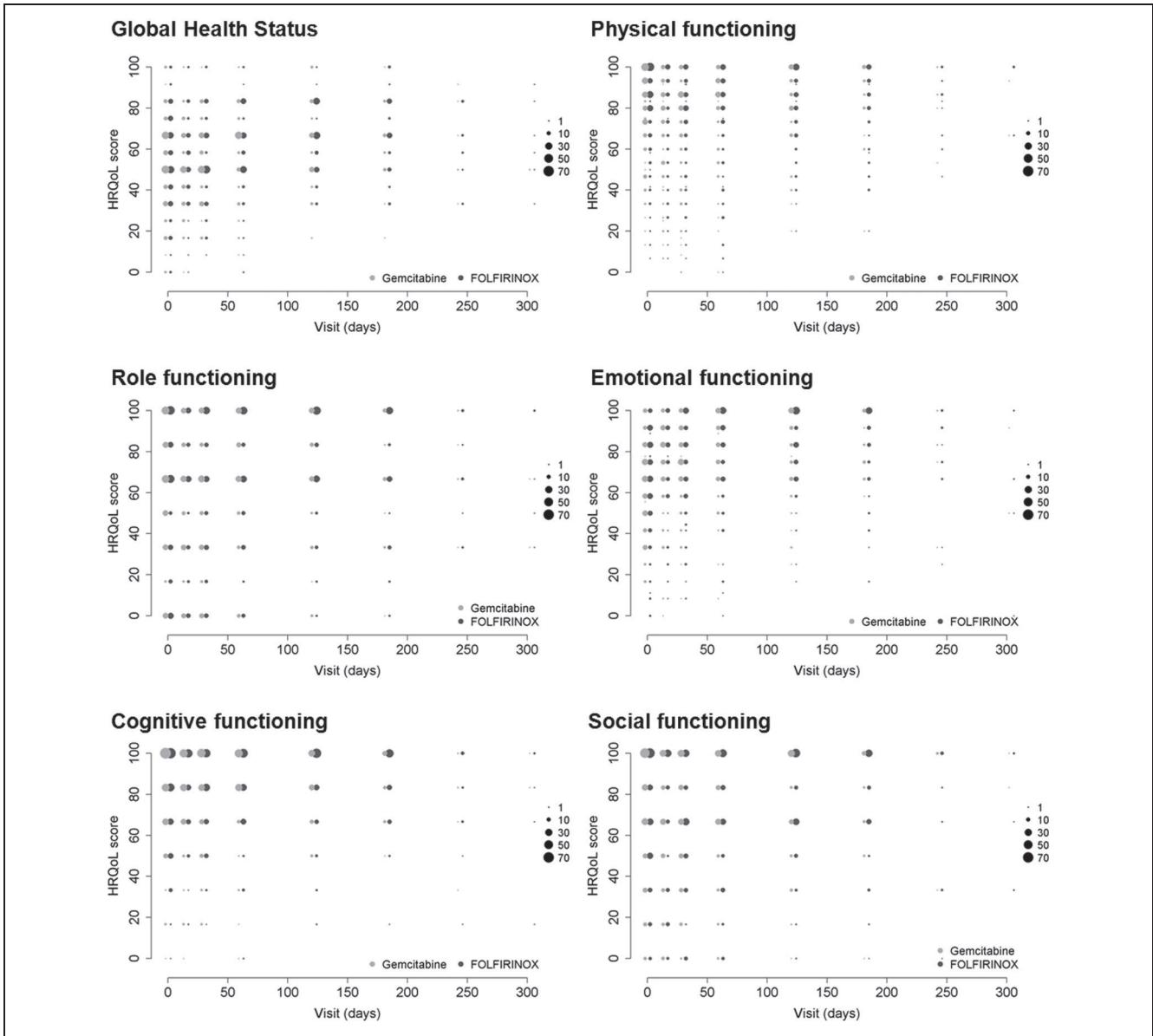


Figure 1. Description of the variable *S* over time for the gemcitabine (pale gray) and the FOLFIRINOX (dark gray) arms, concerning the functioning dimensions and the global health status. The dot size is proportional to the frequency of the scores obtained.

the Likert scale with 4 categories of response. Visually, the 2 treatment groups presented nearly the same frequency of responses.

Longitudinal Analysis

Tables 1 and 2 compare results from the 2 LMM and LPCM models by means of estimates of fixed parameters, standard deviation (SE), and the associated *P* values of the Wald test, respectively, for the

functional scales, GHS/QoL (Table 1), and the symptoms scales (Table 2).

A general pattern emerged for both methods. A significant time effect (LMM and LPCM: *P* < 0.001) on HRQoL perception was highlighted for several dimensions: emotional functioning, pain, insomnia, loss of appetite, and constipation. The previous symptom scales showed that patients felt an improvement of their HRQoL over time with a decrease of the symptomatic level. The negative slope β_2 was

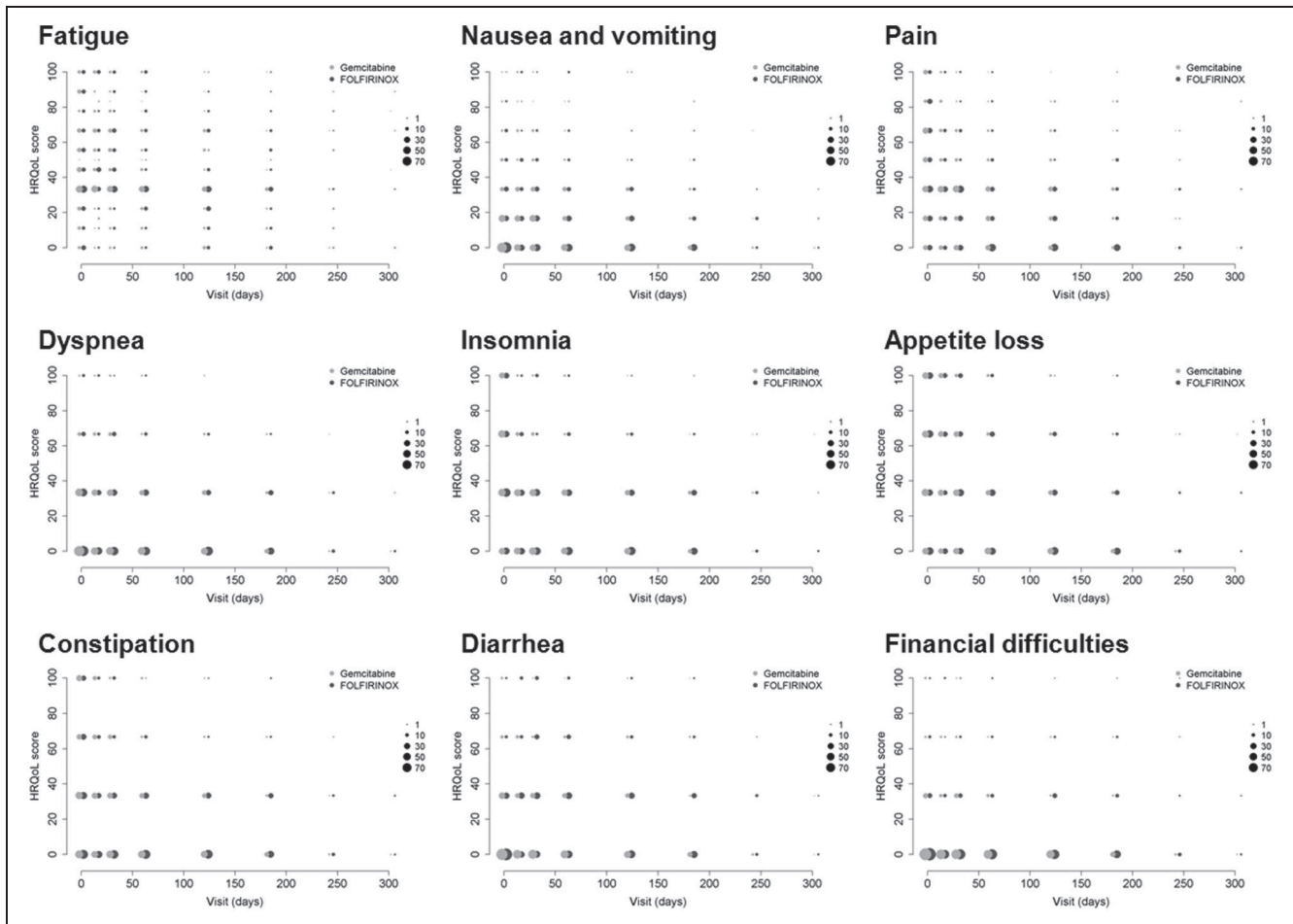


Figure 2. Description of the variable S over time for the gemcitabine (pale gray) and the FOLFIRINOX (dark gray) arms, concerning the symptomatic dimensions. The dot size is proportional to the frequency of the scores obtained.

associated with a decrease of the latent variable (θ_{it}) and of the score over time. The symptom perception also dropped, and patients showed less discomfort over time. Concerning emotional functioning, we obtained opposite slopes with the 2 models, but this does not necessarily mean the results were contradictory. Indeed, the calculated score was designed so as to take high values with a high level of functional capacity and a good HRQoL. The ordinal scale of item responses was then reversed, as done elsewhere, which was the cause of the apparently opposite results. Using our approach, when the latent trait θ decreased for the functioning scales, the HRQoL level increased. Regarding the fixed parameter related to time, the 2 approaches showed inconsistent results for 3 scales: global health status (LMM $P = 0.002$; LPCM $P = 0.147$), fatigue (LMM $P = 0.006$; LPCM $P = 0.283$), and nausea and vomiting (LMM $P = 0.039$;

LPCM $P = 0.337$). The LMM showed a significant change over time of HRQoL perception for these scales while the LPCM did not detect any change. This pattern of estimates could be explained by the relationship between the latent trait (θ) value and the difficulty parameters versus the scoring procedure. Indeed, these 3 HRQoL dimensions have at least 2 items for which the value of some difficulty parameters from different items of the same dimension is very close (Table 3; e.g., $\delta_{10,4} \approx \delta_{12,4}$ for fatigue or $\delta_{29,4} \approx \delta_{30,4}$ for the global health status). Thus, a lower variation (nonsignificant in the LPCM) of θ due to the time effect may explain the item responses unit change concerning these 2 items. For the LMM, a unit change over time items from the same dimension significantly changes the score.

Concerning the 2 others parameters β_1 and β_3 , enabling an HRQoL perception difference between

Table 2 Results of QLQ-C30 Symptoms Scales and Financial Difficulties

	LMM			LPCM		
	Coefficient	SE	P	Coefficient	SE	P
Fatigue						
β ₂	-0.0502	0.0183	0.006	-0.0025	0.0023	0.283
β ₃	0.0038	0.0217	0.860	-0.0009	0.0030	0.753
Nausea/vomiting						
β ₂	-0.0297	0.0143	0.039	-0.0015	0.0015	0.337
β ₃	-0.0046	0.0159	0.774	-0.0020	0.0018	0.273
Pain						
β ₂	-0.0826	0.0181	<0.001	-0.0089	0.0021	<0.001
β ₃	-0.0364	0.0201	0.071	-0.0053	0.0025	0.033
Dyspnea						
β ₂	-0.0129	0.0164	0.431	-0.0017	0.0021	0.433
β ₃	-0.0235	0.0191	0.218	-0.0032	0.0025	0.207
Insomnia						
β ₂	-0.1103	0.0193	<0.001	-0.0103	0.0023	<0.001
β ₃	0.0081	0.0214	0.706	0.0011	0.0023	0.643
Loss of appetite						
β ₂	-0.1289	0.0215	<0.001	-0.0080	0.0017	<0.001
β ₃	-0.0149	0.0241	0.537	-0.0011	0.0019	0.569
Constipation						
β ₂	-0.1124	0.0190	<0.001	-0.0084	0.0019	<0.001
β ₃	0.0107	0.0204	0.599	-0.0001	0.0020	0.977
Diarrhea						
β ₁	9.0733	2.5793	<0.001	0.5167	0.1749	0.003
β ₂	0.0143	0.0185	0.439	0.0001	0.0018	0.965
β ₃	-0.0273	0.0228	0.232	-0.0003	0.0018	0.852
Financial difficulties						
β ₂	-0.0045	0.0125	0.717	-0.0115	0.0074	0.124
β ₃	0.0261	0.0155	0.093	0.0093	0.0055	0.091

Note: LMM = linear mixed model; LPCM = longitudinal partial credit model; β₁ = effect difference at baseline between FOLFIRINOX and gemcitabine groups; β₂ = slope (evolution) of health-related quality-of-life perception for the gemcitabine group; β₃ = time-treatment interaction; β₂ + β₃ = slope (evolution) of health-related quality-of-life perception for the FOLFIRINOX group; SE = standard error; P = P value.

both treatment groups, only the diarrhea symptom revealed a significantly different perception (β₁) between the 2 groups at baseline (LMM *P* < 0.001; LPCM *P* = 0.003). At the first visit, patients in the experimental arm presented with a higher diarrhea level than patients of the standard arm, in which no one had received any treatment. The difference remained constant over time; no change in the diarrhea perception was significant. A significant distinction between the 2 treatment arms was observed for pain symptom (LMM *P* = 0.071; LPCM *P* = 0.033). The LPCM showed that pain perception decreased significantly over time and especially in the experimental arm (FOLFIRINOX). The difference between the effects of the 2 treatments was not significant for the HRQoL dimension using the LMM.

Figures 3 and 4 show the HRQoL evolution over time for both treatment groups. The HRQoL (θ_{it}) level

was specific for each patient *i*, dimension, and time *t*. The results in Tables 1 and 2 confirm if the observed trends in Figures 3 and 4 were or were not significant. Concerning the HRQoL scales, the pattern is the same for both treatment groups (Figures 3 and 4), except for the pain symptom scale.

Concerning the item parameters (Table 3), they are specific to the items and do not change over time in the LPCM. Regarding an individual item, these parameters enable the prediction of the response categories with the highest probability of response, given the latent trait value. Regarding several items, they allow for the comparison of the response difficulty of one item with another at a given capacity level. The fact that the data are categorical ordinal suggests that the difficulty parameters should be ordered. However, there is no constraint for these parameters in the LPCM, and it is therefore possible

Table 3 Difficulty Parameter Estimates for Item j from the Item Response Theory Model

	$\delta_{j,2}$	$\delta_{j,3}$	$\delta_{j,4}$	$\delta_{j,5}$	$\delta_{j,6}$	$\delta_{j,7}$
Global health status/QoL						
Item 29	-2.876	-2.408	-1.174	0.356	1.951	4.229
Item 30	-3.274	-2.200	-1.168	0.298	1.575	4.034
Physical functioning						
Item 1	-0.809	1.432	2.218	—	—	—
Item 2	-1.204	0.880	1.745	—	—	—
Item 3	1.657	1.992	3.271	—	—	—
Item 4	0.194	1.850	2.645	—	—	—
Item 5	3.964	3.524	4.394	—	—	—
Role functioning						
Item 6	-1.353	1.139	2.195	—	—	—
Item 7	-1.058	1.055	2.040	—	—	—
Emotional functioning						
Item 21	-0.995	1.486	3.200	—	—	—
Item 22	-1.464	0.992	1.898	—	—	—
Item 23	-0.382	1.777	2.890	—	—	—
Item 24	-0.201	1.686	2.575	—	—	—
Cognitive functioning						
Item 20	0.673	2.154	3.060	—	—	—
Item 25	0.818	2.920	4.623	—	—	—
Social functioning						
Item 26	-0.162	2.032	3.592	—	—	—
Item 27	-0.563	1.177	2.179	—	—	—
Fatigue						
Item 10	-3.793	0.468	2.156	—	—	—
Item 12	-2.690	0.739	2.169	—	—	—
Item 18	-3.132	0.530	2.448	—	—	—
Nausea/vomiting						
Item 14	0.050	1.725	2.092	—	—	—
Item 15	1.529	2.319	2.497	—	—	—
Pain						
Item 9	-1.662	0.994	1.514	—	—	—
Item 19	-0.822	1.175	1.991	—	—	—
Dyspnea						
Item 8	0.342	2.521	3.666	—	—	—
Insomnia						
Item 11	-0.804	1.229	1.765	—	—	—
Appetite loss						
Item 13	-0.8823	0.178	1.157	—	—	—
Constipation						
Item 16	-0.03693	1.086	1.198	—	—	—
Diarrhea						
Item 17	0.8405	1.7984	2.1774	—	—	—
Financial difficulties						
Item 28	3.7996	6.6507	8.7671	—	—	—

to observe inconsistent rankings. For example, concerning item 5, “physical dimension,” the fact that $\delta_{5,3} < \delta_{5,2} < \delta_{5,4}$ suggests that category 2 for this item is invariably less likely than an answer of category

1, 3, or 4, whatever the value of the latent variable. For several health dimensions, HRQoL perception improves over time, which suggests that patients tended to choose lower response categories over time.

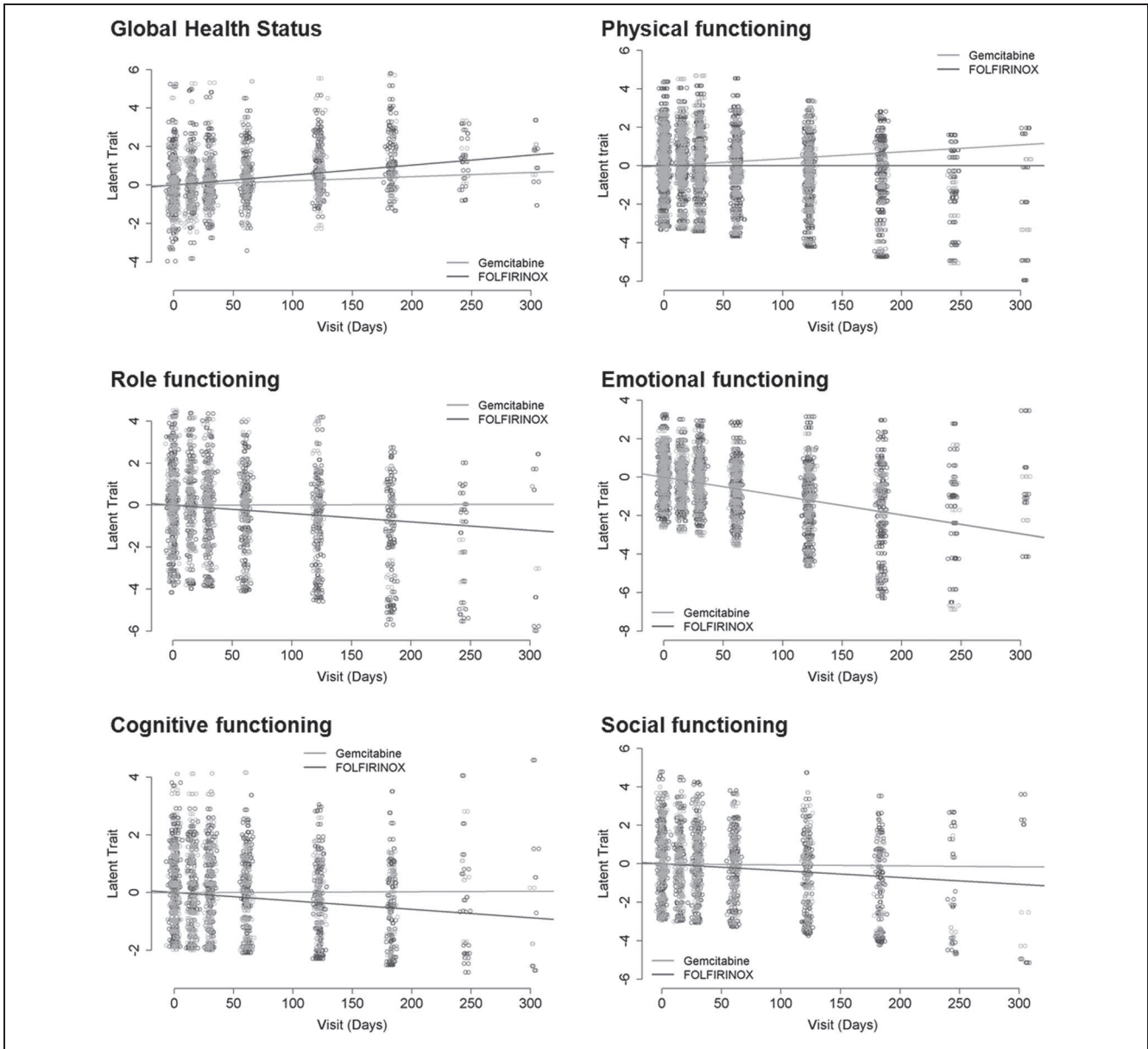


Figure 3 Health-related quality of life (HRQoL) evolution over time for the gemcitabine (pale gray) and the FOLFIRINOX (dark gray) arms. Each point represents the specific value of the patients' HRQoL at different visits and the 2 lines the average trends of HRQoL perception for the 2 treatment groups over time. A high value of latent trait represents a low HRQoL level for the studied functioning dimensions and a high HRQoL level for the global health status dimension.

DISCUSSION

Statistical research on longitudinal analysis of HRQoL continues to pose a major challenge to improving patients' care and to better evaluating the impact of treatments on their everyday life. The methodology proposed in this study is suitable for the

patient-reported outcomes. Despite the fact that the LPCM is more complex than the models currently used, it seems more appropriate because it is built on raw data and not on the summary variable *S*. As the LPCM has, to our knowledge, never been used to analyze the HRQoL data from cancer clinical trials, we have studied the robustness of its estimates

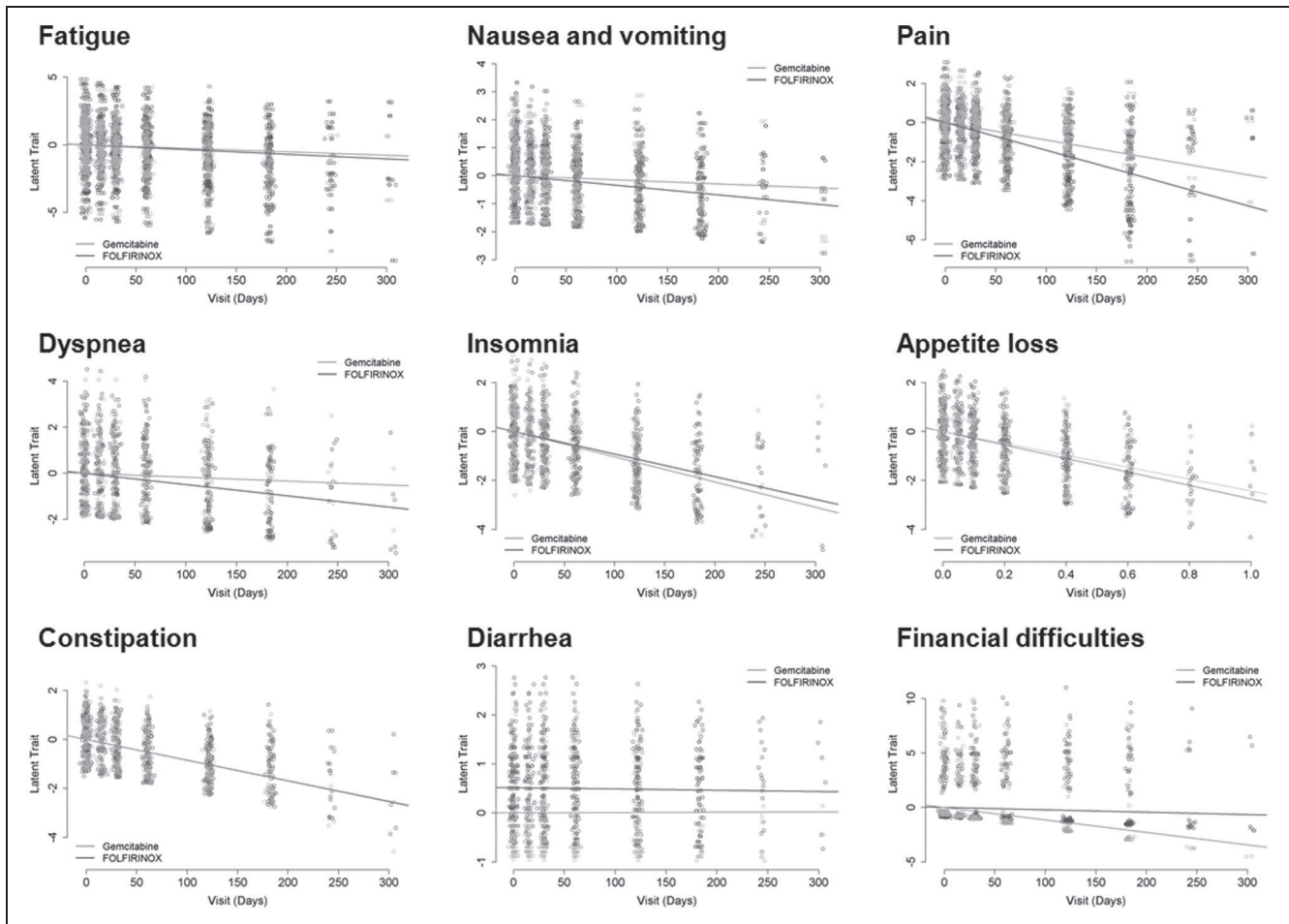


Figure 4 Health-related quality of life (HRQoL) evolution over time for the gemcitabine (pale gray) and the FOLFIRINOX (dark gray) arms. Each point represents the specific value of the patients' HRQoL at different visits and the 2 lines the average trends of HRQoL perception for the 2 treatment groups over time. A high value of latent trait represents a low HRQoL level for each dimension.

following a cross-validation study. The LPCM was fitted on randomly selected samples ($N = 100$) among the initial sample (sample size equals two-thirds of the full sample). This additional study (see Supplementary Figures 1 to 5) showed that the LPCM was stable regarding its parameter estimations and its clinical conclusions.

The HRQoL score is not a Gaussian variable because its distribution is not symmetric and its categorical aspect invariably features in the majority of health scales. It is important to consider the questionnaire responses as qualitative data, even when the score is reported. Indeed, the score and raw data have the same characteristics when we consider the few items of one health dimension (diarrhea, dyspnea, constipation, appetite loss, insomnia, and financial difficulties symptom). The use of LMM to study

the HRQoL score does not seem appropriate because the basic statistical assumptions do not check (the HRQoL score must be a continuous Gaussian variable). Our IRT model may be more complex than the classical models but shows the same qualities and gives similar results, and it has the additional advantage of being more precise and appropriate because of the use of the raw data directly.

The HRQoL study with the mixed regression models does not detect substantial differences between the 2 treatment arms for the PRODIGE 4/ACCORD 11 trial data. Both models show a similar impact of the 2 treatments on patients' HRQoL level. Only 2 HRQoL dimensions of the EORTC QLQ-C30 (pain and diarrhea) are significantly different between the gemcitabine and FOLFIRINOX arms. From LPCM results, the pain level decreased significantly faster

over time in the FOLFIRINOX arm compared with the gemcitabine arm. For this dimension, the LMM result was almost significant for the different evolution between both treatments. This could be an explanation because the LMM is based on a summary variable, which cannot provide enough information to detect a difference. Regarding the diarrhea symptom, a bias concerning the sample was detected that may explain the difference observed. At baseline, patients treated with gemcitabine reported a lower symptomatic level than those treated with FOLFIRINOX. This difference was constant over time, and their diarrhea perception did not change. This result is unexpected because this difference at baseline is observed before treatment administration and both groups are randomized from the same population. On this point, we surmise that the diarrhea perception level remained constant over time, irrespective of the treatment arm.

For most functions or symptoms, analyses with both LMM and LPCM seem to offer a similar quantification of time or treatment effects. Considering the results based on time-to-event approach with the TUDD analysis,⁷ a difference between the 2 treatment arms in favor of FOLFIRINOX was detected for the following dimensions: global health status, physical functioning, role functioning, cognitive functioning, social functioning, fatigue, nausea and vomiting, pain, and dyspnea. These conflicting results may be explained because the mixed model approach examines the different effects of the 2 treatments over time, while TUDD analysis compares the deterioration of HRQoL with the number of events occurring in each arm. The HRQoL level is the same for both treatment arms, but the number of observed events is different between the two of them. Mixed models and TUDD approaches show results that may sometimes seem contradictory, but their interpretations are different and, from our perspective, complementary. An example can be seen by comparing 2 similar cancer clinical trials investigating the effect of bevacizumab. In the first trial,³¹ HRQoL analysis through TUDD showed that the bevacizumab group had a later deterioration of QoL in contrast to patients in the standard group. Conversely, in the second trial,³² HRQoL analysis using the LMM showed a worse HRQoL over time in the bevacizumab group. The 2 conclusions are contradictory, but it is noteworthy that the 2 approaches should not be interpreted in a similar way because of the 2 different methodologies used.

Many HRQoL dimensions are affected by time, which complicates the interpretation of any changes. For example, the perception of an increase in the

HRQoL level may be due to a treatment, whereas the perception of a decrease in this level may be explained by disease progression. HRQoL evolution can be reflected by several patterns and not only by the disease or treatment; a phenomenon called "response shift"³³ can be at the origin of this change. This phenomenon is due to the patient's adaptation to her or his illness and/or treatment. Response shift is then a potential source of bias during the analysis of the different treatments' effects on HRQoL. Indeed, HRQoL is self-evaluated by the patient and may therefore be influenced by psychological phenomena or environmental factors. Finally, the observed shift may not necessarily correspond to a time or treatment effect if response shift occurs. Its effect can bias conclusions of the analysis because it can either over- or underestimate the treatment effect. While methodological and statistical methods exist to characterize the occurrence of this effect,³⁴ no recommendation has been proposed to take into account its occurrence in a longitudinal analysis.

From a statistical point of view, the major argument in favor of the use of LPCM instead of the LMM is that the data are naturally categorical. This allows a more proper and robust statistical analysis, although at the cost of a more complex clinical interpretation. Also, the LMM does not take into account the floor and ceiling effects that can result from biased estimations. With these effects, the response variable would rather have the characteristics of a binary variable: the weakest (strongest) category versus the others associated with the floor (respectively, ceiling) effect. For example, the dimension associated with financial difficulties is subjected to the floor effect (Figure 2). These effects highlight the categorical aspect of the data and separate even more the LMM from the basic assumptions. Conversely, another advantage of LPCM is the possibility of applying directly the item responses in a more accurate manner and working directly with the raw data. LPCM also includes item parameters, allowing for a more articulate interpretation of the underlying effects. Indeed, a sharper analysis can be made on a specific item with the IRT model and not just on the whole dimension.¹⁸

Another statistical problem that is not considered in the analysis of HRQoL data stems from multiple testing. We tested simultaneously at least 15 dimensions without adjusting the type I error in order to obtain a global type I error of 5%. If the 2 tests were to give conflicting results, we would suspect that the tests might not be significant in the absence of a suitable type I error adjustment.

In cancer clinical trials, the most popular method currently used for the longitudinal analysis of HRQoL is the LMM.³⁵ Time-to-event approaches are increasingly employed.^{14,36,37} In this article, we presented another strategy using IRT through the analysis of HRQoL data in the PRODIGE 4/ACCORD 11 trial. In addition to the mixed models discussed above, this alternative strategy offers as a benefit its suitability to fitting the structure of HRQoL data. Indeed, by being discretized, the score estimated in the classical approach presents problems related to the assumption underlying the use of LMM or treating one item dimension with the TUDD approach. However, each model used for longitudinal analyses of HRQoL has its own advantages, and the IRT approach seems the most robust from a methodological and psychometric point of view.

ACKNOWLEDGMENTS

We thank Dr. Hélène de Forges for her editorial assistance.

REFERENCES

1. Fiteni F, Westeel V, Pivot X, Borg C, Vernerey D, Bonnetain F. Endpoints in cancer clinical trials. *J Visc Surg*. 2014;151(1):17–22.
2. Anota A, Barbieri A, Savina M, et al. Comparison of three longitudinal analysis models for the health-related quality of life in oncology: a simulation study. *Health Qual Life Outcomes*. 2014; 12(1):1326.
3. Hardouin J-B, Blanchin M, Feddag M-L, Néel TL, Perrot B, Sébille V. Power and sample size determination for group comparison of patient-reported outcomes using polytomous Rasch models. *Stat Med*. 2015;34:2444–55.
4. Beitz J, Gnecco C, Justice R. Quality-of-life end points in cancer clinical trials: the U.S. Food and Drug Administration perspective. *J Natl Cancer Inst Monogr*. 1996;20:7–9.
5. Payne SA. A study of quality of life in cancer patients receiving palliative chemotherapy. *Soc Sci Med*. 1992;35(12):1505–9.
6. Conroy T, Desseigne F, Ychou M, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N Engl J Med*. 2011; 364(19):1817–25.
7. Gourgou-Bourgade S, Bascoul-Mollevis C, Desseigne F, et al. Impact of FOLFIRINOX compared with gemcitabine on quality of life in patients with metastatic pancreatic cancer: results from the PRODIGE 4/ACCORD 11 randomized trial. *J Clin Oncol Off J Am Soc Clin Oncol*. 2013;31(1):23–9.
8. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med*. 1997;16(20):2349–80.
9. Fairclough DL. Design and Analysis of Quality of Life Studies in Clinical Trials. Boca Raton (FL): Chapman and Hall; 2002.
10. Bonnetain F, Dahan L, Maillard E, et al. Time until definitive quality of life score deterioration as a means of longitudinal analysis for treatment trials in patients with metastatic pancreatic adenocarcinoma. *Eur J Cancer*. 2010;46(15):2753–62.
11. Anota A, Hamidou Z, Paget-Bailly S, et al. Time to health-related quality of life score deterioration as a modality of longitudinal analysis for health-related quality of life studies in oncology: do we need RECIST for quality of life to achieve standardization? *Qual Life Res*. 2015;24(1):5–18.
12. Hamidou Z, Dabakuyo TS, Mercier M, et al. Time to deterioration in quality of life score as a modality of longitudinal analysis in patients with breast cancer. *Oncologist*. 2011;16(10):1458–68.
13. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol*. 1998;16(1):139–44.
14. Burris HA, Lebrun F, Rugo HS, et al. Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. *Cancer*. 2013;119(10): 1908–15.
15. Grilli L, Rampichini C. Multilevel models for ordinal data. In: Kenett RS, Salini S, eds. *Modern Analysis of Customer Surveys*. New York: John Wiley & Sons; 2011. p 391–411. Available from: URL: <http://onlinelibrary.wiley.com/doi/10.1002/9781119961154.ch19/summary>.
16. Rasch G. On General Laws and the Meaning of Measurement in Psychology. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4. Berkeley: University of California Press; 1961. p 321–33. Available from: URL: <http://projecteuclid.org/euclid.bsm/1200512895>.
17. Masters G. A Rasch model for partial credit scoring. *Psychometrika*. 1982;42(2):149–74.
18. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*. 2007;16(suppl 1):5–18.
19. 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. *J Natl Cancer Inst*. 1993;85(5):365–76.
20. Fayers PM, Machin D. *Quality of Life: Assessment, Analysis and Interpretation*. New York: Wiley; 2000. Available from: URL: <http://www3.interscience.wiley.com/cgi-bin/bookhome/91014951>.
21. Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A. *EORTC QLQ-C30 Scoring Manual (3rd edition)*. Brussels: EORTC 2001 ed2001.
22. De Ayala RJ. *The Theory and Practice of Item Response Theory*. New York: Guilford Press; 2009.
23. de Boeck P, Wilson M. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer; 2004.
24. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther*. 2014;36(5):648–62.
25. Glas CAW, Geerlings H, van de Laar MAFJ, Taal E. Analysis of longitudinal randomized clinical trials using item response models. *Contemp Clin Trials*. 2009;30(2):158–70.
26. Brown H, Prescott R. *Applied Mixed Models in Medicine*. 2nd ed. New York: Wiley-Blackwell; 2006.

27. [Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. Boca Raton \(FL\): CRC Press; 2012.](#)
28. [Agresti A. Analysis of Ordinal Categorical Data. New York: John Wiley & Sons; 2010.](#)
29. [Bacci S. Analysis of longitudinal HrQoL using latent regression in the context of Rasch modeling. In: Huber C, Limnios N, Mesbah M, Nikulin M, eds. Mathematical Methods in Survival Analysis, Reliability and Quality of Life. New York: Wiley; 2008. p 275–90.](#)
30. [Van der Linden WJ, Hambleton RK. Handbook of Modern Item Response Theory. New York: Springer; 1997.](#)
31. [Chinot OL, Wick W, Mason W, et al. Bevacizumab plus radiotherapy–temozolomide for newly diagnosed glioblastoma. N Engl J Med. 2014;370\(8\):709–22.](#)
32. [Gilbert MR, Dignam JJ, Armstrong TS, et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. N Engl J Med. 2014;370\(8\):699–708.](#)
33. [Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med. 1999;48\(11\):1531–48.](#)
34. [Anota A, Bascoul-Mollevi C, Conroy T, et al. Item response theory and factor analysis as a mean to characterize occurrence of response shift in a longitudinal quality of life study in breast cancer patients. Health Qual Life Outcomes. 2014;12:32.](#)
35. [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. J Clin Oncol. 2014;32\(13\):1309–16.](#)
36. [Kabbinar FF, Wallace JF, Holmgren E, et al. Health-related quality of life impact of bevacizumab when combined with irinotecan, 5-fluorouracil, and leucovorin or 5-fluorouracil and leucovorin for metastatic colorectal cancer. Oncologist. 2008;13\(9\):1021–9.](#)
37. [Wu Y, Amonkar MM, Sherrill BH, et al. Impact of lapatinib plus trastuzumab versus single-agent lapatinib on quality of life of patients with trastuzumab-refractory HER2+ metastatic breast cancer. Ann Oncol. 2011;22\(12\):2582–90.](#)

Sélection conceptuelle de modèle(s) IRT

Dans ce chapitre, les modèles IRT sont replacés dans le contexte des GLMM pour données catégorielles. En effet, les modèles IRT représentent un sous-ensemble de ces modèles et se distinguent par une paramétrisation spécifique du prédicteur linéaire. Nous proposons une sélection conceptuelle des modèles IRT, orientée pour l'analyse spécifique de la HRQoL en cancérologie. Cette sélection est effectuée sur la base d'arguments pratiques et méthodologiques mis en exergue par les propriétés des modèles ainsi que les attentes et objectifs cliniques.

Sommaire

3.1	Introduction	58
3.2	GLMM pour données catégorielles	59
3.2.1	Modèle cumulatif	62
3.2.2	Modèle adjacent	63
3.2.3	Modèle séquentiel	64
3.3	Sélection conceptuelle	65
3.4	Discussion	68
3.5	Article : <i>Selection of item response theory models for the longitudinal analysis of health-related quality of life in cancer clinical trials</i>	69

3.1 Introduction

Plusieurs freins à l'utilisation des modèles IRT pour les analyses longitudinales subsistent, notamment le manque d'exemples concrets, d'accessibilité computationnelle, ou encore d'arguments pour le choix du modèle IRT pour données polytomiques. Dans la littérature, différentes modélisations sont présentées dans ce contexte, sans pour autant qu'il n'y ait de critères ou d'arguments permettant de justifier le choix d'un modèle par rapport à un autre. De nombreux travaux présentent les différentes modélisations IRT telles qu'elles sont présentées dans la section 2.2 (Boeck and Wilson, 2004; Bacci et al., 2012; Titman et al., 2013). Par ailleurs, l'utilisation de ces modèles pour des analyses concrètes est rare et le choix conceptuel du modèle IRT est la plupart du temps peu expliqué ou alors statisticien-dépendant. Les études réalisées se concentrent sur un modèle particulier, que ce soit sur le PCM (Hardouin et al., 2012; Bock et al., 2013; Verhagen and Fox, 2013; Anota et al., 2014a; Guilleux et al., 2015; Barbieri et al., 2015), le GRM (Douglas, 1999; Liu and Hedeker, 2006; van Nispen et al., 2007; Edelen and Reeve, 2007), ou encore le SRM (Tutz, 1990). Un argument est mis en avant quant à l'utilisation du modèle PCM, est sa propriété d'objectivité spécifique. Elle réfère à l'indépendance entre le niveau du trait latent obtenu et l'instrument de mesure, et caractérise les modèles de la famille de Rasch (Hardouin et al., 2015)¹.

Aujourd'hui, les modèles IRT sont de plus en plus plébiscités dans les analyses longitudinales de données issues de questionnaire (Gorter et al., 2015; Barbieri et al., 2015). Les modèles IRT sont des GLMM pour données catégorielles (Boeck and Wilson, 2004; Agresti, 2007) avec une paramétrisation spécifique du prédicteur linéaire permettant la distinction des différentes sources de la variable réponse (items) (voir Figure 3.1). Cette paramétrisation permet de contourner la difficulté de considérer un modèle à équations structurelles où les réponses aux items seraient des multi-réponses aléatoires. Ces ajouts de paramètres d'item permettent de manière habile de considérer la donnée comme étant issue d'une seule variable réponse. Dans la littérature, plusieurs modèles IRT sont référencés mais peu d'indications sont disponibles sur le choix du modèle IRT le plus pertinent dans notre situation, c'est-à-dire pour l'analyse longitudinale de la HRQoL en cancérologie. Pour y palier, ces modèles ont été repositionnés dans le contexte des GLMM dont les propriétés et définitions sont plus discutées.

Les travaux sur les GLMM ont été initiés dans les années 1980 et résultent d'une double extension. D'une part, une extension en termes de loi a été envisagée dans le but de prendre en compte des données non gaussiennes dans les modèles de régression. Nelder and Wedderburn (1972) ont alors introduit la famille des modèles linéaires généralisés (*generalized linear models*, GLM) il y a plus de quarante ans. D'autre part, une extension en termes d'introduction d'effets aléatoires, qui permet de séparer les différentes sources de variabilité induites par les données. L'ajout des effets aléatoires se fait dans le prédicteur linéaire η qui contient alors une partie fixe et une partie

1. En d'autres mots, les paramètres de difficulté ne changent pas en fonction des observations lorsque le même questionnaire est utilisé;

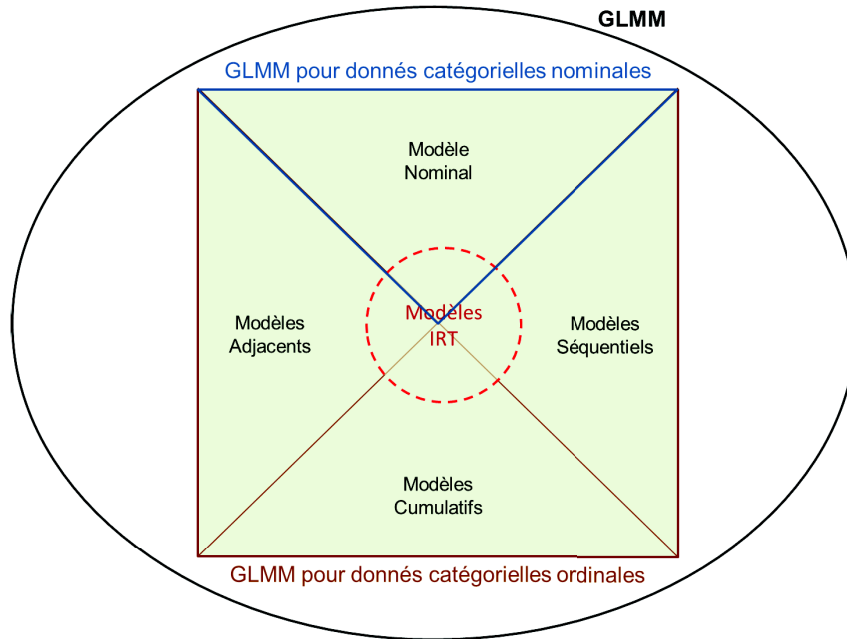


FIGURE 3.1: Schéma représentant la famille des GLMM (en noir) dans laquelle est comprise la famille des GLMM pour données catégorielles (dans le carré : en bleu pour données nominales et en orange pour données ordinales). Les modèles IRT (en rouge) sont un sous-ensemble de modèles GLMM pour données catégorielles qui se distinguent par une paramétrisation spécifique permettant la considération de plusieurs échelles catégorielles.

aléatoire tel que :

$$\eta_\xi = X\beta + U\xi, \quad (3.1)$$

où l'hypothèse de normalité sur les effets aléatoires ξ est maintenue. Conditionnellement à ξ , les différentes réalisations de la variable réponse sont indépendantes et de même loi appartenant à la famille exponentielle, et l'espérance mathématique conditionnelle $\mathbb{E}(Y|\xi) = \mu_\xi$ est reliée au prédicteur linéaire par la fonction de lien g tel que :

$$\eta_\xi = g(\mu_\xi) \quad (3.2)$$

Tout comme les GLM, les GLMM se définissent suivant le prédicteur linéaire et la fonction de lien, mais dans un raisonnement conditionnel à ξ^2 .

3.2 GLMM pour données catégorielles

Jusqu'aux années 1960, les méthodes statistiques pour l'analyse des données catégorielles étaient peu élaborées. Depuis, de nombreuses recherches ont permis de développer pleinement ces méthodes. Elles ont été notamment motivées par des études

2. Conditionnellement à ξ , le GLMM conserve toutes les propriétés du GLM ;

en sciences sociales (économie, psychologie, médecine, ...) utilisant des échelles catégorielles. Nous distinguons deux types de données issues de ces échelles catégorielles : les données ordinales et les données nominales. Les données ordinales font référence à un ensemble de réponses possibles (supérieures à deux) où un ordre naturel existe entre celles-ci. Parmi les données ordinales, nous retrouvons par exemple l'évaluation d'un symptôme lors d'un traitement contre une maladie (avec les catégories absent, faible, moyen, fort), ou encore le niveau d'étude (avec les catégories brevet, baccalauréat, licence, master). Concernant les données nominales, elles caractérisent les réponses à plusieurs catégories de réponse (supérieures à deux) où aucun ordre naturel n'existe entre ces catégories. Par exemple, nous pouvons citer le type de musique préféré (rock, jazz, classique, autres), ou encore la couleur (naturelle) des cheveux (brun, blond, roux, châtain).

Dans le cadre des GLMM pour des réponses multinomiales, la variable réponse $Y_{iv}^{(j)}$, à M_j valeurs possibles, est réécrite en une version informative de la variable réponse comme un vecteur réponse $\mathbf{y}_{iv}^{(j)} = (y_{iv1}^{(j)}, \dots, y_{iv,M_j-1}^{(j)})$ où :

$$y_{ivm}^{(j)} = \begin{cases} 1 & \text{si } Y_{iv}^{(j)} = m, \quad m = 1, \dots, M_j - 1 \\ 0 & \text{sinon} \end{cases}, \quad (3.3)$$

pour un sujet i à la visite v pour l'item j .

Dans l'IRT, il est fréquent de considérer la première modalité ($m = 1$) de réponse comme la modalité de référence. Pour des raisons pratiques et communes aux notations des GLMM pour réponses multinomiales, la modalité M_j sera la modalité de référence dans cette partie. La réponse tronquée pour cette modalité s'exprime en fonction des autres telle que $y_{iv(M_j)}^{(j)} = 1 - \sum_{k=1}^{M_j-1} y_{ivk}^{(j)}$. La distribution de la famille exponentielle appropriée pour ces données est la distribution multinomiale. Dans le cadre des GLMM, l'intérêt se concentre essentiellement sur les probabilités. Le vecteur tronqué (Fahrmeir and Tutz, 2001) (version informative) des probabilités est $\boldsymbol{\pi}_{iv}^{(j)} = (\pi_{iv1}^{(j)}, \dots, \pi_{iv,M_j-1}^{(j)})$ (avec $\pi_{ivM_j}^{(j)} = 1 - \sum_{k=1}^{M_j-1} \pi_{ivk}^{(j)}$). La distribution multinomiale (pour un sujet i à la visite v pour l'item j) est :

$$f(\mathbf{y}_{iv}^{(j)}; \boldsymbol{\pi}_{iv}^{(j)}) = \left(\prod_{m=1}^{M_j-1} \pi_{ivm}^{(j) y_{ivm}^{(j)}} \right) \left(1 - \sum_{k=1}^{M_j-1} \pi_{ivk}^{(j)} \right)^{1 - \sum_{k=1}^{M_j-1} y_{ivk}^{(j)}} \quad (3.4)$$

La fonction de lien g permet dans les GLMM de passer de l'espérance mathématique de la variable réponse au prédicteur linéaire. Pour une réponse multinomiale, le prédicteur linéaire n'est pas directement relié à la probabilité de réponse, mais à une transformation particulière d'un rapport des différentes probabilités, tel que :

$$g(\boldsymbol{\pi}_{iv}^{(j)}) = \boldsymbol{\eta}_{iv}^{(j)},$$

où pour le cas du modèle de régression logistique multinomiale pour données nomi-

nales (Agresti, 2002), la composante $m \in \{1, \dots, M_j - 1\}$ est définie par :

$$g_m \left(\boldsymbol{\pi}_{iv}^{(j)} \right) = \log \left(\frac{\pi_{ivm}^{(j)}}{\pi_{ivM_j}^{(j)}} \right) = \eta_{ivm}^{(j)}. \quad (3.5)$$

La fonction de lien g résulte alors de la combinaison du rapport de probabilité r et de la fonction de répartition F utilisée. Peyhardi et al. (2015) ont proposé une nouvelle manière de définir les GLM pour données catégorielles, caractérisée par le triplet (r, F, Z) : rapport de probabilité (r), fonction de répartition (F) et matrice design (Z). Cette nouvelle spécification des modèles permet une meilleure compréhension de leur construction et donc une meilleure interprétation des résultats. Les propriétés des modèles paraissent également plus intuitives. Par la suite, nous utilisons cette spécification des modèles. Cependant, la dernière composante sera caractérisée par le prédicteur linéaire (η). Nous parlerons alors de design proportionnel ou non proportionnel du modèle faisant référence respectivement aux modèles à rapport de côtes proportionnelles ou non.

Pour des données catégorielles, il existe quatre familles de modèles. La première est la famille des modèles référents utilisée pour les données nominales. Ils sont définis par l'équation (3.5). Lorsqu'il s'agit de données ordinales, trois familles sont proposées : la famille des modèles cumulatifs (Samejima, 1969; McCullagh, 1980), la famille des modèles adjacents (Masters, 1982; Agresti, 2002) et enfin la famille des modèles séquentiels (Tutz, 1990). Nous retrouvons respectivement les trois célèbres modèles IRT (GRM, PCM et SRM) faisant référence aux trois familles. Ces modèles IRT ont tous été définis suivant la CdF (*cumulative distribution function*, CdF) logistique. Ces différentes familles de modèles se distinguent par le rapport de probabilité $\mathbf{r} \left(\boldsymbol{\pi}_{iv}^{(j)} \right) = \left(r_1 \left(\boldsymbol{\pi}_{iv}^{(j)} \right), \dots, r_{M_j} \left(\boldsymbol{\pi}_{iv}^{(j)} \right) \right)'$ qui les caractérise. Chaque modèle est entièrement défini par son triplet (r, F, η) tel que :

$$r_m \left(\boldsymbol{\pi}_{iv}^{(j)} \right) = F \left(\eta_{ivm}^{(j)} \right), \quad m = 1, \dots, M_j - 1, \quad (3.6)$$

avec $\eta_{ivm}^{(j)} = \delta_{jm} + X_{iv}^{(j)}\beta + U_{iv}\xi_i = \delta_{jm} + \theta_{iv}$ et $\pi_{ivm}^{(j)} = \Pr \left(Y_{iv}^{(j)} = m | \theta_{iv}, \delta_j \right)$. De manière générale, nous pouvons mettre en évidence la relation entre la fonction de lien $g_m(\cdot)$ et le triplet :

$$\begin{aligned} g_m \left(\boldsymbol{\pi}_{iv}^{(j)} \right) &= F^{-1} \circ r_m \left(\boldsymbol{\pi}_{iv}^{(j)} \right) \\ &= \eta_{ivm}^{(j)}. \end{aligned}$$

Les données de questionnaires considérées étant par définition ordinales, seuls les modèles adjacents, cumulatifs et séquentiels sont présentés par la suite. Dans un premier temps, ces différents modèles sont présentés en fonction de leur rapport de probabilité définissant leur famille d'appartenance, ainsi que par l'expression des différentes probabilités de réponse pour une CdF quelconque. Les propriétés intéressantes de ces modèles pour l'analyse longitudinale de la HRQoL seront également

mises en avant. Pour alléger les notations, les indices associés aux individus, aux visites et aux items seront supprimés. Considérons $\eta_m = \delta_m + \theta$ le prédicateur linéaire avec δ_m l'intercept lié à la catégorie de réponse $m = 1, \dots, M - 1$ et θ la pente (la partie commune à l'ensemble des catégories de réponse).

3.2.1 Modèle cumulatif

Un modèle de régression pour données ordinales est souvent associé au modèle cumulatif (McCullagh, 1980) qui, comme son nom l'indique, est construit sur les probabilités cumulées³. En considérant le triplet (r, F, η) , ce modèle est défini par :

$$r_m(\boldsymbol{\pi}) = \pi_1 + \dots + \pi_m = F(\eta_m). \quad (3.7)$$

Ce modèle fait référence au principe de seuils (Hedeker and Gibbons, 1994), où les seuils sont supposés croissants tels que $\delta_1 < \dots < \delta_{M-1}$. Les différentes probabilités de réponse sont alors définies pour toutes F par :

$$\pi_m = \Pr(Y = m|\eta) = F(\eta_m) - F(\eta_{m-1}). \quad (3.8)$$

Ce modèle a de nombreuses propriétés dont certaines sont intéressantes pour notre objectif d'étude (McCullagh, 1980; Peyhardi et al., 2015) :

- Interprétation au travers d'une variable latente sous-jacente au modèle ;
- Le modèle est invariant sous la permutation inverse⁴ à condition que la Cdf utilisée soit symétrique ;
- Le modèle est invariant sous le regroupement de deux catégories adjacentes.

Selon la première propriété énoncée, il existe donc une variable latente \tilde{Y} sous-jacente au modèle qui est la version continue de la variable aléatoire réponse catégorielle Y . D'un point de vue de l'IRT, cette variable continue inobservable correspond tout simplement au paramètre individuel θ du modèle auquel un terme d'erreur est ajouté. Du point de vue des GLMM, la variable latente continue est supposée suivre un LMM où sa décomposition linéaire ne prend pas en compte des seuils référant aux modalités de réponse. Pour les deux points de vue, elle est définie par :

$$\begin{aligned} IRT : \quad \tilde{Y} &= \theta && +\varepsilon \\ GLMM : \quad \tilde{Y} &= X\beta + U\xi && +\varepsilon, \end{aligned}$$

où tous les éléments du vecteur ε sont indépendants et distribués suivant F et $\xi \sim \mathcal{N}(0, \Sigma)$ avec $\Sigma = \text{diag}\{\sigma_k^2 A_k\}$ pour k effets aléatoire considérés et A_k une matrice connue. La variable latente continue \tilde{Y} est modélisée par un LMM. La catégorie de réponse associée à Y est alors déterminée à partir de la série supposée croissante des seuils $-\infty = \delta_0 < \delta_1 < \dots < \delta_M = +\infty$. Les équivalences entre Y et \tilde{Y} sont les suivantes :

$$\{Y = m\} \Leftrightarrow \{\delta_{m-1} \leq \tilde{Y} < \delta_m\}. \quad (3.9)$$

3. Bien qu'il n'y ait pas un rapport de probabilité mais une somme, nous utiliserons le terme "rapport" par abus de langage ;

4. l'ordre des catégories de réponse est inversé ;

Nous venons de présenter le modèle cumulatif connu sous un design proportionnel (le *proportional odds model*), où la pente est commune pour l'ensemble des catégories de réponse ($\theta = \theta_1 = \dots = \theta_M$). Dans le cas d'un modèle à design non proportionnel ($\theta_1 \neq \theta_2 \neq \dots \neq \theta_M$), la contrainte établie sur les paramètres de seuils devient $-\infty < \delta_1 + \theta_1 < \dots < \delta_{M-1} + \theta_{M-1} < \infty$, qui est plus difficile à vérifier, et rend le modèle non estimable dans certains cas (Peyhardi et al., 2015).

3.2.2 Modèle adjacent

Les modèles adjacents représentent la deuxième famille de GLMM pour données ordinales la plus utilisée en pratique. Ils sont construits sur des liens entre les modalités adjacentes. Le modèle adjacent est conceptuellement proche du modèle référence (3.5). Chaque catégorie de réponse est comparée à la modalité adjacente et non à une modalité de référence. Cela revient à appliquer le modèle pour données binaires à chaque paire de modalités de réponse adjacente. Considérant le triplet (r, F, η) , ce modèle est défini par :

$$r_m(\boldsymbol{\pi}) = \frac{\pi_m}{\pi_m + \pi_{m+1}} = F(\eta_m) \quad (3.10)$$

Dans la littérature, les modèles adjacents sont présentés essentiellement avec la CdF logistique (Masters, 1982; Agresti, 2010; Hardouin et al., 2015), ou selon son rapport de probabilité (Agresti, 2010). Cependant, ce modèle peut être défini suivant une CdF F quelconque. Notons $F_m = F(\eta_m)$ quel que soit $m = 1, \dots, M-1$ où les probabilités vérifient le lien donné par (3.10). Du rapport de probabilité (3.10), nous déduisons :

$$\begin{aligned} \pi_m &= F_m(\pi_m + \pi_{m+1}), \quad m = 1, \dots, M-1 \\ \Leftrightarrow \pi_m &= \frac{F_m}{1 - F_m} \pi_{m+1}. \end{aligned}$$

Par récurrence,

$$\begin{aligned} \pi_m &= \left(\frac{F_m}{1 - F_m} \right) \times \dots \times \left(\frac{F_{M-1}}{1 - F_{M-1}} \right) \pi_M \\ &= \pi_M \prod_{k=m}^{M-1} \left(\frac{F_k}{1 - F_k} \right). \end{aligned} \quad (3.11)$$

Étant donné $\pi_M = 1 - \sum_{m=1}^{M-1} \pi_m$ et $\sum_{m=1}^{M-1} \pi_m = \pi_M \sum_{m=1}^{M-1} \prod_{k=m}^{M-1} \left(\frac{F_k}{1 - F_k} \right)$, les probabilités se définissent donc par :

$$\begin{cases} \pi_M &= \frac{1}{1 + \sum_{m=1}^{M-1} \prod_{k=m}^{M-1} \left(\frac{F_k}{1 - F_k} \right)} \\ \pi_m &= \frac{\prod_{k=m}^{M-1} \left(\frac{F_k}{1 - F_k} \right)}{1 + \sum_{m=1}^{M-1} \prod_{k=m}^{M-1} \left(\frac{F_k}{1 - F_k} \right)}, \quad m = 1, \dots, M-1 \end{cases} \quad (3.12)$$

Ce modèle est moins restrictif que le modèle cumulatif car il n'y a pas de contraintes sur les paramètres. En effet, les paramètres δ ne sont pas contraints à suivre un

certain ordre selon les modalités de réponse. Dans le cas d'un design proportionnel, l'interprétation des effets est générale bien que le modèle ne soit pas interprétable via une variable latente. En effet, une augmentation de θ_{iv} entraînera une augmentation des probabilités associées aux fortes modalités de réponse et inversement. De plus, une propriété commune avec le modèle cumulatif est la propriété d'invariance sous la permutation inverse.

3.2.3 Modèle séquentiel

Les modèles séquentiels pour données ordinales ont été proposés par Gerhard Tutz (Tutz, 1990). L'idée sous-jacente est de considérer des données ordinales résultant d'un mécanisme séquentiel où les différentes modalités ordonnées sont atteintes successivement. Considérant le triplet (r, F, η) , ce modèle est défini par :

$$r_m(\boldsymbol{\pi}) = \frac{\pi_m}{\pi_m + \dots + \pi_M} = F(\eta_m), \quad (3.13)$$

et les différentes probabilités de réponse sont alors définies pour toutes F par (Linden and Hambleton, 1997; Fahrmeir and Tutz, 2001) :

$$\pi_m = \Pr(Y = m|\eta) = F(\eta_m) \prod_{k=1}^{m-1} \{1 - F(\eta_k)\}. \quad (3.14)$$

De la même manière que le modèle cumulatif, cette modélisation peut aussi être motivée par l'existence de variables latentes continues sous-jacentes au modèle. Le processus aléatoire continu séquentiel $(\tilde{Y}_m)_{m=1, \dots, M-1}$ a une forme linéaire telle que pour tout $m = 1, \dots, M - 1$,

$$\tilde{Y}_m = -\theta + \varepsilon_m,$$

où les éléments de ε_m sont indépendants et distribués suivant F . Ensuite, le mécanisme de réponse commence par la catégorie 1 et la première étape est définie par :

$$\{Y = 1\} \Leftrightarrow \{\tilde{Y}_1 \leq \delta_1\},$$

et si $\tilde{Y}_1 \leq \delta_1$, alors le processus s'arrête. Sinon,

$$\{Y = 2\} \Leftrightarrow \{\tilde{Y}_2 \leq \delta_2\} \cap \{\tilde{Y}_1 > \delta_1\},$$

et ainsi de suite, de telle sorte que

$$\{Y = m\} \Leftrightarrow \{\tilde{Y}_m \leq \delta_m\} \cap \bigcap_{k=1}^{m-1} \{\tilde{Y}_k > \delta_k\}.$$

Le mécanisme séquentiel est supposé être une décision binaire à chaque étape où seulement le résultat final est observé. Ces modèles séquentiels sont beaucoup moins intuitifs que les modèles cumulatifs ou les modèles adjacents. Par conséquent, leur utilisation est quasi nulle en pratique.

3.3 Sélection conceptuelle

Dans cette section, nous résumons la sélection conceptuelle de modèle IRT proposée dans l'article 3.5. Cet article reprend les modèles définis précédemment pour établir un argumentaire construit sur le triplet (r, F, η) afin de dégager les modèles les plus adéquats issus des différentes familles pour l'analyse longitudinale de la HRQoL. Les principales caractéristiques des modèles sont décrites dans la Table 3.1.

Le fait que les données de HRQoL soient toujours ordonnées, les modèles référents ne sont pas considérés. Ensuite, d'un point de vue appliqué, la propriété d'invariance sous permutation inverse est intéressante. Cette propriété se déduit de la fonction de lien du modèle, soit de l'association entre le rapport de probabilité et la CdF. En effet, elle se vérifie seulement sous les rapports de probabilité qui définissent les modèles cumulatifs et adjacents en association à une CdF symétrique. Cette propriété est importante en pratique pour l'application du modèle à des données et l'interprétation des résultats. Les items pour les dimensions fonctionnelles sont codifiés de la même manière que les items pour les dimensions symptomatiques (une catégorie de réponse élevée est connotée négativement). Pour qu'une valeur élevée de la variable latente représente une capacité fonctionnelle élevée, il est donc nécessaire de renverser l'ordre de l'échelle. Pour les mêmes raisons que pour les calculs des scores de HRQoL (équations (1.3) et (1.4)), il est intéressant de pouvoir renverser l'ordre des réponses suivant la volonté d'interprétation des cliniciens. De plus, les résultats et conclusions seront d'autant plus comparables et homogènes entre les différentes approches et applications. Les modèles séquentiels n'ont pas cette propriété d'invariance sous la permutation inverse, nous ne les considérerons pas par la suite.

TABLE 3.1: Récapitulatif des différentes caractéristiques des GLMM pour données ordinales.

Caractéristiques du modèle	Modèles			
	Référent	Adjacent	Cumulatif	Séquentiel
Invariant sous toutes permutations	<i>oui</i> ¹	<i>oui</i> ¹ / <i>non</i>	<i>non</i>	<i>non</i>
Données ordinales	<i>non</i>	<i>oui</i>	<i>oui</i>	<i>oui</i>
Réversible*	<i>oui</i>	<i>oui</i>	<i>oui</i>	<i>non</i>
Interprétation en variable latente	<i>non</i>	<i>non</i>	<i>oui</i>	<i>oui</i>
Toujours défini	<i>oui</i>	<i>oui</i>	<i>oui/non</i> ²	<i>oui</i>

¹ : seulement sous la distribution logistique ² : pour quelques modèles à rapport de côtes non proportionnelles

* : pour distributions symétriques

Toujours dans notre contexte, un argument supplémentaire intéressant est la facilité d'interprétation des résultats. Les modèles cumulatifs ont un avantage certain sur les modèles adjacents : l'interprétation via une variable latente. La Figure 3.2 illustre les résultats de la dimension de douleur de l'essai Prodiges 4/ACCORD 11, associés à l'item 9. Pour cette dimension, il y a une diminution significative de la perception de douleur. De plus, cette diminution est d'autant plus importante pour le groupe Folfirnox (en marron) que pour le groupe Gemcitabine (en bleu) (Figure 3.2a). Il y a plus de modalités de réponse faibles qui sont observées dans le groupe expérimental à 4 mois (Figure 3.2b). Cette différence s'accroît significativement au

cours du temps. La variable latente θ s'interprète de la même manière qu'une variable réponse modélisée par un LMM, ce qui est beaucoup plus familier et intuitif pour les cliniciens. Les paramètres d'item permettent de faire directement le lien (équation 3.9) entre la variable latente et la variable réponse ordinaire Y .

Concernant le modèle adjacent, l'interprétation des coefficients obtenus avec le modèle adjacent est similaire à celle du modèle cumulatif. Le modèle adjacent a pour avantage d'être toujours défini contrairement au modèle cumulatif. En effet, si un design non proportionnel est considéré dans l'analyse, la contrainte des paramètres liés aux catégories de réponse peut être délicate à vérifier pour ce dernier. Le modèle adjacent est moins contraignant et permet donc une plus grande souplesse pour une analyse double explicative.

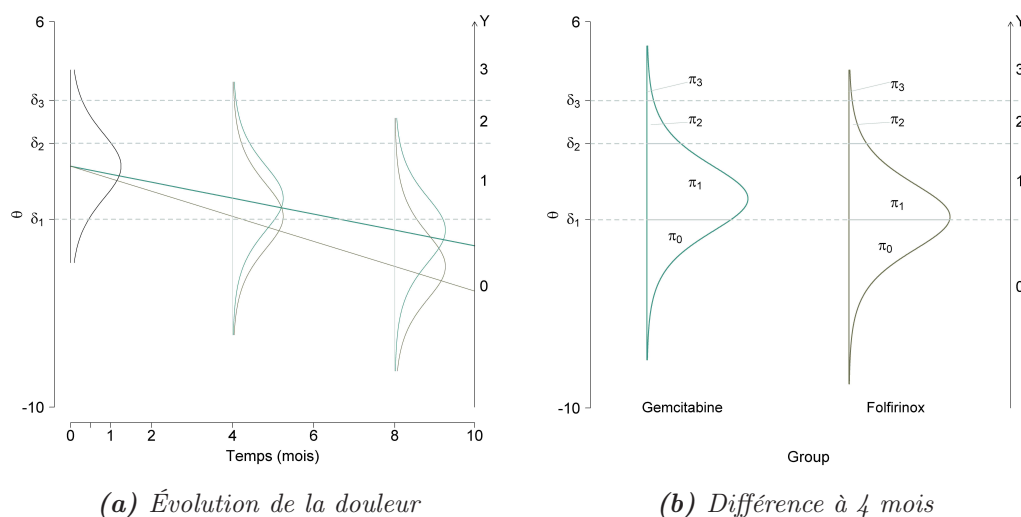


FIGURE 3.2: Visualisation graphique des résultats du modèle cumulatif pour la dimension de douleur sur l'essai Pordige4/Accord11. Lien entre la variable latente θ et la variable réponse ordinaire Y est rendu possible grâce aux paramètres de difficulté (ceux de l'item 9) : $\delta_1 = -2.2$, $\delta_2 = 0.95$ et $\delta_3 = 2.73$.

Intéressons-nous à présent au choix de F . Parmi les fonctions les plus utilisées, considérons les fonctions symétriques et non symétriques. Aucun modèle IRT n'a été présenté suivant une fonction non symétrique. En effet, nous allons expliquer par la suite pourquoi est-il avantageux de considérer une fonction symétrique dans notre contexte. Considérons quatre CdF différentes : les distributions logistique et gaussienne (les deux CdF symétriques les plus connues), et les distributions Gumbel min et Gumbel max (asymétriques). Les deux dernières sont définies telles que : $F(u) = \exp(-\exp(-u))$ pour la distribution Gumbel max et $F(u) = 1 - \exp(-\exp(u))$ pour la Gumbel min.

Nous avons étudié l'impact de la CdF sur le choix d'un modèle IRT (cumulatif ou adjacent) sur des données réelles (les données de HRQoL de l'essai Prodiges 4/ACCORD11 ont été utilisées) pour six dimensions de HRQoL. Le prédicteur linéaire utilisé est le même pour tous les modèles et correspond à celui donné par l'équation

TABLE 3.2: Estimations de -2ℓ associées aux différents modèles adjacents et cumulatifs et aux différentes CdF. Six dimensions de HRQoL sont utilisées : statut global de santé (GHS), fonction physique (PF2), fonction cognitive (CF), fatigue (FA), douleur (PA) et diarrhée (DI). Pour les cases vides, le modèle n'a pas convergé.

Famille de modèles	CdF F	Dimensions de HRQoL					
		GHS	PF2	CF	FA	PA	DI
Modèles Adjacents	Logistique	6633.2	9319.2	3826.5	6930.0	5008.8	2238.2
	Gaussienne	6651.4	9315.7	3823.5	6948.4	5012.0	2237.8
	Gumbel min	-	9400.5	3834.5	-	5014.6	2239.8
	Gumbel Max	6643.2	9299.4	3819.7	6927.0	5015.6	2239.2
Modèles Cumulatifs	Logistique	6558.1	9323.5	3818.0	6913.4	4998.0	2220.5
	Gaussienne	6628.0	9331.1	3812.5	6947.0	5009.4	2227.8
	Gumbel min	6642.2	9500.6	3835.4	6990.0	5004.9	2216.5
	Gumbel Max	6553.9	9309.8	3821.0	6924.9	5031.9	2246.8

(2.10) sans la prise en compte de la covariable liée au groupe. Étant donné le même prédicteur linéaire pour tous les modèles IRT testés, la sélection de la CdF la plus appropriée aux vues des données est basée sur la valeur de la vraisemblance. Les analyses ont été réalisées avec le logiciel SAS (version 9.3) (Institute, 2011; Boeck and Wilson, 2004). La Table 3.2 renvoie les différentes valeurs de -2ℓ (ℓ représentant la log-vraisemblance) selon le modèle (adjacent ou cumulatif) et la CdF utilisés. Nous remarquons que pour les dimensions de fatigue et de douleur, la CdF logistique est préférée pour le modèle cumulatif, alors que la CdF Gumbel max semble plus adaptée pour le GHS et la dimension physique, la CdF gaussienne pour la fonction cognitive et la CdF Gumbel min pour la diarrhée. Le choix de la CdF est également différent pour les modèles adjacents suivant les dimensions.

Quelle que soit F , l'information apportée par les effets fixes est la même. Le choix de F découle essentiellement des données analysées afin de les ajuster au mieux. Cela peut avoir un intérêt pour des objectifs de prédiction, mais pour notre étude, cela ne s'avère pas primordial. De plus, nous avons vu l'intérêt d'utiliser une CdF symétrique afin que le modèle soit invariant sous la permutation inverse. Les distributions non symétriques ne sont donc pas considérées par la suite et les CdF Logistique et gaussienne sont toutes deux favorisées.

Enfin, nous n'avons pas discuté dans les modèles IRT du fait que différentes réponses aux items pouvaient être en adéquation avec différentes CdF, étant donné que ce sont des réponses aléatoires différentes. La différenciation des CdF se situe sur la pente en fonction de θ (voir Figure 3.3a). En IRT, cette distinction entre les items est réalisée par l'intermédiaire des paramètres de discrimination (voir Figure 3.3b), qui à partir d'une CdF symétrique choisie pour la modélisation, joueront sur l'inclinaison de la pente. Nous retrouvons bien la construction astucieuse des modèles IRT où les paramètres d'item permettent de distinguer la variable réponse selon les items afin d'associer à chaque item un modèle qui lui est spécifique.

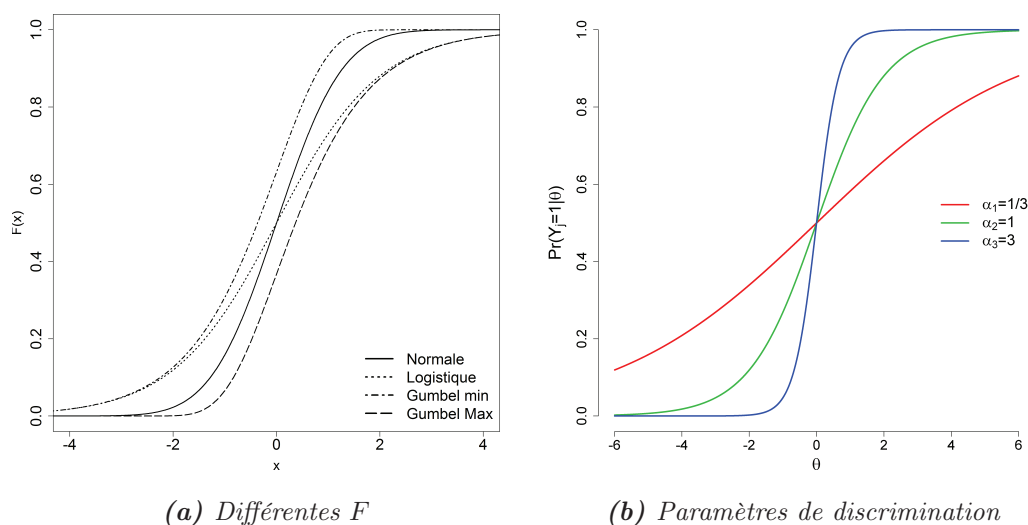


FIGURE 3.3: Liaison entre la CdF et le paramètre de discrimination. La Figure 3.3a montre les différentes pentes des F considérées et la Figure 3.3b montre les différentes pentes suivant les différentes valeurs du paramètre discrimination de la CdF logistique.

3.4 Discussion

La présentation des modèles IRT d'un point de vue des GLMM pour données ordinales est originale. Les termes employés ne sont pas les mêmes, et faire le lien entre ces deux approches permet une meilleure compréhension des modèles. En IRT, la paramétrisation spécifique du prédicteur linéaire est une manière astucieuse de considérer une seule variable réponse catégorielle, où la modélisation est personnalisée pour chaque item par l'intermédiaire de paramètres spécifiques à l'item. Pour l'analyse longitudinale de la HRQoL, nous recommandons les modèles IRT appartenant aux familles de modèles adjacents ou cumulatifs, associés à une CdF symétrique. Le choix entre ces deux familles est ensuite discutable. En termes d'interprétation directe des différents effets introduits dans le modèle de régression, les modèles cumulatifs sont les plus appropriés. En effet, la considération de la variable latente sous-jacente au modèle est plus explicite et permet d'analyser le concept étudié de la même manière que pour un LMM. Cela est un avantage certain pour présenter la méthodologie et les résultats aux cliniciens. Dans le cas de l'analyse longitudinale de la HRQoL dans les essais cliniques, seuls les modèles à rapport de cotes proportionnelles sont considérés. Ainsi, vérifier la contrainte des seuils ordonnés ne pose pas de problème. Nous recommandons alors l'utilisation du modèle cumulatif.

Le choix d'une fonction de répartition F symétrique est guidé par un argument pratique qui découle de la propriété de réversibilité. L'aspect pratique réside dans le fait de renverser l'ordre de l'échelle des items fonctionnels sans modifier le modèle sous-jacent. La seule conséquence à ce renversement est que le signe des effets fixes du modèle est opposé. L'interprétation des résultats est alors identique à celle des résultats obtenus sur les scores HRQoL, à savoir : une augmentation de la capacité

fonctionnelle renvoie à une amélioration de la HRQoL, alors qu'une augmentation de la variable symptomatique renvoie à une détérioration de la HRQoL. Ensuite, le choix de la distribution symétrique (logistique, gaussienne) n'est pas primordial. En effet, l'ajustement est réalisé par l'intermédiaire des paramètres de discrimination. Concernant ces paramètres, les résultats de la Table 3.2 montrent qu'il serait intéressant de les considérer comme différents selon la dimension de HRQoL. Par ailleurs, l'utilisation de la même CdF pour l'ensemble des analyses est recommandée.

Un frein à mentionner dans l'utilisation de modèle IRT est l'aspect computationnel. Leur temps de calcul et la disponibilité de programmes limitent leur utilisation. Bien qu'ils soient des GLMM, le nombre de données et de paramètres est bien plus conséquent qu'un GLMM classique pour données ordinales. L'estimation de ces GLMM passe par la maximisation de la vraisemblance (algorithmes EM, Newton Raphon, Marquardt et autres) ainsi que par des méthodes numériques d'intégration (méthodes de quadrature, méthode de Monte-Carlo, et autres). Pour des modèles IRT pour données polytomiques, la fonction de vraisemblance à maximiser est :

$$\mathcal{L}(\lambda; y) = \prod_{i=1}^n \int \prod_{v=1}^{n_i} \prod_{j=1}^J \prod_{m=1}^{M_j} \pi_{ivm}^{(j)}(\theta_{iv}, \delta_j)^{y_{ivm}^{(j)}} \phi(\xi_i | \Sigma) d\xi_i \quad (3.15)$$

où λ représente l'ensemble des paramètres et $\pi_{ivm}^{(j)}(\theta_{iv}, \delta_j)$ défini par (3.12) ou (3.8) pour respectivement un modèle adjacent ou un modèle cumulatif. Cela rajoute le produit sur les J items qui n'est pas le cas pour les GLMM classiques. L'estimation est donc bien moins aisée que pour l'approche CTT. Pour l'étude d'une population homogène avec un suivi relativement court (classiquement observé dans les essais cliniques), les conclusions ne sont pas différentes entre les deux approches (Anota et al., 2014a; Barbieri et al., 2015). L'approche CTT par LMM présente alors des avantages d'un point de vue pratique qui ne sont pas négligeables, malgré des hypothèses du modèle peu adaptées aux scores de HRQoL.

3.5 Article : Selection of item response theory models for the longitudinal analysis of health-related quality of life in cancer clinical trials

Item response models for the longitudinal analysis of health-related quality of life in cancer clinical trials

Antoine Barbieri^{*,1,2,3}, Jean Peyhardi^{1,4,5}, Thierry Conroy^{6,7},
Sophie Gourgou², Christian Lavergne^{3,8} and Caroline Mollevi^{2,9,10}

Abstract

Statistical researches regarding health-related quality of life is a major challenge to better evaluate the impact of the treatments on their everyday life and to improve patients' care. In the literature, the mixed models based on the item response theory are proposed to analyze directly HRQoL data from the questionnaires given to the patients. First, we used a recent classification of generalized linear models for categorical data to discuss about a selection of item response models for the longitudinal analysis of health-related quality of life in cancer clinical trials. Through methodological and practical arguments, the adjacent and cumulative models seem particularly suitable for this specific context. Then, a simulation study is carried out to compare the linear mixed model classically used to the discussed item response models on the random part of mixed models. Finally, we proposed a longitudinal analysis of HRQoL data from cancer clinical trial using an item response cumulative model.

Keywords: Health-related quality of life; Item response theory; Mixed models; Ordinal categorical data; Longitudinal analysis.

1 Introduction

Endpoints refer to biological and clinical measures to assess the efficiency of new therapeutic strategies. The overall survival endpoint is the gold standard to show a clinical benefit of these strategies and treatments. Therapeutic treatments being more efficient and increasing the patients' lifetime, the overall survival endpoint may become insufficient to show a significant difference between two treatments. It is then necessary to consider a longer follow-up or a larger cohort of patients to have a sufficient number of events and a good statistical power (Fiteni et al., 2014), both representing considerable costs. Thus, to conclude to the benefit of a new treatment, other endpoints have emerged and the health-related quality of life (HRQoL) is currently one of the most important. In cancer clinical trials, the patient-reported

*Corresponding author: Antoine.Barbieri@umontpellier.fr

¹Université de Montpellier, Montpellier, France

²Institut régional du Cancer Montpellier (ICM) - Val d'Aurelle, Biometrics Unit, Montpellier, France

³Institut Montpelliérain Alexander Grothendieck (IMAG), Montpellier, France

⁴CIRAD, AGAP and Inria, Virtual Plants, France

⁵Institut de génomique fonctionnelle, Montpellier, France

⁶National Quality of Life in Oncology Platform, France

⁷Institut de Cancérologie de Lorraine, Nancy, France

⁸Université Paul-Valéry Montpellier 3, Montpellier, France

⁹IRCM, Institut de Recherche en Cancérologie de Montpellier, Montpellier, France

¹⁰INSERM, U1194, Montpellier, France

outcomes are increasingly used to analyze a clinical benefit for medical decision-making (Fiteni et al., 2014). The HRQoL endpoint may seem more pertinent to show the interest of a new therapy in some cases such as the palliative or geriatric situations. However, there are conceptual and methodological brakes underlying to the concept and the assessment of HRQoL. Indeed, HRQoL is a multidimensional concept regarding the physical, psychological and social functions as well as symptoms associated with the disease and treatments. Another conceptual brake is the subjectivity of its measurement. Indeed, patients report their feelings about their HRQoL thanks to self-reported questionnaires. Both arguments preclude the use of HRQoL as sole primary endpoint in clinical trials.

In oncology, HRQoL is assessed using a general questionnaire for a set of different cancers, and an additional specific questionnaire associated with each type of cancer (Aronson et al., 1993; Cella et al., 1993). Each questionnaire decomposes the HRQoL to measure several under-concepts (dimensions of HRQoL) which themselves comprise one or several items. The items are built on the Likert scales in which the response variable is ordinal. Thus, considering several items for a given dimension, HRQoL data are composed of multiple ordinal responses. Also, the questionnaires are filled by the subjects themselves, and collected at different times defined in the trial protocol (usually at inclusion, during treatment and follow-up). These repeated measures are used to assess the evolution of the subject's HRQoL over time. In Europe, these questionnaires are developed and validated by the European organization for research and treatment of cancer (EORTC). The standard questionnaire currently used in oncology is the EORTC Quality of Life Questionnaire - Core 30 (EORTC QLQ-C30) (Aronson et al., 1993), together with the scoring procedure proposed by the EORTC (Fayers et al., 2001). The score is then calculated for each dimension and for each subject, corresponding to the average of the item responses for a single dimension, and expressed on a scale ranging from 0 to 100. The interpretation is such that high functional scores reflect good functional capacities and a good HRQoL level, and conversely, high symptomatic scores represent strong symptoms and point out difficulties. The use of scoring procedures is common in practice because the statistical methods for quantitative variables are more powerful and easier to implement and interpret (Grilli and Rampichini, 2011). But, in a Likert scale, the gap which separates each adjacent category of response ("not at all", "a little", "quite a bit" and "very much") may not be the same, and the HRQoL score calculation does not take into account this characteristic. Another drawback in the score use is that subjects could have different item outcomes and obtain the same score. In this situation, the score does not make a distinction between these subjects (Gorter et al., 2015).

The longitudinal statistical models classically used in oncology are performed on the summary score through using the linear mixed models (LMM) or time-to-event models (Anota et al., 2014). In the LMM, the variable associated with the HRQoL score is considered as a Gaussian variable while it presents the characteristics of an ordinal variable, being non-continuous and bounded. These models allow to take into account the correlation introduced by repeated measurements on the same patient (collection of the HRQoL questionnaires over time) and different covariates such as time, treatment group, age... However, the use of the LMM for HRQoL analysis is scientifically questionable because the score is not continuous. Also, the LMM do not consider the bounded characteristic of the HRQoL score, ranged between 0 and 100. Moreover, many symptomatic dimensions are composed of only one item, the HRQoL score has exactly the same properties than ordinal categorical data, and using the LMM is not appropriated. Thus, if a ceiling or floor effect is observed, the categorical feature is even more marked when one of the two extreme categories is over-represented. The second approach for the longitudinal analysis of HRQoL is based on the time-to-event models: the time-to-deterioration (TTD) and the time-until-definitive-deterioration (TUDD) (Hamidou et al., 2011). Survival approaches are often used and

thus well-known in the oncologic field, and are appreciated for their easiness to interpret result and their good understanding by clinicians. In these models, an event is classically defined by the (definitive or not) deterioration of the HRQoL score between baseline and a follow-up time, given a minimally clinically important difference (Anota et al., 2013). The lack of homogeneity of the methods used for the HRQoL data analyses in different oncology clinical trials is also a real obstacle to the comparison of results. Indeed, the LMM and TUDD approaches show results which may sometimes seem contradictory, and with different interpretations, but they may also be complementary. An example can be taken comparing two similar cancer clinical trials investigating the effect of bevacizumab. In the first trial (Chinot et al., 2014), HRQoL analysis through TUDD showed that the bevacizumab group had a later deterioration of HRQoL compared with patients in the standard group. Conversely, in the second trial (Gilbert et al., 2014), HRQoL analysis using the LMM showed a worse HRQoL overtime in the bevacizumab group.

Interest in the HRQoL endpoint is growing rapidly in cancer clinical trials and it is essential to use a suitable methodology to analyze HRQoL data, taking into account the data properties (repeated measurements of the ordinal multiple responses). In our study, we first focused on the different and most adapted models to analyze HRQoL from raw data, *i.e.* directly on the item responses. Studies on psychometric properties from questionnaires such as the one used for HRQoL have been ongoing for a long time (Edelen and Reeve, 2007; Jafari et al., 2012), known as the item response theory (IRT). The IRT models link the individual's item responses and the latent variable which represents the studied HRQoL concept. They are generalized linear mixed models (GLMM) for ordinal data with a particular parameterization of the linear predictor. The interest for this kind of model to analyze the data, including the longitudinal analyzes, is increasing (Titman et al., 2013; Hardouin et al., 2015; Gorter et al., 2015). However, to our knowledge, there is no work that discusses of the choice of one of the different IRT models over the others for HRQoL longitudinal analysis. We propose in section 2 a conceptual selection of these models through practical and methodological arguments. For this selection, we replaced IRT models in the general linear mixed models (GLMM) context using the new specification of generalized linear models (GLM) for categorical responses, proposed by Peyhardi et al. (2015). Then, as some comparison simulations have already been performed between the IRT models and the LMM on their capacity to detect fixed effects, we focus in section 3 on the sensibility of these models to detect the random effects. Finally, section 4 presents an application of the chosen IRT model on real data from a multicenter randomized phase III clinical trial in first-line metastatic pancreatic cancer patients (Conroy et al., 2011).

2 Conceptual selection of IRT models for the longitudinal analysis of HR-QoL in cancer clinical trials

HRQoL raw data are repeated measurements of ordinal multiple responses. The GLMM for ordinal responses seem well suitable to analyze this kind of data. The use of random effects takes into account the inter-patient variability and the correlation between the repeated measurements for each single patient. The IRT models have been increasingly used to analyze health data deriving from self-questionnaires made of polytomous responses (Hardouin et al., 2012; Anota et al., 2014; Barbieri et al., 2015). These models turn out to be GLMM for polytomous data with a specific parameterization of the linear predictor taking into account the multiple outcomes. For ordinal responses, three families of regression models are described: the families of adjacent models (Masters, 1982; Agresti, 2010), cumu-

lative models (Samejima, 1968; McCullagh, 1980), and sequential models (Tutz, 1990; Fahrmeir and Tutz, 2001). Many IRT models are proposed for the analysis of this kind of data, often with no explanation regarding the choice of one model over another. In this section, we use the new specification of the general linear models (GLM) for ordinal data proposed by Peyhardi et al. (2015) to talk about the relevance of the models adopted in the context of the HRQoL longitudinal analysis in cancer clinical trials. Each model is defined according to three components (r, F, Z): the ratio of probabilities (r), the cumulative distribution function (F), and the parameterization of the linear predictor determined by the design matrix (Z). For the GLMM framework, we extended this new specification to the quadruplet (r, F, Z, U) with Z the design matrix of fixed effects and U the design matrix of random effects. The relationship between these components is determined by $R = \mathcal{F}(Z\beta + U\xi)$. Given the linear predictor $\eta = Z\beta + U\xi$ and $\pi_{iv}^{(j)} = (\pi_{iv0}^{(j)}, \dots, \pi_{iv, M-1}^{(j)})$ the truncated vector¹ of conditional probabilities with $\pi_{im}^{(j)} = \Pr(Y_{ij} = m | \xi_i)$ the conditional probability that subject i ($i = 1, \dots, n$) selects the category $m \in \{0, \dots, M_j\}$ for item j ($j = 1, \dots, J$) at visit v ($v = 1, \dots, V$) given individual random effect, we defined:

- $R = \left\{ r_m \left(\pi_{iv}^{(j)} \right) \right\}_{i,j,v,m}$;
- $\mathcal{F} \left\{ \left(\eta_{ivm}^{(j)} \right)_{i,j,v,m} \right\} = \left\{ F \left(\pi_{ivm}^{(j)} \right) \right\}_{i,j,v,m}$ where $\eta = \left(\eta_{ivm}^{(j)} \right)_{i,j,v,m}$.

After a presentation of the IRT parameterization used concerning the linear predictor, we compare different polytomous IRT models on the basis of the link function (ratio of probabilities and the cumulative distribution function (CdF)) using methodological and practical arguments.

2.1 The IRT parameterization of the linear predictor

The IRT probabilistic models emerged following the works of Georg Rasch (Rasch, 1960) on dichotomous responses, and were then extended to ordinal responses. Considering the three families of adjacent, cumulative and sequential models, there are three associated famous IRT models (Boeck and Wilson, 2004; Bacci et al., 2012), respectively the graded response model (Samejima, 1968), the (generalized) partial credit model (Masters, 1982; Muraki, 1992), and the sequential model (Tutz, 1990). These models link the individual's item responses to the unidimensional latent variable which represents a concept not directly measurable. In oncology framework, the concept is HRQoL relatively to one specific HRQoL dimension.

From the IRT, the specific parameterization of the linear predictor $\eta_{im}^{(j)}$ is built into two parts: the individual part and the item part. The best-known way is to consider the following decomposition:

$$\eta_{im}^{(j)} = \alpha_j (\theta_i - \delta_{jm}), \quad (1)$$

where θ_i is associated with an unidimensional random variable (currently assuming to be standard normal for identifiability) representing the latent value for the subject i , δ_{jm} and α_j being the item parameters. Generally called difficulty parameter, δ_{jm} is the intercept (or threshold) associated with the item j for the category $m \in \{1, \dots, M_j\}$. The parameter α_j is called the discrimination parameter of item j , and

¹This truncated vector is sufficient to characterize the categorical distribution since $\pi_{ivM}^{(j)} = 1 - \sum_{c=0}^{M-1} \pi_{ivc}^{(j)}$;

represents the sensitivity of each response probability according to the value of the latent trait. Indeed, the more the discrimination parameter value is high, the more the item allows well discriminating two individuals with a close latent trait value. However, the predictor is no longer linear for IRT models using discrimination parameters because it includes a product of parameters. Thus, these models do not belong to the class of GLMM (Liu and Hedeker, 2006).

In oncology, HRQoL analysis is carried out using IRT models which do not include the discrimination parameters (fixed to one for all items). Thus, these IRT models are within the class of GLMM. Concerning the longitudinal analysis, several studies proposed to extend some IRT models using the linear decomposition of the latent variable θ with fixed and random effects (Hardouin et al., 2012; Verhagen and Fox, 2013; Huber et al., 2013):

$$\theta_{iv} = x'_{iv}\beta + u'_v\xi_i, \quad (2)$$

with the vector β is associated with the fixed effects, the vector ξ_i with the subject-specific random effects and the index v the current visit. In the equation (2), θ_{iv} is thus the estimation of latent processus at the visit v .

2.2 The probability ratio: structure of the models

The linear predictor is not directly related to the response probability but to a particular transformation ratio. The choice of the ratio is related to the nature of response from the ordering assumption among categories. Thus, reference ratio (Peyhardi et al., 2015) for nominal response is excluded in this work because the HRQoL responses are ordinal. Let's consider the simple situation from GLM with one item with $(M + 1)$ response categories. The three model families for ordinal data are distinguished by the choice of the ratio of probabilities $r(\boldsymbol{\pi}) = (r_0(\boldsymbol{\pi}), \dots, r_{M-1}(\boldsymbol{\pi}))$. Each model is summarized by M equations $\{r_m(\boldsymbol{\pi}) = F(\eta_m^*)\}_{m=0, \dots, M-1}$ with $\eta_m^* = \delta_m - \theta$, highlighting the decomposition of the link function which is determined through the ratio of probabilities and the CdF. Indeed, we may distinguish different ratios of probabilities for these different families, respectively, for the cumulative models,

$$r_m(\boldsymbol{\pi}) = \pi_0 + \dots + \pi_m, \quad m = 0, \dots, M - 1; \quad (3)$$

for the adjacent models,

$$r_m(\boldsymbol{\pi}) = \frac{\pi_m}{\pi_m + \pi_{m+1}}, \quad m = 0, \dots, M - 1; \quad (4)$$

and, for the sequential models,

$$r_m(\boldsymbol{\pi}) = \frac{\pi_m}{\pi_m + \dots + \pi_M}, \quad m = 0, \dots, M - 1.$$

In the IRT, adjacent and cumulative models are usually presented given the reverse permutation (Samejima, 1968; Masters, 1982; Bacci et al., 2012). This permutation is defined as the reversal of category order (McCullagh, 1980). Assuming that the considered CdF is symmetric (i.e. the corresponding probability density function is symmetric about the y-axis), these models are invariant under this permutation (Peyhardi et al., 2015). For our application context, this is as an advantage for result interpretation. A lower item-response category reflects a lower level of the symptomatic dimensions whereas it represents a higher level of capacity for the functional dimensions. The reverse permutation

for the functional dimensions makes it easier and intuitive for clinicians to present their results. This allows the homogenization of the result interpretation as it is done in the scoring procedure proposed by the EORTC (for functional dimensions, the score scale is reversed compared with the item responses categories order) (Fayers et al., 2001). However, sequential models correspond to process ordering and reversing the process may change its nature. These models are not reversible (i.e. no invariant under the reverse permutation). Thus, sequential models will not be used and only the adjacent and cumulative models which correspond to scale ordering as used for HRQoL measurements, will be consider.

The cumulative models also have additional properties (McCullagh, 1980), including that they are invariant when successive categories are gathered. Thus, if one category is not observed, it can be combined with its successive categories without changing the model. Another advantage of the cumulative models is their interpretation through a continuous latent variable. Indeed, the continuous latent variable \tilde{Y} underlying the model exists and a direct link with the response variable Y through the thresholds presumed to be strictly increasing ($-\infty = \delta_0 < \delta_1 < \dots < \delta_M < \delta_{M+1} = +\infty$) is such as:

$$\{Y = m\} \quad \text{if} \quad \left\{ \delta_m < \tilde{Y} \leq \delta_{m+1} \right\}, \quad m = 0, \dots, M,$$

where $\tilde{Y} = \theta + \varepsilon$ and ε is the error term distributed following the CdF. Here, the latent variable \tilde{Y} represents HRQoL and its interpretation is then equivalent to the interpretation of the response variable using a LMM.

However, an advantage of the adjacent models is that there is no constraint affecting the model estimation. Nonetheless, the cumulative models have to respect constraints, which can make difficult the model estimation, particularly in the case of non-proportional design of linear predictor (Peyhardi et al., 2015). For the proportional design, a common slope (θ) is considered for all categories, else the slope is dependent of the category (θ_m). Let the simple parameterization of the linear predictor $\eta_m = \theta_m - \delta_m$ for $m \in \{1, \dots, M\}$ where θ_m and δ_m are the slope and the intercept associated with the category m , respectively. Considering proportional design ($\theta = \theta_1 = \dots = \theta_M$), the cumulative models refer to the principle of thresholds (McCullagh, 1980; Hedeker and Gibbons, 1994) with the constraint they have to be strictly increasing such as $-\infty < \delta_1 < \dots < \delta_M < +\infty$. Considering the non-odd proportional models, the constraint then becomes $-\infty < \eta_M < \dots < \eta_1 < +\infty$ which is more difficult to verify. For the longitudinal analysis of HRQoL data in oncology, the proportional design is considered and to verify the constraint only on the threshold is easier.

Table 1 summarizes some characteristics of these three families of models which are important for the longitudinal analysis of HRQoL in cancer clinical trials. In this context, a proportional design of the linear predictor is preferred. Under this parameterization, there is few difficulties to respect the cumulative models constraints and to estimate them. The adjacent models seem more flexible than cumulative models because they are always define for all linear predictor. But, their interpretation of the results using the cumulative model is more intuitive than adjacent models.

2.3 The cumulative distribution function

The latest component discussed in the IRT model selection is the CdF. Each model probability can be defined with any CdF. As commonly seen in the IRT models, and thank to the reversibility property, the adjacent and cumulative models are used in descending order. For the cumulative model, the probabilities

Table 1: Summary of the characteristics for the three model families

Characteristics	Models		
	Adjacent	Cumulative	Sequential
Reversibility	<i>yes</i>	<i>yes</i>	<i>no</i>
Interpretation using the latent variable	<i>no</i>	<i>yes</i>	<i>yes</i>
Always defined	<i>yes</i>	<i>yes</i> (<i>no</i> ¹)	<i>yes</i>

¹: for some non proportional design models

are defined from the equation (3) and given the F as:

$$\begin{cases} \pi_0 &= 1 - F(\eta_1) \\ \pi_m &= F(\eta_m) - F(\eta_{m+1}), \quad m = 1, M - 1 \\ \pi_M &= F(\eta_M) \end{cases} \quad (5)$$

Then, general expression of the sequential model whatever the CdF used is such that

$$\pi_m = F(\eta_m) \prod_{k=1}^{m-1} \{1 - F(\eta_k)\},$$

where $m = 1, \dots, M$ and $\prod_{k=1}^0 \{.\} = 1$ (Fahrmeir and Tutz, 2001).

Such general equations have never been presented for adjacent models, only described with the logistic CdF (Masters, 1982; Muraki, 1992; Fahrmeir and Tutz, 2001; Agresti, 2010; Hardouin et al., 2012; Anota et al., 2014). However, the different response probabilities can be presented from the adjacent ratio and according to a general CdF (F):

$$\begin{cases} \pi_0 = \frac{1}{1 + \sum_{m=1}^M \prod_{k=1}^m \left(\frac{F(\eta_k)}{1 - F(\eta_k)} \right)} \\ \pi_m = \frac{\prod_{k=1}^m \left(\frac{F(\eta_k)}{1 - F(\eta_k)} \right)}{1 + \sum_{m=1}^M \prod_{k=1}^m \left(\frac{F(\eta_k)}{1 - F(\eta_k)} \right)}, \quad m = 1, \dots, M \end{cases} \quad (6)$$

The CdF choice is especially used to best fit the data. Let's four CdF from two different kinds: the most commonly used symmetric distributions, the logistic and Gaussian distributions, and the two asymmetric distributions, the Gumbel min and Gumbel max distributions. The two later distributions are respectively defined by $F(\eta) = \exp(-\exp(-\eta))$ for the Gumbel max distribution and by $F(\eta) = 1 - \exp(-\exp(\eta))$ for the Gumbel min distribution.

Figure 1a shows different slopes depending on the particular CdF. The CdF allows to take into account the influence of linear predictor (η) change on the response probability evolution. In general IRT parameterization (equation 1), the slope adjustment is managed by the discrimination parameter. Depending on different discrimination parameter values, Figure 1b presents the CdF logistic according to the individual latent variable. This item parameter has the task of fitting the CdF for each considered item to distinguish more the different response variable.

In the literature, the cumulative model is presented according to the use of several of the previously mentioned CdF (Samejima, 1968; Fahrmeir and Tutz, 2001; Liu and Hedeker, 2006), while the adjacent

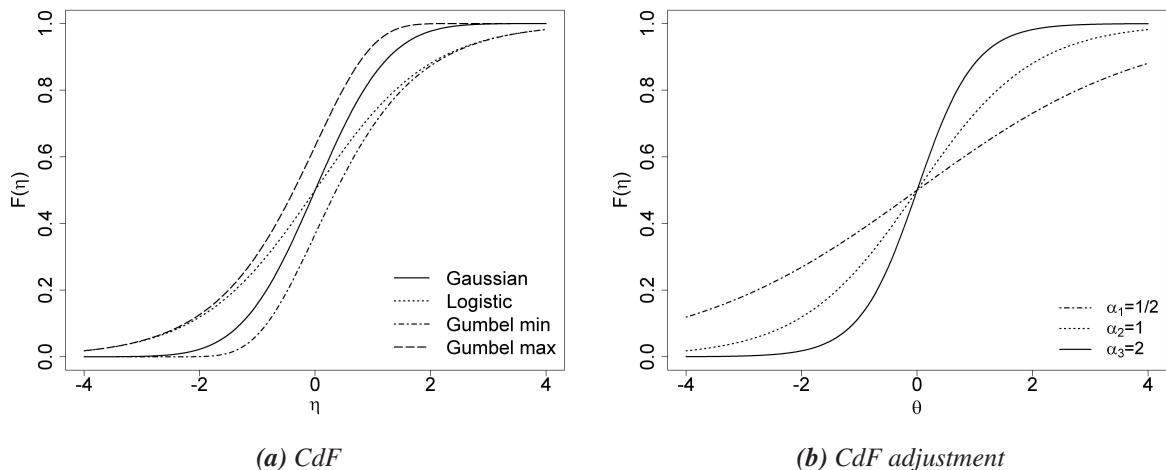


Figure 1: Relationship between the Cdf and the IRT parameterization for dichotomous items where $F(\eta) = F(\alpha_j(\theta - \delta_j))$. Figure 1a presents the different Cdf for one item j given $\alpha_j = 1$ and $\delta_j = 0$. Figure 1b presents the logistic Cdf adjustment for three items ($j = 1, 2, 3$) with different α_j and $\delta_j = 0$, given the linear predictor defined in equation (1).

models are most often presented with the logistic Cdf. However, for both, we recommend the use of a symmetric Cdf. As seen in Figure 2b, the IRT parametrization is a subtle and effective way to take into account the multiple item outcome characteristics in GLMM for categorical responses. In the context of the HRQoL in clinical trial, the HRQoL dimension considers a small set of items which are correlated and measure an unique latent variable. The discrimination parameter is routinely not use in this kind of analysis.

Relatively to the literature, Table 2 shows the specification of the famous polytomous IRT models following the different components. For IRT models being within the class of GLMM, we proposed to define them with the four components (r, F, Z_q, U_r) . The kind of considered location item parameters can be indicated by the index q where $q = 1$ for including only difficulty parameters. Let $q = 2$ for considering the rating scale model (Andrich, 1978) parameterization where the difficulty parameters are common for all items and one shift parameter is considered for each item. Regarding the random part, the number of random effects can be then indicate by the index r . For the classical IRT parametrization presented in Table 2, only one random effect ($r = 1$) is taken into account : the capacity parameter θ . For IRT models including discrimination parameters for each item, we proposed to replace the components Z and U by a component specifying that the predictor is no longer linear (nl), such as (r, F, nl) .

3 Simulation study

In the previous section, we focused on the use of the mixed models for ordinal data analysis and their relevance in the HRQoL analysis in oncology was discussed. Some comparisons studies exist between these different approaches (Blanchin et al., 2011; Anota et al., 2014), mainly on the fixed part of the

Table 2: Specification of the famous IRT model following the components : (r, F, Z_p, U_r) for the GLMM and (r, F, nl) for IRT model with no longer linear predictor. Index p denotes the number of kind of item parameters considered in the IRT model and r the number of random effect.

IRT models	$\eta_{im}^{(j)}$	(r, F, Z_p, U_r)
Rating scale model	$\theta_i - (\delta_m + \tau_j)$	<i>(adjacent, logistic, Z_2, U_1)</i>
Partial credit model	$\theta_i - \delta_{jm}$	<i>(adjacent, logistic, Z_1, U_1)</i>
Sequential Rasch model	$\theta_i - \delta_{jm}$	<i>(sequential, logistic, Z_1, U_1)</i>
Graded response model	$\alpha_j (\theta_i - \delta_{jm})$	<i>(cumulative, logistic, nl)</i>
Generalized partial credit model	$\alpha_j (\theta_i - \delta_{jm})$	<i>(adjacent, logistic, nl)</i>

mixed models. Anota et al. (2014) had shown an equivalent capacity to detect a fixed effect for the LMM and for one of the IRT models. Indeed, even if the LMM take into account the HRQoL score, which is a summary variable, this approach is at least equivalent to the IRT models in terms of power.

In this simulation study, the adjacent and cumulative models with the same parameterization of the linear predictor and the logistic CdF were used (as usually in the IRT models). The aim of the following section is to reinforce these comparisons between the LMM and the IRT models on the random part of the mixed models. The datasets were simulated from an IRT model (adjacent and cumulative models). Regarding the parameterization, two subject-specific random effects ξ_{i0} and ξ_{i1} were considered, respectively associated with the intercept and the slope. Of course, the usefulness of the random effect introduction in the model is strongly associated with the observed data. HRQoL is a subjective endpoint, and the individual random effect ξ_{i0} is thus entirely justified. Indeed, it is easy to imagine that each patient has a different level of HRQoL at baseline. The random slope is more questionable, indeed, the assumption that the specific HRQoL evolution of one single patient diverges from the average evolution for the whole population, is less obvious than the previous one. In this section, the capacity of the mixed models to detect the slope random effect was thus studied. No group effect was considered in this simulation study.

3.1 Design

We want to study the capacity of each model to detect the random effect ξ_{i1} associated with time (random slope). The two subject-specific random effects are considered independent where $\xi_{i0} \sim \mathcal{N}(0, \sigma_0^2)$ and $\xi_{i1} \sim \mathcal{N}(0, \sigma_1^2)$. The following model choice study is performed on the basis of the Bayesian information criteria (BIC) where two models were considered: \mathcal{M}_2 with the two random effects (r, F, Z_1, U_2) and \mathcal{M}_1 excluding the random slope (r, F, Z_1, U_1) . For the IRT models, the linear decomposition of the latent trait θ_{iv} only took into account the time as a fixed effect. The two considered models with proportional design are:

$$\begin{aligned} \mathcal{M}_2 : \theta_{iv} &= (t_v - t_0) \beta_1 + \xi_{i0} + (t_v - t_0) \xi_{i1} \\ \mathcal{M}_1 : \theta_{iv} &= (t_v - t_0) \beta_1 + \xi_{i0} \end{aligned} \quad (7)$$

In order to best reflect the EORTC QLQ-C30 questionnaire, the most frequent HRQoL dimension with two items ($j = 1, 2$) comprising four response categories ($m \in \{0, \dots, M\}$ with $M = 3$), was used to design the simulation study. A sample size of three hundred subjects ($i = 1, \dots, n$ with $n = 300$)

and eight follow-up time ($v = 0, \dots, 7$), as for the trial presented in the previous section, were considered. The datasets were simulated from a multinomial distribution. The different response probabilities $\left\{ \pi_{ivm}^{(j)} = \Pr \left(Y_{iv}^{(j)} = m | \theta_{iv}, \delta_j \right) \right\}$ concerning the subject i for item j were determined by equation (6) for the adjacent model and by equation (5) for the cumulative model, given:

- the item parameters $(\delta_{j1}, \delta_{j2}, \delta_{j3})_{j=1,2}$;
- the latent trait (θ_{iv}) deduced in accordance with equation (7);
- the logistic CdF,

$$F \left(\eta_{ivm}^{(j)} \right) = \frac{\exp \left(\eta_{ivm}^{(j)} \right)}{1 + \exp \left(\eta_{ivm}^{(j)} \right)},$$

where $\eta_{ivm}^{(j)} = \theta_{iv} - \delta_{jm}$.

The values of the parameters used were deduced from the pain symptom data of the clinical trial presented in the previous section. We considered two kinds of difficulty parameters: near $\delta^{ne} = (\delta_1^{ne}, \delta_2^{ne})$ and far $\delta^{fa} = (\delta_1^{fa}, \delta_2^{fa})$. These parameter values were chosen in order to illustrate several scenarios described in Table 3. The different scenarios were due with the different associations between the model used to simulate the data, $(adjacent, logistic, Z_1, U_r)_{r=1,2}$ or $(cumulative, logistic, Z_1, U_r)_{r=1,2}$, and the different considered values of the difficulty parameters. Table 3 shows the simulated responses expected at baseline ($t = 0$). The responses simulated across time depended of the considered coefficient β_1 . Each scenario was simulated $N = 500$ times.

Table 3: Values of difficulty parameters used to simulate the data and expected responses at t_0 under each studied scenarios.

Models	Difficulty parameters	
	$(r, F, Z_1, U_r)_{r=1,2}$	$\delta_1^{ne} = (-1.6, 1, 1.45)$ $\delta_2^{ne} = (-0.8, 1.15, 1.9)$
$(adjacent, logistic, Z_1, U_r)_{r=1,2}$	balanced responses	focus on center categories (1 and 2)
$(cumulative, logistic, Z_1, U_r)_{r=1,2}$	focus on extreme categories (0,1 and 3)	balanced responses

Concerning the LMM, the scoring procedure proposed by the EORTC was considered (Fayers et al., 2001), and the score associated with a symptomatic dimension was first calculated using the simulated data. Let the two simulated ordinal outcomes $y_{iv}^{(1)}$ and $y_{iv}^{(2)}$ concerning the individual i at the visit v , the related score was:

$$S_{iv} = \left(\frac{\sum_{j=1}^{J=2} y_{iv}^{(j)}}{2} \right) \frac{100}{M}$$

Similarly to the parameterization in equation (7), we took into account the related choice model with:

$$\begin{aligned} \mathcal{M}_2 : S_{iv} &= \beta_0^l + (t_v - t_0) \beta_1^l + \xi_{i0} + (t_v - t_0) \xi_{i1} + \varepsilon_{iv} \\ \mathcal{M}_1 : S_{iv} &= \beta_0^l + (t_v - t_0) \beta_1^l + \xi_{i0} + \varepsilon_{iv} \end{aligned}$$

where β_0^l is the fixed parameter associated with the intercept, the ξ^l are the random effects normally distributed with the mean equals to zero and $\varepsilon_{iv} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ the error term.

3.2 Results

Table 4 shows the capacity of the three models (adjacent model, cumulative model and LMM) to detect the random slope given different scenarios (Table 1). When we simulated the data under \mathcal{M}_2 according to the random effect variances estimated from real data, each model detected the random slope (ξ_{i1}) in 100% of cases whatever the different situations. On the contrary, under \mathcal{M}_1 , the simulated model \mathcal{M}_1 was correctly chosen in most cases. For all simulations under \mathcal{M}_1 , the cumulative model seemed to detect the random slope although it was not included in the simulation step. Moreover, the IRT model which was not used to simulate the data, wrongly detected this random effect given a negative value of β_1 and the difficulty parameter coefficients δ^{ne} . This could be explained by the fact that the difficulty parameters were not uniformly separated around zero and also because they were too close. Indeed, given $\beta_1 < 0$, the probabilities to observe the upper categories were very small over time and under-represented in comparison with the lower categories (as illustrated in Figure 2a). In the specific case where $\beta_1 = -0.3$, the IRT model which did not simulate the data could not explain the different outcomes only with the fixed effect and the random intercept, and it compensated the lack of information with the random slope. We then could expect symmetric results from β_1 (positive values) considering the opposite sign of the difficulty parameters because of the reversibility property of the IRT models.

On the contrary, the LMM was stable and thus allowed making the good choice of model whatever the β_1 values and the IRT model used to simulate the data. Concerning the IRT models where only one model out of the two detected the random effect ξ_{i1} , the most suitable model seemed the one not detecting this random effect.

Table 4: Frequency (on $N = 500$ datasets) of the \mathcal{M}_1 selection according to the BIC, given $t_v = (0, 1, 2, 4, 6, 8, 10, 12)$ and $\sigma_0^2 = 1.5$. For $r = 1, 2$, the (*adjacent, logistic, Z_1, U_r*) models and the (*cumulative, logistic, Z_1, U_r*) models are denoted respectively by AM and CM. For the random component, U_1 if $\sigma_1^2 = 0$ and U_2 if $\sigma_1^2 > 0$.

Simulated		Scenarios											
Model		AM using δ^{ne}			CM using δ^{fa}			CM using δ^{ne}			AM using δ^{fa}		
σ_1^2	β_1	LMM	AM	CM	LMM	AM	CM	LMM	AM	CM	LMM	AM	CM
0.2	-0.3	0	0	0	0	0	0	0	0	0	0	0	0
0.2	0.3	0	0	0	0	0	0	0	0	0	0	0	0
0	-0.5	97.67	99.29	56.49	100	94.63	92.98	100	61.33	95.71	100	99.66	89.54
0	-0.3	99.00	100	33.04	100	88.63	93.30	100	36.33	94.91	100	100	83.33
0	-0.2	100	99.62	49.28	100	94.56	93.81	100	71.67	95.77	100	99.64	79.02
0	-0.1	98.67	95.65	94.78	100	98.65	89.62	100	98.98	90.41	100	100	88.10
0	0.0	95.60	100	94.55	99.00	99.66	91.75	99.00	99.66	89.71	97.00	99.66	94.42
0	0.1	83.00	100	94.78	93.33	100	92.63	97.00	100	90.91	87.33	100	94.69
0	0.3	98.33	99.64	90.61	100	99.64	89.05	100	100	93.67	100	99.65	93.78
0	0.5	100	100	94.29	100	99.32	94.71	100	100	97.61	100	100	97.19

The capacity of the different models to detect the random slope when its variance value changes, is presented in Table 5. All models were sensitive to the signal-to-noise ratio. Indeed, the more β_1 increased, the less the random effect provided information. This was well characterized as the capacity to detect the random effect for greater variances when the signal was strong. In this case, the signal

provided the essential information explaining the different responses. In the model comparison, the LMM was less sensitive than the IRT models. Indeed, the LMM detected the random slope for a greater variance of this one whatever the β_1 value. This result was expected because the LMM is based on the HRQoL score which is a summary variable with less information than the raw data. Thus, the IRT models are more sensitive in all cases. Comparing the two IRT models, there is a tendency for the random slope model to be preferred under the cumulative model regardless of whether it is the true model or not. On the contrary, in the specific case where $\beta_1 = -0.3$, the IRT model used to simulate the data was less efficient than the other IRT model which detected a random slope to remedy the lack of information. This was coherent with our previous results shown in Table 4. Finally, the more β_1 was close to zero, the more the models detected the random slope for a low variance.

Table 5: Frequency (on $N = 500$ datasets) of the \mathcal{M}_2 selection according to the BIC, given $t_v = (0, 1, 2, 4, 6, 8, 10, 12)$ and $\sigma_0^2 = 1.5$. For the (*adjacent, logistic, Z_1, U_r*) models and the (*cumulative, logistic, Z_1, U_r*) models are denoted respectively by AM and CM. For the random component, U_1 if $\sigma_1^2 = 0$ and U_2 if $\sigma_1^2 > 0$.

Simulated		Scenarios											
Model		AM using δ^{ne}			CM using δ^{fa}			CM using δ^{ne}			AM using δ^{fa}		
β_1	σ_1^2	LMM	AM	CM	LMM	AM	CM	LMM	AM	CM	LMM	AM	CM
1	0.01	0	2.33	24.92	0	5.03	6.94	0	2.69	3.70	0.33	6.44	24.75
	0.02	0	21.40	54.67	0	37.58	44.11	0	17.73	18.12	0	50.00	77.00
	0.03	0	61.00	90.97	0	75.67	80.00	0	41.33	45.58	0	86.33	98.33
	0.05	0	97.67	99.66	0	100	100	0.33	89.00	90.00	0	99.33	100
	0.2	39.33	100	100	40.67	100	100	10.67	100	100	57.67	100	100
	0.5	100	100	100	100	100	100	100	100	100	100	100	100
0.3	0.002	16.67	6.33	21.40	0	2.03	3.97	0	3.06	3.94	11.00	11.04	15.25
	0.005	72.33	86.33	92.67	30.67	55.33	59.00	0	32.33	46.00	85.67	87.33	91.67
	0.008	97.67	100	100	86.00	97.3	98.00	4.00	76.33	88.33	99.33	99.67	100
	0.01	100	100	100	96.33	99.67	99.33	17.33	94.00	97.00	100	100	100
	0.02	100	100	100	100	100	100	96.67	100	100	100	100	100
-0.3	0.002	0.67	4.36	61.43	0	54.00	5.09	0	93.33	1.82	0	2.07	18.57
	0.005	5.67	62.33	79.00	0	95.67	40.40	0	99.67	33.22	0	56.00	48.33
	0.008	23.67	96.33	97.33	0	100	86.67	0	100	82.67	1.67	96.33	86.33

From the simulation results, we can propose a last argument in order to choose the best model. In case where only one of the two models detects the random slope, the model that does not detect the random effect should be preferred. This model is the most suitable to fit the data.

4 Application to HRQoL data

In this section, we performed a longitudinal analysis of the HRQoL data from a multicenter randomized phase III clinical trial in first-line metastatic pancreatic cancer patients: PRODIGE4/ACCORD11 (Conroy et al., 2011). Three hundred and forty-two patients were randomly assigned to Folfirinix (experimental arm) versus Gemcitabine (control arm) regimens. The detailed inclusion and exclusion criteria, the study design and protocol, the treatment, the compliance to the questionnaires, and the HRQoL anal-

yses have previously been published (Conroy et al., 2011; Gourgou-Bourgade et al., 2012; Barbieri et al., 2015). The patients filled the EORTC QLQ-C30 questionnaire themselves at different follow-up times defined in the protocol: at baseline, day 15, day 30, and at months 2, 4, 6, 8, and 10. The different measuring times reflected the longitudinal aspect of the HRQoL and allowed the assessment of the change of HRQoL for each dimension.

Analyses were performed using the SAS software (version 9.3) (Institute, 2011; Boeck and Wilson, 2004). For all previous arguments, the cumulative models are preferred for the longitudinal analysis of the HRQoL. There, we considered the (*cumulative,logistic,Z₁,U₂*) model with to analyze the data. In HRQoL study in oncology, the analysis is carried out for each HRQoL dimension. Given one HRQoL dimension with few correlated items, the discrimination parameters could be considered equals to one for each item. The distinction between the multiple-item responses is only achieved though difficulty parameters (thresholds) (Anota et al., 2014; Barbieri et al., 2015). For the HRQoL longitudinal analysis with the subject i ($i = 1, \dots, 342$), the visit v ($v = 1, \dots, 8$), the item j with M_j response categories, the (*cumulative,logistic,Z₁,U₂*) model is defined by:

$$\Pr\left(Y_{iv}^{(j)} \geq m | \theta_i\right) = \frac{\exp\left(\eta_{ivm}^{(j)}\right)}{1 + \exp\left(\eta_{ivm}^{(j)}\right)}, \quad (8)$$

with the following linear predictor is considered in the analyses:

$$\begin{cases} \eta_{ivm}^{(j)} = \theta_{iv} - \delta_{jm} \\ \theta_{iv} = g_i \beta_1 + (t_v - t_0) \beta_2 + g_i (t_v - t_0) \beta_3 + \xi_{i0} + (t_v - t_0) \xi_{i1} \end{cases} \quad (9)$$

where:

- δ_{jm} is the difficulty parameter (threshold) associated with the category m of item j ;
- t_v is the date of the visit v , and t_0 is the date of baseline;
- $g_i = 1$ if the patient i belongs the experimental group (Folfinrox), $g_i = 0$ if the patient i belongs the control group (Gemcitabine);
- β_1 is the effect difference at baseline between Folfinrox and Gemcitabine groups;
- β_2 is the slope (evolution) of health-related quality-of-life perception for the Gemcitabine group;
- $\beta_2 + \beta_3$ is the slope (evolution) of health-related quality-of-life perception for the Folfinrox group;
- ξ_{i0} and ξ_{i1} are respectively the subject-specific random effects associated with the intercept and the slope such as $(\xi_{i0}, \xi_{i1})' \sim \mathcal{N}(\mathbf{0}, \Sigma)$, Σ being the unstructured covariance matrix.

These HRQoL data have been already analyzed with different approaches. Gourgou-Bourgade et al. (2012) have presented the results using time-to-event models. They concluded for a better HRQoL in the Folfinrox arm than the Gemcitabine arm. Then, Barbieri et al. (2015) have presented the results through the LMM and the partial credit model extended for the longitudinal analysis (*adjacent,logistic,Z₁,U₂*). The conclusions of both mixed models are similar.

For the (*cumulative, logistic, Z_1, U_2*) model, Table 6 shows the estimations of fixed parameters, their standard deviation and the associated P-value of the Wald test. Concerning the functional dimension, we performed a reverse permutation on the functional scales for an intuitive interpretation. This allows to consider that an increase of the latent trait θ is associated with an increase of the functional capacity (improvement of HRQoL) or increase of the symptoms (deterioration of HRQoL). For all HRQoL dimensions, there should be no difference at baseline ($\beta_1 = 0$) in a randomized clinical trial. However, we observed a significant difference concerning the diarrhea symptom between the two groups at baseline ($p = 0.007^{**}$). It referred to a difference between the two arms at day 15, day 30 (during treatment period) but not necessarily at baseline. Then, the perception of diarrhea symptom remained higher in the Folfirinox arm over time. This result is expected because the Folfirinox is more toxic than Gemcitabine and is known to cause more diarrhea symptom.

Regarding the other dimensions, the HRQoL changed over time for several dimensions (emotional functioning, pain, insomnia, constipation and appetite loss) returning a significant improvement in terms of HRQoL perception. Only the pain showed a significantly different evolution between the two arms ($p = 0.04$). Indeed, the patients receiving the Folfirinox had a perception of pain which decreased significantly more over time than the patients receiving the Gemcitabine.

Among the advantages of the cumulative models, there are the interpretation and the illustration more intuitive. The constraints on the item parameter in these models allow an interpretation through the latent variable (e.g. comparing the proportions of the response categories for one specific item over time or between different groups during a fixed time). Figure 2 presents the evolution concerning the probability of response either over time (Figure 2a) or between group (Figure 2b). This example is illustrated through the first item of the pain symptom of the clinical trial previously presented. The probability (π_m) for a patient to response the category m corresponds to the area under the curve delimited by the horizontal lines. For both groups, Figure 2a shows that the probability to choose the categories 2 or 3 decreased over time while the probability to choose the category 0 increased. At baseline, the response proportion for the categories 0, 1, 2 and 3 were respectively $\pi_0 = 0.10$, $\pi_1 = 0.62$, $\pi_2 = 0.22$ and $\pi_3 = 0.06$ for each group. Then, the evolution of the proportions showed a decrease of the level of pain between the baseline and the 4-month visit, and, finally, a decrease of the latent trait over time. Likewise, Figure 2b shows the different response proportions between the two groups at four months. In the control group, the proportions were $\pi_0 = 0.29$, $\pi_1 = 0.61$, $\pi_2 = 0.08$ and $\pi_3 = 0.02$ for the categories 0, 1, 2 and 3. In the experimental group, they were $\pi_0 = 0.47$, $\pi_1 = 0.48$, $\pi_2 = 0.04$ and $\pi_3 = 0.01$. The probability to response category 3 was the lowest whatever the group, but was even less likely for patients in experimental group than in control group. In contrary, the probability to response category 0 was more likely in experimental group than in control group. The observed gap corresponded to the difference between the two linear predictors associated with each group only four months after the baseline. One of the interests of this illustration concerns the clinical interpretation. The IRT models thus offer a complete analysis: the general analysis of a HRQoL dimension and the specific analysis for each item (Edelen and Reeve, 2007).

5 Discussion

We have explored the suitable mixed models for the longitudinal analysis of the HRQoL in oncology. This data coming from questionnaires through Likert scales, we focused on regression models for ordinal

Table 6: Estimations of fixed effect parameters $(\beta_p)_{p=1,2,3}$ of the $(cumulative, logistic, Z_1, U_2)$ model. All HRQoL dimensions of the EORTC QLQ-C30 are considered.

HRQoL Dimensions	Coefficient	Standard error	Pvalue
Global Health Status			
β_2	0.098	0.070	0.166
β_3	0.130	0.085	0.128
Physical functioning			
β_2	-0.150	0.077	0.051
β_3	0.122	0.098	0.212
Role functioning			
β_2	-0.011	0.081	0.892
β_3	0.157	0.103	0.131
Emotional functioning			
β_2	0.335	0.070	<.001***
β_3	0.001	0.086	0.992
Cognitive functioning			
β_2	-0.002	0.054	0.972
β_3	0.088	0.067	0.189
Social functioning			
β_2	0.010	0.073	0.888
β_3	0.116	0.093	0.211
Fatigue			
β_2	-0.087	0.085	0.308
β_3	-0.033	0.107	0.761
Nausea and vomiting			
β_2	-0.052	0.060	0.393
β_3	-0.069	0.072	0.336
Pain			
β_2	-0.330	0.076	<.001***
β_3	-0.188	0.092	0.040*
Dyspnea			
β_2	-0.060	0.075	0.420
β_3	-0.093	0.088	0.295
Insomnia			
β_2	-0.359	0.080	<.001***
β_3	0.046	0.083	0.627
Appetite loss			
β_2	-0.354	0.072	<.001***
β_3	-0.026	0.080	0.747
Constipation			
β_2	-0.325	0.077	<.001***
β_3	0.003	0.083	0.974
Diarrhea			
β_1	0.739	0.272	0.007**
β_2	0.018	0.067	0.792
β_3	-0.026	0.076	0.786
Financial difficulties			
β_2	-0.522	0.282	0.066
β_3	0.302	0.208	0.146

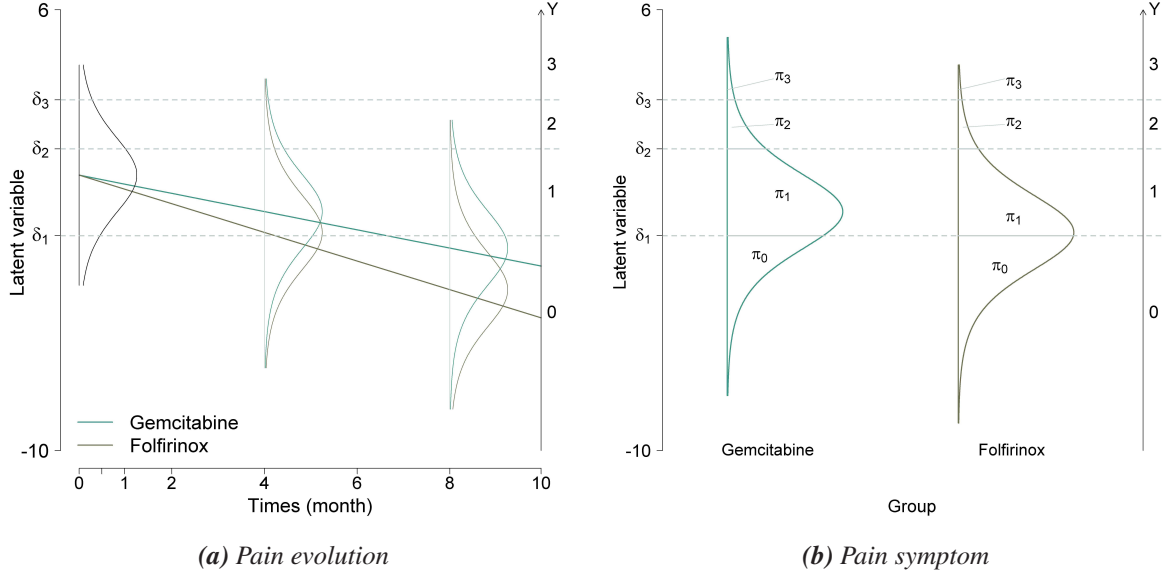


Figure 2: Interpretation of the (cumulative, logistic, Z_1, U_2) model through its underlying latent variable for the first item concerning the pain symptom. The estimated difficulty parameter are $\delta_1 = -2.1$, $\delta_2 = 1$ and $\delta_3 = 2.75$. Figure 2a: the different HRQoL evolution of the latent variable and response variable Y between the two groups. Figure 2b: the different proportions (π_m) of different responses Y between the two groups four month after the Baseline (t_4).

data. These models have been specified with three components, the linear predictor parameterization, the ratio of probabilities and the CdF (Peyhardi et al., 2015). In oncology, the analysis being performed on multiple-item measurements associated with one HRQoL dimension (Fayers et al., 2001), the specific IRT parameterization of the linear predictor is thus used. The item parameters allow to distinguish the outcomes from different items which measure an unique unidimensional latent variable. This latent variable was decomposed linearly to take into account the different covariates in the fixed part of the model and to incorporate subject-specific random effects. The analysis with IRT models is the richer because they are based on raw data (Gorter et al., 2015). The analysis can be made on one specific item through the item parameters or on the whole HRQoL dimension (Edelen and Reeve, 2007). Indeed, these models take into consideration all available information from the data, it is why the use of this kind of model is more and more studied (Gorter et al., 2015).

Then, concerning the choice of the model family, the cumulative and adjacent models are preferred. From the ratios of probabilities which characterize them and the symmetric CdF, the practical property of the invariant seems important to interpret the results. The cumulative models also assume an underlying continuous latent variable that is associated with a linear mixed regression model (McCullagh, 1980; Hedeker and Gibbons, 1994). This allows a better interpretation and illustration of the results such as the easy analysis of the evolution of the response proportions of the different categories over time or between groups, given one item. The adjacent models show the advantage not to have any constraint for the model estimation. These models can thus be preferred when the regression is performed on the item part of the linear predictor, given non-proportional design. Finally, the choice of the CdF essentially depends on the

observed data and properties which interest the users. These IRT models are reversible only if the CdF is symmetric. For the longitudinal analysis of HRQoL in oncology, we recommend the use of commonly symmetric CdF: the logistic and the Gaussian distributions. There is no argument for the choice between the two CdF. Indeed, the CdF adjustment for each item is achieved by discrimination parameters.

The simulation study showed that the IRT model capacity to detect the random effect was better than the LMM currently used. This result seems natural because the LMM is based on the study of a summary variable with less information. Thus, the variability from data is also reduced. Of course, the usefulness of the random effect introduction in the model is strongly associated with the observed data. Moreover, the more the difficulty parameters were distinct and the influence of covariates was stronger, the less the random effect provided information. All these results confirmed that the IRT models allow a more detailed analysis to interpret the results from a specific dimension or item. Whatever the IRT model used to generate the data, the LMM remained competitive through these simulations. However, the IRT model, that did not generate data, seemed more sensitive to the random slope than the other IRT model used to simulate the dataset. Indeed, in some cases, it tended to detect the random slope while it did not exist. In case where one of the two models detects the random slope, we recommend the use of the model not detecting the effect as it is the most suitable to fit the data. On the opposite, if both models are identically sensitive to its presence, we suggest using the IRT model following the previous practical and methodological arguments.

An aspect that remains to be discussed is the multidimensional aspect of HRQoL. Nowadays in oncology, the different dimensions are analyzed independently of one another, and this causes the problem of multiple tests. Moreover, the number of items within each HRQoL dimension is small (between 1 and 5 items). We can thus imagine an unidimensional latent trait for two or several HRQoL dimensions. Indeed, the discrimination parameters could be considered as equal between items from the same dimension. Finally, this allows to consider a set of HRQoL dimensions to assess a common concept like such as the symptomatic status concerning all symptomatic dimensions. Another approach to consider the all HRQoL dimensions would be the use of structural equation modeling. This would allow to summarize the information through some factors to explain the evolution of the global HRQoL and the potential structural link between the latent variables.

Acknowledgment

This study was supported by a grant from the French Public Health Research Institute (www.iresp.net) under the 2012 call for projects as part of the 2009-2013 Cancer Plan. We thank Dr. H el ene de Forges for her editorial assistance and UNICANCER for the data from PRODIGE4/ACCORD11 clinical trial which is used in this paper.

References

- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., and de Haes, J. C. (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. *Journal of the National Cancer Institute*, 85(5):365–376.
- Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. John Wiley & Sons.

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4):561–573.
- Anota, A., Barbieri, A., Savina, M., Pam, A., Gourgou-Bourgade, S., Bonnetain, F., and Bascoul-Mollevis, C. (2014). Comparison of three longitudinal analysis models for the health-related quality of life in oncology: a simulation study. *Health and Quality of Life Outcomes*, 12.
- Anota, A., Hamidou, Z., Paget-Bailly, S., Chibaudel, B., Bascoul-Mollevis, C., Auquier, P., Westeel, V., Fiteni, F., Borg, C., and Bonnetain, F. (2013). Time to health-related quality of life score deterioration as a modality of longitudinal analysis for health-related quality of life studies in oncology: do we need RECIST for quality of life to achieve standardization? *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*.
- Bacci, S., Bartolucci, F., and Gnaldi, M. (2012). A class of Multidimensional Latent Class IRT models for ordinal polytomous item responses. *arXiv:1201.4667 [stat]*. arXiv: 1201.4667.
- Barbieri, A., Anota, A., Conroy, T., Gourgou-Bourgade, S., Juzyna, B., Bonnetain, F., Lavergne, C., and Bascoul-Mollevis, C. (2015). Applying the Longitudinal Model from Item Response Theory to Assess the 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*.
- Blanchin, M., Hardouin, J.-B., Neel, T. L., Kubis, G., Blanchard, C., Mirallié, E., and Sébille, V. (2011). Comparison of CTT and Rasch-based approaches for the analysis of longitudinal Patient Reported Outcomes. *Statistics in Medicine*, 30(8):825–838.
- Boeck, P. d. and Wilson, M. (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. Springer, New York.
- Cella, D. F., Tulsky, D. S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., Silberman, M., Yellen, S. B., Winicour, P., and Brannon, J. (1993). The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 11(3):570–579.
- Chinot, O. L., Wick, W., Mason, W., Henriksson, R., Saran, F., Nishikawa, R., Carpentier, A. F., Hoang-Xuan, K., Kavan, P., Cernea, D., Brandes, A. A., Hilton, M., Abrey, L., and Cloughesy, T. (2014). Bevacizumab plus Radiotherapy–Temozolomide for Newly Diagnosed Glioblastoma. *New England Journal of Medicine*, 370(8):709–722.
- Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Gourgou-Bourgade, S., de la Fouchardière, C., Bennouna, J., Bachet, J.-B., Khemissa-Akouz, F., Péré-Vergé, D., Delbaldo, C., Assenat, E., Chauffert, B., Michel, P., Montoto-Grillot, C., and Ducreux, M. (2011). FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *The New England journal of medicine*, 364(19):1817–1825.
- Edelen, M. O. and Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 16 Suppl 1:5–18.

- Fahrmeir, L. and Tutz, G. (2001). Multivariate Statistical Modelling Based on Generalized Linear Models. Springer.
- Fayers, P. M., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., and Bottomley, A. o. b. o. t. E. Q. o. L. G. (2001). EORTC QLQ-C30 Scoring Manual (3rd edition), volume Brussels: EORTC 2001.
- Fiteni, F., Westeel, V., Pivot, X., Borg, C., Vernerey, D., and Bonnetain, F. (2014). Endpoints in cancer clinical trials. Journal of Visceral Surgery, 151(1):17–22.
- Gilbert, M. R., Dignam, J. J., Armstrong, T. S., Wefel, J. S., Blumenthal, D. T., Vogelbaum, M. A., Colman, H., Chakravarti, A., Pugh, S., Won, M., Jeraj, R., Brown, P. D., Jaeckle, K. A., Schiff, D., Stieber, V. W., Brachman, D. G., Werner-Wasik, M., Tremont-Lukats, I. W., Sulman, E. P., Aldape, K. D., Curran, W. J., and Mehta, M. P. (2014). A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma. New England Journal of Medicine, 370:699–708.
- Gorter, R., Fox, J.-P., and Twisk, J. W. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. BMC Medical Research Methodology, 15(1):55.
- Gourgou-Bourgade, S., Bascoul-Mollevi, C., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Boige, V., Bérille, J., and Conroy, T. (2012). Impact of FOLFIRINOX Compared With Gemcitabine on Quality of Life in Patients With Metastatic Pancreatic Cancer: Results From the PRODIGE 4/ACCORD 11 Randomized Trial. Journal of clinical oncology: official journal of the American Society of Clinical Oncology.
- Grilli, L. and Rampichini, C. (2011). Multilevel Models for Ordinal Data. In Kenett, R. S. and Salini, S., editors, Modern Analysis of Customer Surveys, pages 391–411. John Wiley & Sons, Ltd.
- Hamidou, Z., Dabakuyo, T. S., Mercier, M., Fraisse, J., Causeret, S., Tixier, H., Padeano, M.-M., Loustalot, C., Cuisenier, J., Sauzedde, J.-M., Smail, M., Combier, J.-P., Chevillote, P., Rosburger, C., Arveux, P., and Bonnetain, F. (2011). Time to Deterioration in Quality of Life Score as a Modality of Longitudinal Analysis in Patients with Breast Cancer. The Oncologist, 16(10):1458–1468.
- Hardouin, J.-B., Audureau, E., Leplège, A., and Coste, J. (2012). Spatio-temporal Rasch analysis of quality of life outcomes in the French general population. Measurement invariance and group comparisons. BMC Medical Research Methodology, 12(1):182.
- Hardouin, J.-B., Blanchin, M., Feddag, M.-L., Néel, T. L., Perrot, B., and Sébille, V. (2015). Power and sample size determination for group comparison of patient-reported outcomes using polytomous Rasch models. Statistics in Medicine, pages n/a–n/a.
- Hedeker, D. and Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. Biometrics, 50(4):933–944.
- Huber, C., Limnios, N., Mesbah, M., and Nikulin, M. (2013). Mathematical Methods in Survival Analysis, Reliability and Quality of Life. John Wiley & Sons.
- Institute, S. A. S. (2011). SAS/STAT 9.3 User's Guide: Mixed Modeling (Book Excerpt). SAS Institute.

- Jafari, P., Bagheri, Z., Ayatollahi, S. M., and Soltani, Z. (2012). Using Rasch rating scale model to reassess the psychometric properties of the Persian version of the PedsQLTM 4.0 Generic Core Scales in school children. Health and Quality of Life Outcomes, 10(1):27.
- Liu, L. C. and Hedeker, D. (2006). A mixed-effects regression model for longitudinal multivariate ordinal data. Biometrics, 62(1):261–268.
- Masters, G. (1982). A rasch model for partial credit scoring. Psychometrika, 42(2):149–174.
- McCullagh, P. (1980). Regression models for ordinal data (with discussion). Journal of the Royal Statistical Society, Series B, 42:109–142.
- Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. Applied Psychological Measurement, 16(2):159–176.
- Peyhardi, J., Trottier, C., and Guédon, Y. (2015). A new specification of generalized linear models for categorical responses. Biometrika, page asv042.
- Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Danmarks Paedagogiske Institut.
- Samejima, F. (1968). Estimation of Latent Ability Using a Response Pattern of Graded Scores1. ETS Research Bulletin Series, 1968(1):i–169.
- Titman, A. C., Lancaster, G. A., and Colver, A. F. (2013). Item response theory and structural equation modelling for ordinal data: Describing the relationship between KIDSCREEN and Life-H. Statistical Methods in Medical Research.
- Tutz, G. (1990). Sequential item response models with an ordered response. British Journal of Mathematical and Statistical Psychology, 43(1):39–55.
- Verhagen, J. and Fox, J.-P. (2013). Longitudinal measurement in health-related surveys. A Bayesian joint growth model for multivariate ordinal responses. Statistics in Medicine, 32(17):2988–3005.

Deuxième partie

Méthodes complémentaires pour la prise en compte de nouvelles structures dans l'étude de la HRQoL

Étude globale et longitudinale par modèles à équation structurelle et à effets aléatoires

L'objectif de ce chapitre est d'envisager l'étude de la HRQoL de manière globale et non plus dimension par dimension. Nous proposons un modèle à équation structurelle applicable à chaque temps de mesure permettant d'une analyse globale de la HRQoL, et d'autre part l'estimation de deux facteurs sujet-spécifiques reflétant respectivement les dimensions fonctionnelles et symptomatiques. Une étude de simulation permet de vérifier la programmation et de dégager des recommandations pour l'analyse. Un avantage de ce modèle est de répondre au problème de la multiplicité des tests pour l'étude du critère de HRQoL. Les analyses longitudinales peuvent être réduites à celles du statut global de santé, du statut fonctionnel (facteur fonctionnel) et du statut symptomatique (facteur symptomatique). Une application sur les données de l'essai CO-HO-RT permettra d'illustrer l'approche proposée.

Sommaire

4.1	Introduction	94
4.2	Approche par modèles à équation structurelle	96
4.2.1	Estimation du modèle via l'algorithme EM	99
4.2.2	Analyse de sensibilité par simulations	107
4.3	Application	114
4.3.1	Résultats	115
4.3.2	Compléments sur les données réelles	116
4.4	Discussion	117
4.5	Article : <i>EM algorithm estimation of a structural equation model for the longitudinal study of the quality of life</i>	119

4.1 Introduction

L'analyse de la HRQoL telle qu'elle est actuellement réalisée dans les essais cliniques pose le problème de la multiplicité des tests. En considérant seulement le questionnaire EORTC QLQ-C30 sans module additionnel, quinze analyses indépendantes sont réalisées, soit une par dimension de HRQoL. En effet, l'influence des variables explicatives, tel que le traitement, est testée pour chacune des dimensions de HRQoL de manière indépendante. Sans ajustement du risque de première espèce, le risque global α_G de conclure à tort une différence sur les $k = 15$ dimensions est : $\alpha_G = 1 - (1 - \alpha)^k \approx 0.54$ pour $\alpha = 5\%$.

Un moyen de palier à ce problème est de réaliser une analyse ciblée suivant les traitements testés. En effet, dans un essai clinique randomisé, la connaissance des cliniciens sur les traitements peut permettre de focaliser l'analyse sur une ou plusieurs dimensions particulières de HRQoL. Par exemple, dans l'essai clinique Prodigé 4/ACCORD 11, le Folfirinox est reconnu pour accentuer le symptôme de diarrhée. Connaître la (les) toxicité(s) et les effets secondaires va alors permettre de cibler une (des) dimension(s) particulière(s) de HRQoL pour l'analyse, en plus du statut global de santé (*global health statut*, GHS) qui est toujours analysé. Dans ce cas, un faible nombre d'analyses indépendantes est réalisé et la modélisation étudiée dans le chapitre précédent est adaptée. Très récemment, un design associant la HRQoL à l'OS comme co-critère principal a été proposé dans lequel l'analyse de la HRQoL est basée sur trois dimensions $k = 3$ de HRQoL réduisant ainsi la dépense de α_G de manière raisonnable. Le choix de ces dimensions s'est dirigé naturellement sur le GHS qui représente la HRQoL globale, la fonction physique qui comprend le plus d'items et la fatigue qui reste le symptôme le plus général et le plus représenté en nombre d'items.

Cependant, si aucune hypothèse n'est formulée a priori par le clinicien, l'analyse s'effectue alors de manière globale, sur chaque dimension de la HRQoL de manière indépendante. Dans ce chapitre, nous nous plaçons dans cette dernière situation où aucune connaissance *a priori* des effets secondaires ou des toxicités n'est suffisante pour focaliser l'analyse sur quelques dimensions spécifiques. Sans une connaissance *a priori* des dimensions particulières à surveiller, il est difficile de choisir celles qui sont les plus pertinentes à analyser. Quelle que soit la méthode statistique utilisée, cela engendre un problème de tests multiples.

Comme alternative, nous proposons une nouvelle approche globale qui permet de tenir compte conjointement de l'aspect longitudinal et de la nature multidimensionnelle de la HRQoL. L'idée est proche de celle des modèles IRT où nous souhaitons estimer des variables latentes reflétées par un ensemble de dimensions de HRQoL, et non d'items. En effet, l'objectif est l'estimer de traits communs à plusieurs variables qui ne sont pas directement mesurables, tout en gardant la structure conceptuelle de données issues de QLQ-C30. Pour cela, nous nous sommes placés sous l'approche CTT où les scores de HRQoL sont considérés comme une bonne représentation du niveau de HRQoL. Dans ce chapitre, nous nous intéressons aux modèles à équations structurelles (*equation structural modeling*, SEM) qui sont souvent associés aux

modèles IRT dans la littérature (Rabe-Hesketh and Skrondal, 2008; Titman et al., 2013). Les modèles IRT tout comme les SEM font partie de la classe des modèles à variables latentes, au même titre que les modèles de régression multi-niveaux ou encore les modèles à facteurs (Rabe-Hesketh et al., 2004). Les modèles IRT peuvent être considérés comme une classe particulière des SEM. En effet, ils comprennent un facteur unidimensionnel où les variables réponses observées sont ordinales (voir chapitre 2). À l’opposé, les SEM sont généralement associés aux données continues car ils sont plus complexes que les modèles IRT. Ils ont tendance à avoir une variable latente multi-dimensionnelle (plusieurs facteurs), et formalisent les liens de dépendance des variables numériques observées (modèle de mesure), tout en supposant une relation entre ces variables latentes (modèle structurel).

Dans la littérature, deux familles concurrentes de méthodes traitant des modèles à équations structurelles sont présentées : les méthodes à facteurs et les méthodes à composantes. La première famille fondée sur l’estimation du maximum de vraisemblance a d’abord été développée par Joreskog (1970) et mise en œuvre dans le logiciel LISREL. Le terme de facteur renvoie à celui de variable latente supposée être normalement distribuée et reflétée par les variables observées. L’estimation des paramètres (coefficients du prédicteur linéaire et des variances des erreurs) est basée sur la structure de la matrice de covariance des données du modèle. Cependant, cette approche n’inclue pas l’estimation des valeurs spécifiques des facteurs. La seconde famille est celle des méthodes à composantes. Celle-ci a été initialement développée par Wold sous l’appellation de la méthode PLS (*Partial Least Squares*) (Wold, 1966; Noonan and Wold, 1977; Wold, 1985). Cette approche suppose que chaque variable latente est une composante, à savoir une combinaison linéaire des variables observées correspondantes. Cette hypothèse est plus forte que la simple hypothèse de distribution de facteurs, mais elle permet d’estimer facilement la variable latente.

Ces deux approches présentent chacune un avantage : l’utilisation de facteurs pour la méthode LISREL (qui ne suppose pas la contrainte de linéarité), et l’estimation spécifique des variables latentes dans la méthode PLS. Un travail réalisé par Bry et al. (2015) a consisté à joindre ces deux avantages. Dans leur approche, les paramètres du SEM et les valeurs sujet-spécifiques des facteurs sont estimés sans la considération d’une contrainte linéaire. Ceci est rendu possible par l’intermédiaire d’un l’algorithme EM. En collaboration avec les auteurs de ces travaux et tout particulièrement avec Myriam Tami, nous avons développé une nouvelle approche basée sur une estimation de modèles d’équation structurelle pour l’analyse de la HRQoL via l’algorithme EM. Cette approche que nous proposons se résume en deux points :

1. Appliquer le SEM à chaque visite sur l’ensemble des dimensions observées de HRQoL pour :
 - Décrire globalement l’influence de variables explicatives sur chaque score de HRQoL ;
 - Reconstruire les deux facteurs associés respectivement aux dimensions fonctionnelles et aux dimensions symptomatiques ;
2. Modéliser le GHS, les facteurs (fonctionnel et symptomatique) au cours du temps au travers d’un LMM.

Ce travail a donné lieu à un article inséré à la fin de ce chapitre en section 4.5. L'utilisation des LMM pour l'analyse longitudinale étant présentée dans cet article et en section 1.4.1 de ce manuscrit, nous n'y reviendrons pas dans ce chapitre. En revanche, nous nous attarderons sur l'estimation et la modélisation du SEM. Après avoir décrit le SEM ainsi que son estimation par l'algorithme EM, nous présentons une étude de simulation afin d'étudier la sensibilité du SEM en fonction du nombre de variables observées ou de sujets pris en compte. Cette étude de simulation s'étendra également sur l'influence des variables explicatives dans le modèle, ainsi que l'intérêt de l'estimation des facteurs dans l'étude spécifique de la HRQoL. Enfin, nous reviendrons sur l'application présentée dans l'article 4.5 afin d'en compléter et discuter les résultats.

4.2 Approche par modèles à équation structurelle

Un SEM formalise les relations de dépendance de beaucoup de variables numériques mesurées ou observées par un plus petit nombre de variables non observées, appelées variables latentes. Chaque variable latente est supposée être reflétée par un ensemble spécifique de variables observées. En outre, un SEM est défini suivant deux types d'équations : les équations de mesure et les équations structurelles. Chaque équation de mesure concerne au moins une variable latente correspondant à un ensemble de variables observées. Puis, une équation structurelle officialise une relation supposée entre les variables latentes.

Le SEM que nous proposons est intuitif et découle de la construction même des questionnaires de HRQoL du groupe EORTC. La Figure 4.1 illustre la décomposition du concept de HRQoL en plusieurs dimensions suivant les recommandations de l'EORTC pour le questionnaire générique EORTC QLQ-C30 auquel est ajouté le module complémentaire EORTC QLQ-BR23. Deux familles de dimension émergent : la première concerne les dimensions fonctionnelles et la seconde celle des dimensions symptomatiques. En plus de ces deux familles, une dimension représentant le concept en général de HRQoL, le GHS, est considérée séparément.

La Figure 4.2 schématise le SEM sous-jacent à la conception du questionnaire. Soit, une variable y à expliquer, qui représente la dimension associée au GHS, dépendante de deux groupes explicatifs X^1 et X^2 , représentés respectivement par les q_1 variables associées aux dimensions fonctionnelles et par les q_2 variables associées aux dimensions symptomatiques. Ces deux groupes sont appelés groupes de mesure. Nous supposons qu'il existe une variable latente notée $(f^r)_{r=1,2}$, appelée facteur, qui est reflétée par les variables observées du groupe de mesure $(X^r)_{r=1,2}$, ainsi que les $(qT^r)_{r=1,2}$ variables explicatives potentielles $((T^r)_{r=1,2})$ (traitements, centre, âge, ...). Le SEM proposé est formulé selon le système d'équations suivant :

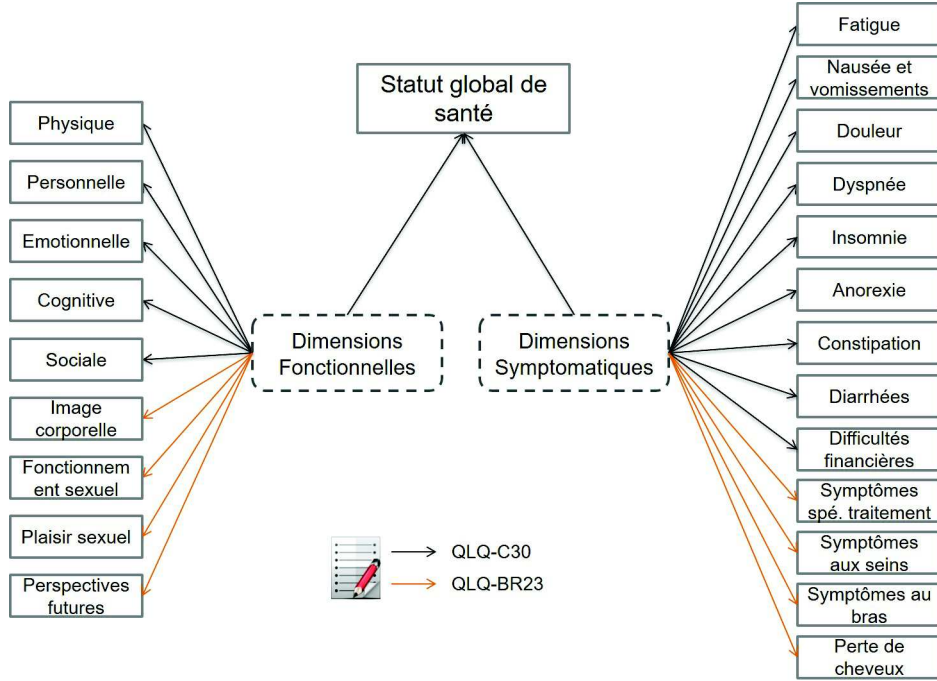


FIGURE 4.1: Décomposition intuitive de la HRQoL suivant la construction du questionnaire générique EORTC QLQ-C30 et du questionnaire spécifique du cancer du sein EORTC QLQ-BR23.

$$\begin{cases}
 \underbrace{y}_{(n \times 1)} = \underbrace{T}_{(n \times q_y)} \underbrace{d}_{(q_y \times 1)} + \underbrace{f^1}_{(n \times 1)} \underbrace{c^1}_{(1 \times 1)} + \underbrace{f^2}_{(n \times 1)} \underbrace{c^2}_{(1 \times 1)} + \underbrace{\varepsilon^Y}_{(n \times 1)} \\
 \underbrace{X^1}_{(n \times q_1)} = \underbrace{T^1}_{(n \times q_{T1})} \underbrace{D^1}_{(q_{T1} \times q_1)} + \underbrace{f^1}_{(n \times 1)} \underbrace{(a^1)'}_{(1 \times q_1)} + \underbrace{\varepsilon^1}_{(n \times q_1)} \\
 \underbrace{X^2}_{(n \times q_2)} = \underbrace{T^2}_{(n \times q_{T2})} \underbrace{D^2}_{(q_{T2} \times q_2)} + \underbrace{f^2}_{(n \times 1)} \underbrace{(a^2)'}_{(1 \times q_2)} + \underbrace{\varepsilon^2}_{(n \times q_2)}
 \end{cases} \quad (4.1)$$

où n est le nombre d'unités statistiques observées (sujets), y est la variable réponse de l'équation structurelle, $X^r = [(x_1^r)', (x_2^r)', \dots, (x_n^r)']'$ la matrice associée aux variables du groupe de mesure r ($r = 1, 2$). Le vecteur d et les matrices $(D^r)_{r=1,2}$ représentent les paramètres inconnus associés aux variables explicatives, et $T = [(t_1)', (t_2)', \dots, (t_n)']'$ et $T^r = [(t_1^r)', (t_2^r)', \dots, (t_n^r)']'_{r=1,2}$ les matrices de design respectives. Les variables explicatives ne sont pas nécessairement les mêmes pour expliquer le bloc de mesures associé aux dimensions fonctionnelles, celui associé aux dimensions symptomatiques ou encore la variable réponse y . Les scalaires $(c^r)_{r=1,2}$ représentent l'intensité de lien entre y et les différents facteurs. De la même manière, $(a^r)_{r=1,2}$ sont les coefficients pondérateurs entre les variables observées d'un bloc de mesure et le facteur respectif. Les hypothèses de ce modèle de régression à facteur sont les suivantes :

- Les observations i ($i = 1, \dots, n$) sont indépendantes ;
- $\forall r \in \{1, 2\}, f_i^r \sim \mathcal{N}(0, 1)$;

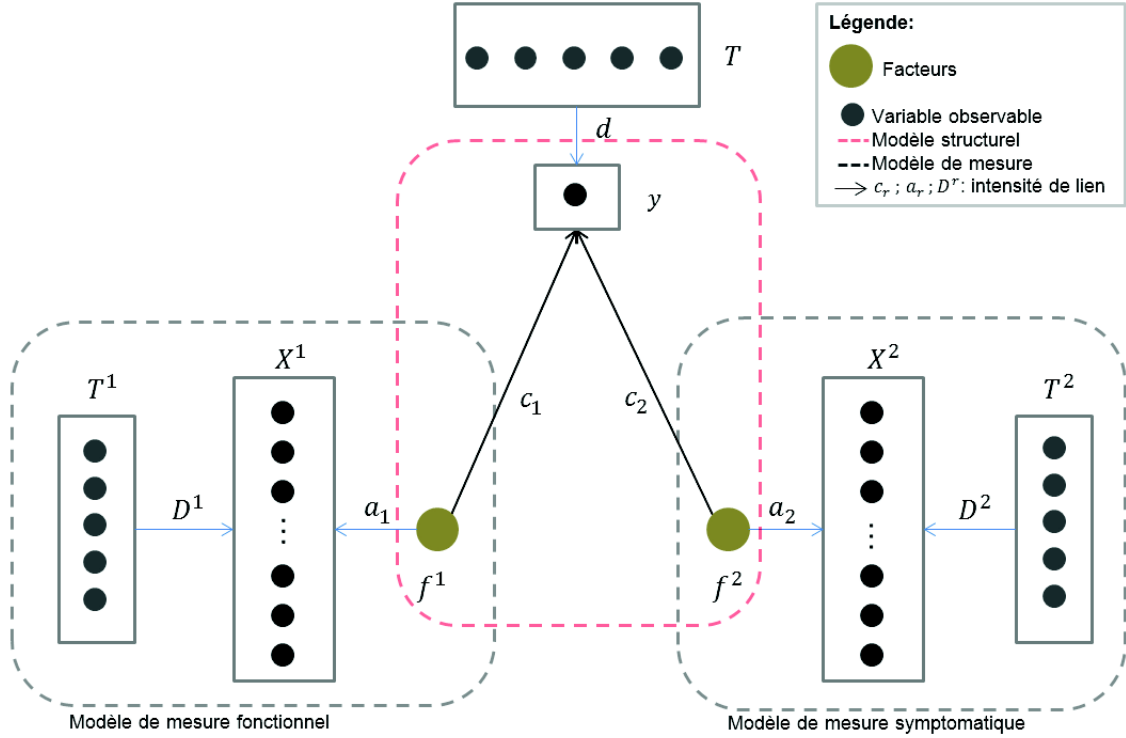


FIGURE 4.2: Schéma représentant le SEM avec les différents liens considérés.

- $\forall r \in \{1, 2\}$, $\varepsilon_i^r \sim \mathcal{N}(0, \psi_r)$, tel que $\psi_r = \sigma_r^2 Id_{q_r}$, soit toutes les variables observées d'un même bloc de mesure sont supposées de même variance ;
- $\varepsilon_i^Y \sim \mathcal{N}(0, \sigma_y^2)$;
- $\varepsilon_i^Y, \varepsilon_i^1, \varepsilon_i^2$ sont mutuellement indépendants de f_ℓ^1 et $f_{\ell'}^2$ pour tout i, ℓ, ℓ' ;
- Les facteurs f^1 et f^2 sont supposés indépendants ;
- Toutes les variables sont supposées gaussiennes¹ et indépendantes conditionnellement aux facteurs.

Pour un sujet $i (i = 1, \dots, n)$, le modèle se réécrit comme suit :

$$\begin{cases} x_i^{1'} = t_i^{1'} D^1 + f_i^1 a^1 + \varepsilon_i^{1'} & (4.2a) \\ x_i^{2'} = t_i^{2'} D^2 + f_i^2 a^2 + \varepsilon_i^{2'} & (4.2b) \\ y_i' = t_i' d + f_i^1 c^1 + f_i^2 c^2 + \varepsilon_i^{y'} & (4.2c) \end{cases}$$

où $a^1 = (a_1^1, \dots, a_{q_1}^1)'$, $a^2 = (a_1^2, \dots, a_{q_2}^2)'$, c^1 et c^2 réels.

Dans ce modèle, nous sommes en présence de données observées $z = (y, X^1, X^2)$ et de données cachées $h = (f^1, f^2)$. Les facteurs sont considérés comme des variables latentes et peuvent être associés à des données cachées, non directement mesurables. Le paramètre θ représente l'ensemble des paramètres du modèle à estimer,

1. Hypothèse de confort dans le cadre de la HRQoL ;

soit $\theta = (d, D^1, D^2, a^1, a^2, c^1, c^2, \sigma_1^2, \sigma_2^2, \sigma_y^2)$. D'après le modèle (4.1) et la règle des distributions gaussiennes, les distributions sont les suivantes :

- $y_i | f_i^1, f_i^2 \sim \mathcal{N}(d't_i + f_i^1 c^1 + f_i^2 c^2, \sigma_y^2)$;
- $\forall r \in \{1, 2\}, x_i^r | f_i^r \sim \mathcal{N}((D^r)' t_i + a^r f_i^r, \psi_r)$;
- $\forall r \in \{1, 2\}, f_i^r \sim \mathcal{N}(0, 1)$.

La densité des données complètes (z, h) , ou vraisemblance complétée, est de la forme :

$$\begin{aligned} \mathcal{L}(\theta; z, h) &= \prod_{i=1}^n p(z_i | h_i, \theta) p(h_i | \theta) \\ &= \prod_{i=1}^n p(y_i, x_i^1, x_i^2 | f_i^1, f_i^2; \theta) p(f_i^1, f_i^2; \theta) \\ &= \prod_{i=1}^n p(x_i^1 | f_i^1; \theta) p(x_i^2 | f_i^2; \theta) p(y_i | f_i^1, f_i^2; \theta) p(f_i^1) p(f_i^2), \end{aligned}$$

ce qui implique la log-vraisemblance complétée suivante :

$$\begin{aligned} \ell(\theta; z, h) &= -\frac{1}{2} \sum_{i=1}^n \{ \ln(\sigma_y^2) + \sigma_y^{-2} (y_i - d't_i - f_i^1 c^1 - f_i^2 c^2)^2 \} \\ &\quad -\frac{1}{2} \sum_{i=1}^n \{ \ln|\psi_1| + (x_i^1 - (D^1)' t_i - a^1 f_i^1)' \psi_1^{-1} (x_i^1 - (D^1)' t_i - a^1 f_i^1) \} \\ &\quad -\frac{1}{2} \sum_{i=1}^n \{ \ln|\psi_2| + (x_i^2 - (D^2)' t_i - a^2 f_i^2)' \psi_2^{-1} (x_i^2 - (D^2)' t_i - a^2 f_i^2) \} \\ &\quad -\frac{1}{2} \sum_{i=1}^n f_i^1 - \frac{1}{2} \sum_{i=1}^n f_i^2 + cte \end{aligned} \tag{4.3}$$

Les données à disposition ne permettent pas l'estimation des paramètres. La vraisemblance est analytiquement impossible à maximiser du fait que nous supposons la présence de facteurs latents (données cachées) caractérisés par le vecteur $f_i = (f_i^1, f_i^2)'$ pour l'unité statistique i .

4.2.1 Estimation du modèle via l'algorithme EM

L'algorithme EM ("*Expectation-Maximization*") est un algorithme itératif construit sur deux étapes successives [Dempster et al. \(1977\)](#). La première étape, l'étape E, revient à considérer l'espérance de la vraisemblance des données complètes conditionnellement aux données observées (z, h) et à la valeur courante du paramètre $\theta^{[s]}$ (à l'itération s) :

$$\mathbb{E}_z^h [\mathcal{L}(\theta; z, h)] = \mathbb{E} [\mathcal{L}(\theta; z, h) | z, \theta^{[s]}]. \tag{4.4}$$

La seconde étape, l'étape M, consiste quant à elle à maximiser l'espérance conditionnelle de la (log-)vraisemblance sachant la reconstruction des données cachées.

Les statistiques exhaustives (f) étant alors estimées par leur espérance conditionnelle (équation (4.4)), la mise à jour des paramètres est donnée par les estimations du maximum de vraisemblance :

$$\begin{aligned} \theta^{[s+1]} &= \operatorname{argmax}_{\theta} \mathbb{E}_z^h [\mathcal{L}(\theta; z, h)] \\ \Leftrightarrow \quad \mathbb{E}_z^h \left[\frac{\partial}{\partial \theta} \ell(\theta; z, h) \right] &= 0 \end{aligned} \quad (4.5)$$

Ces deux étapes sont répétées jusqu'à convergence de l'algorithme. Une fois que l'algorithme a convergé, nous obtenons l'estimation du maximum de vraisemblance des paramètres ($\widehat{\theta}$) mais également les estimations des facteurs $\widetilde{f}_i = (\widetilde{f}_i^1, \widetilde{f}_i^2)'$ pour chaque unité statistique $i (i = 1, \dots, n)$.

Étape E

Dans le cas de ce SEM, toutes les variables (h, z) sont supposées gaussiennes. Sous cette hypothèse, la loi des données cachées sachant les données observées ($p(h|z)$) est explicite d'après les résultats sur le conditionnement de lois gaussiennes. Les espérances conditionnelles des variables latentes étant calculables directement, nous les utilisons comme estimation des statistiques exhaustives manquantes h . Chercher la loi de $h_i|z_i$ revient à chercher la loi de $(f_i^1, f_i^2|y_i, x_i^1, x_i^2)$. Afin de la déterminer, exprimons dans un premier temps la loi jointe de $(f_i^1, f_i^2, y_i, x_i^1, x_i^2)$. Sous les hypothèses définies précédemment, pour $i = 1, \dots, n$:

- la loi de (f_i^1, f_i^2) est :

$$\begin{pmatrix} f_i^1 \\ f_i^2 \end{pmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right); \quad (4.6)$$

- $y_i \sim \mathcal{N}(d'T_i, (c^1)^2 + (c^2)^2 + \sigma_y^2)$;
- $\forall r \in \{1, 2\}, x_i^r \sim \mathcal{N}((D^r)'t_i^r, a^r a^{r'} + \Psi_r)$.

À partir des hypothèses établies précédemment, nous pouvons déduire explicitement la loi jointe des données complètes. Pour cela, les covariances associées doivent être calculées. Par exemple : $\forall r \in \{1, 2\}, \operatorname{Cov}(x_i^r, y_i) = \mathbb{E}(x_i^r y_i) - \mathbb{E}(x_i^r) \mathbb{E}(y_i)$, où :

$$\begin{aligned} x_i^r y_i &= ((D^r)'t_i^r + a^r f_i^r + \varepsilon_i^r) (d't_i + c^1 f_i^1 + c^2 f_i^2 + \varepsilon_i^Y) \\ &= (D^r)'t_i^r d't_i + (D^r)'t_i^r c^1 f_i^1 + (D^r)'t_i^r c^2 f_i^2 + (D^r)'t_i^r \varepsilon_i^Y \\ &\quad + a^r f_i^r d't_i + a^r f_i^r c^1 f_i^1 + a^r f_i^r c^2 f_i^2 + a^r f_i^r \varepsilon_i^Y \\ &\quad + \varepsilon_i^r d't_i + \varepsilon_i^r c^1 f_i^1 + \varepsilon_i^r c^2 f_i^2 + \varepsilon_i^r \varepsilon_i^Y \\ \Leftrightarrow \mathbb{E}(x_i^r y_i) &= (D^r)'t_i^r d't_i + a^r \mathbb{E}((f_i^r)^2) c^r \\ \Leftrightarrow \operatorname{Cov}(x_i^r y_i) &= (D^r)'t_i^r d't_i + a^r \mathbb{E}((f_i^r)^2) c^r - (D^r)'t_i^r d't_i \\ &= a^r \mathbb{V}(f_i^r) c^r \\ &= a^r c^r \end{aligned}$$

Suites aux calculs des covariances nécessaires et d'après les hypothèses du modèle (4.1), la loi jointe des données complètes est :

$$\begin{pmatrix} f_i^1 \\ f_i^2 \\ y_i \\ x_i^1 \\ x_i^2 \end{pmatrix} \sim \mathcal{N} \left(\mu = \begin{bmatrix} 0 \\ 0 \\ d't_i \\ (D^1)'t_i^1 \\ (D^2)'t_i^2 \end{bmatrix}, \Sigma \right)$$

où,

$$\Sigma = \begin{bmatrix} 1 & 0 & c^1 & (a^1)' & 0_{(1,q_2)} \\ 0 & 1 & c^2 & 0_{(1,q_1)} & (a^2)' \\ c^1 & c^2 & (c^1)^2 + (c^2)^2 + \sigma_Y^2 & c^1(a^1)' & c^2(a^2)' \\ a^1 & 0_{(q_1,1)} & c^1 a^1 & a^1(a^1)' + \Psi_1 & 0_{(q_1,q_2)} \\ a^2 & 0_{(q_2,1)} & c^2 a^2 & 0_{(q_2,q_1)} & a^2(a^2)' + \Psi_2 \end{bmatrix}.$$

Pour le calcul de la loi conditionnelle, nous utilisons le résultat suivant.

Si deux variables X_1 et X_2 suivent une loi normale tel que,

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N} \left(\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

où μ_1 est de dimension $q \times 1$, μ_2 est de dimension $p \times 1$, Σ_{11} est de dimension $q \times q$, Σ_{12} est de dimension $q \times p$, Σ_{21} est de dimension $p \times q$ et Σ_{22} est de dimension $p \times p$;

alors, $(X_1|X_2 = x_2) \sim \mathcal{N}(m, \Gamma)$ avec

$$\begin{aligned} m &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \Gamma &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

Par identification, nous avons

$$\begin{cases} \mu_1 = (0 \ 0)' \\ \mu_2 = (d't_i \ (D^1)'t_i^1 \ (D^2)'t_i^2)' \\ \Sigma_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ \Sigma_{22} = \begin{pmatrix} (c^1)^2 + (c^2)^2 + \sigma_Y^2 & c^1(a^1)' & c^2(a^2)' \\ c^1 a^1 & a^1(a^1)' + \Psi_1 & 0_{(q_1,q_2)} \\ c^2 a^2 & 0_{(q_2,q_1)} & a^2(a^2)' + \Psi_2 \end{pmatrix} \\ \Sigma_{12} = \begin{pmatrix} c^1 & (a^1)' & 0_{(1,q_2)} \\ c^2 & 0_{(1,q_1)} & (a^2)'\end{pmatrix} \\ \Sigma_{21} = (\Sigma_{12})' \end{cases}$$

Ainsi, la loi des facteurs conditionnellement aux données observées est :

$$h_i = \begin{pmatrix} f_i^1 \\ f_i^2 \end{pmatrix} | z_i \sim \mathcal{N} \left(m_{i1} = \begin{pmatrix} \mu_{i2} \\ \mu_{i2} \end{pmatrix}, \Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \right)$$

où

$$\begin{cases} m_i &= \mu_1 + \Sigma_{12} [\Sigma_{22}]^{-1} \begin{pmatrix} y_i - d't_i \\ x_i^1 - (D^1)'t_i^1 \\ x_i^2 - (D^2)'t_i^2 \end{pmatrix} \\ \Gamma &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12} \end{cases} \quad (4.7)$$

Étape M

Afin d'estimer les paramètres du modèle, nous cherchons à résoudre l'équation 4.5 associée à l'étape "M" de l'algorithme. Pour cela, la dérivée de la log-vraisemblance complétée (4.3) par rapport à chaque paramètre est nécessaire. Notons $\ell = \ell(\theta; z, h)$.

- Dérivée de (4.3) suivant le paramètre d

$$\begin{aligned} \frac{\partial \ell}{\partial d} &= \frac{\partial}{\partial d} \left[\frac{-1}{2\sigma_y^2} \sum_{i=1}^n (y_i - d't_i - f_i^1 c^1 - f_i^2 c^2)' (y_i - d't_i - f_i^1 c^1 - f_i^2 c^2) \right] \\ &= \frac{\partial}{\partial d} \left[\frac{-1}{2\sigma_y^2} (y - Td - f^1 c^1 - f^2 c^2)' (y - Td - f^1 c^1 - f^2 c^2) \right] \\ &= \frac{1}{\sigma_y^2} (T'y - T'Td - T'f^1 c^1 - T'f^2 c^2) \end{aligned} \quad (4.8)$$

- Dérivée de (4.3) suivant les paramètres $\{c^r\}_{r=1,2}$

$$\begin{aligned} \frac{\partial \ell}{\partial c^r} &= \frac{\partial}{\partial c^r} \left[\frac{-1}{2\sigma_y^2} \sum_{i=1}^n (y_i - d't_i - f_i^1 c^1 - f_i^2 c^2)' (y_i - d't_i - f_i^1 c^1 - f_i^2 c^2) \right] \\ &= \frac{1}{\sigma_y^2} \sum_{i=1}^n f_i^r (y_i - d't_i - f_i^1 c^1 - f_i^2 c^2) \\ &= \frac{1}{\sigma_y^2} ((f^r)' y - (f^r)' Td - (f^r)' f^1 c^1 - (f^r)' f^2 c^2) \end{aligned} \quad (4.9)$$

- Dérivée de (4.3) suivant les paramètres $\{a^r\}_{r=1,2}$

$$\begin{aligned} \frac{\partial \ell}{\partial a^r} &= \frac{\partial}{\partial a^r} \left[-\frac{1}{2\sigma_r^2} \sum_{i=1}^n (x_i^r - (D^r)'t_i^r - a^r f_i^r)' (x_i^r - (D^r)'t_i^r - a^r f_i^r) \right] \\ &= \frac{1}{\sigma_r^2} \left[\sum_{i=1}^n x_i^r f_i^r - \sum_{i=1}^n (D^r)'t_i^r f_i^r - \sum_{i=1}^n a^r (f_i^r)^2 \right] \\ &= \frac{1}{\sigma_r^2} [(X^r)' f^r - (D^r)' T^r f^r - a^r (f^r)' f^r] \end{aligned} \quad (4.10)$$

- Dérivée de (4.3) suivant les paramètres $\{D^r\}_{r=1,2}$

$$\begin{aligned}\frac{\partial \ell}{\partial (D^r)'} &= \frac{\partial}{\partial D^r} \left[-\frac{1}{2\sigma_r^2} \sum_{i=1}^n (x_i^r - (D^r)'t_i^r - a^r f_i^r)' (x_i^r - (D^r)'t_i^r - a^r f_i^r) \right] \\ &= \frac{1}{\sigma_r^2} \left[\sum_{i=1}^n x_i^r (t_i^r)' - \sum_{i=1}^n (D^r)'t_i^r (t_i^r)' - \sum_{i=1}^n a^r f_i^r (t_i^r)' \right] \\ \frac{\partial \ell}{\partial D^r} &= \frac{1}{\sigma_r^2} [(T^r)'X - (T_i^r)'T_i^r D^r - (T_i^r)'f^r(a^r)']\end{aligned}\quad (4.11)$$

- Dérivée de (4.3) suivant le paramètre σ_y^2

$$\begin{aligned}\frac{\partial \ell}{\partial \sigma_y^2} &= \frac{\partial}{\partial \sigma_y^2} \left[-\frac{1}{2} \ln(|\psi_y|) - \frac{1}{2\sigma_y^2} (y - Td - f^1 c^1 - f^2 c^2)' (y - Td - f^1 c^1 - f^2 c^2) \right] \\ &= -\frac{n}{2\sigma_y^2} + \frac{1}{2\sigma_y^4} (y - Td - f^1 c^1 - f^2 c^2)' (y - Td - f^1 c^1 - f^2 c^2)\end{aligned}\quad (4.12)$$

- Dérivée de (4.3) suivant les paramètres σ_r^2 où $r = 1, 2$

$$\begin{aligned}\frac{\partial \ell}{\partial \sigma_r^2} &= \frac{\partial}{\partial \sigma_r^2} \left[-\sum_{i=1}^n \frac{1}{2} \ln(|\psi_1|) - \sum_{i=1}^n \frac{1}{2\sigma_r^2} (x_i^r - (D^r)'t_i^r - a^r f_i^r)' (x_i^r - (D^r)'t_i^r - a^r f_i^r) \right] \\ &= -\frac{nq_r}{2\sigma_r^2} + \frac{1}{2\sigma_r^4} \sum_{i=1}^n (x_i^r - (D^r)'t_i^r - a^r f_i^r)' (x_i^r - (D^r)'t_i^r - a^r f_i^r)\end{aligned}\quad (4.13)$$

Les estimations du maximum de vraisemblance (EMV) du modèle sont données par l'équation (4.5). L'espérance conditionnelle de ces dérivées nécessite la connaissance de quatre espérances spécifiques. D'après l'équation (4.7) et $\forall r = \{1, 2\}$,

$$\begin{aligned}\tilde{f}^r &= \mathbb{E}_z^h(f^r) = m_r \\ \phi_{rr} &= \mathbb{E}_z^h((f^r)'f^r) = n\Gamma_{rr} + (m_r)'m_r \\ \phi_{12} &= \mathbb{E}_z^h((f^1)'f^2) = n\Gamma_{12} + (m_1)'m_2 \\ \phi_{21} &= \mathbb{E}_z^h((f^2)'f^1) = n\Gamma_{21} + (m_2)'m_1\end{aligned}$$

Pour le paramètre d , l'EMV \hat{d} qui annule (4.8) sachant les paramètres $(\hat{c}^r)_{r=1,2}$ est donnée par :

$$\begin{aligned}0 &= \mathbb{E}_z^h \left(\frac{\partial}{\partial d} \ell(\theta; z, h) \right) \\ &= T' (y - T\hat{d} - \tilde{f}^1 \hat{c}^1 - \tilde{f}^2 \hat{c}^2) \\ \Leftrightarrow \hat{d} &= [T'T]^{-1} [T'y - T'\tilde{f}^1 \hat{c}^1 - T'\tilde{f}^2 \hat{c}^2]\end{aligned}\quad (4.14)$$

L'EMV \widehat{c}^1 dépend de \widehat{c}^2 et de (4.14). Soit, d'après (4.9) :

$$\begin{aligned}
 0 &= \mathbb{E}_z^h \left(\frac{\partial}{\partial \widehat{c}^1} \ell(\theta; z, h) \right) \\
 &= \left(\widetilde{f}^1 \right)' y - \left(\widetilde{f}^1 \right)' T \widehat{d} - \phi_{11} \widehat{c}^1 - \phi_{12} \widehat{c}^2 \\
 &= \left(\widetilde{f}^1 \right)' y - \left(\widetilde{f}^1 \right)' T [T' T]^{-1} \left[T' y - T' \widetilde{f}^1 \widehat{c}^1 - T' \widetilde{f}^2 \widehat{c}^2 \right] - \phi_{11} \widehat{c}^1 - \phi_{12} \widehat{c}^2 \\
 &= \left(\widetilde{f}^1 \right)' y - M_1 y - M_1 \widetilde{f}^1 \widehat{c}^1 - M_1 \widetilde{f}^2 \widehat{c}^2 - \phi_{11} \widehat{c}^1 - \phi_{12} \widehat{c}^2,
 \end{aligned}$$

avec $M_1 = \left(\widetilde{f}^1 \right)' T [T' T]^{-1} T'$,

$$\widehat{c}^1 = \left[M_1 \widetilde{f}^1 + \phi_{11} \right]^{-1} \left[\left(\widetilde{f}^1 \right)' y - M_1 y - M_1 \widetilde{f}^2 \widehat{c}^2 - \phi_{12} \widehat{c}^2 \right]$$

En notant $M_2 = \left[M_1 \widetilde{f}^1 + \phi_{11} \right]^{-1}$, l'EMV \widehat{c}^1 est alors :

$$\widehat{c}^1 = M_2 \left(\widetilde{f}^1 \right)' y - M_2 M_1 y - M_2 M_1 \widetilde{f}^2 \widehat{c}^2 - M_2 \phi_{12} \widehat{c}^2 \quad (4.15)$$

L'EMV \widehat{c}^2 dépend quant à lui de \widehat{c}^1 et de \widehat{d} donnés respectivement par les équations (4.15) et (4.14). Soit, nous avons d'après (4.9) :

$$\begin{aligned}
 0 &= \left(\widetilde{f}^2 \right)' y - \left(\widetilde{f}^2 \right)' T \widehat{d} - \phi_{21} \widehat{c}^1 - \phi_{22} \widehat{c}^2 \\
 &= \left(\widetilde{f}^2 \right)' y - \left(\widetilde{f}^2 \right)' T [T' T]^{-1} \left[T' y - T' \widetilde{f}^1 \widehat{c}^1 - T' \widetilde{f}^2 \widehat{c}^2 \right] - \phi_{21} \widehat{c}^1 - \phi_{22} \widehat{c}^2 \\
 &= \left(\widetilde{f}^2 \right)' y - M_3 y - M_3 \widetilde{f}^1 \widehat{c}^1 - M_3 \widetilde{f}^2 \widehat{c}^2 - \phi_{21} \widehat{c}^1 - \phi_{22} \widehat{c}^2 \\
 &= \left(\widetilde{f}^2 \right)' y - M_3 y - M_3 \widetilde{f}^2 \widehat{c}^2 - \phi_{22} \widehat{c}^2 \\
 &\quad - M_3 \widetilde{f}^1 \left[M_2 \left(\widetilde{f}^1 \right)' y - M_2 M_1 y - M_2 M_1 \widetilde{f}^2 \widehat{c}^2 - M_2 \phi_{12} \widehat{c}^2 \right] \\
 &\quad - \phi_{21} \left[M_2 \left(\widetilde{f}^1 \right)' y - M_2 M_1 y - M_2 M_1 \widetilde{f}^2 \widehat{c}^2 - M_2 \phi_{12} \widehat{c}^2 \right]
 \end{aligned}$$

où $M_3 = \left(\widetilde{f}^2 \right)' T [T' T]^{-1} T'$ et en posant $M_4 = \left[M_2 \left(\widetilde{f}^1 \right)' y - M_2 M_1 y \right]$, nous obtenons :

$$\begin{aligned}
 \widehat{c}^2 &= \left[M_3 \widetilde{f}^2 + \phi_{22} - M_3 \widetilde{f}^1 M_2 M_1 \widetilde{f}^2 - M_3 \widetilde{f}^1 M_2 \phi_{12} - \phi_{21} M_2 M_1 \widetilde{f}^2 - \phi_{21} M_2 \phi_{12} \right]^{-1} \\
 &\quad \times \left[\left(\widetilde{f}^2 \right)' y - M_3 y - M_3 \widetilde{f}^1 M_4 - \phi_{21} M_4 \right]
 \end{aligned} \quad (4.16)$$

Sachant les équations (4.14), (4.15) et (4.16), l'expression de $\widehat{\sigma}_y^2$ est solution de l'équation qui annule $\mathbb{E}_z^h \left(\frac{\partial}{\partial \widehat{\sigma}_y^2} \ell(\theta; z, h) \right)$ et est donnée par :

$$\begin{aligned}
 \widehat{\sigma}_y^2 &= \left[y - T \widehat{d} \right]' \left[y - T \widehat{d} \right] - \left[y - T \widehat{d} \right]' \widetilde{f}^1 \widehat{c}^1 - \left[y - T \widehat{d} \right]' \widetilde{f}^2 \widehat{c}^2 \\
 &\quad - \widehat{c}^1 \left(\widetilde{f}^1 \right)' \left[y - T \widehat{d} \right] + \left(\widehat{c}^1 \right)^2 \phi_{11} + \widehat{c}^1 \widehat{c}^2 \phi_{12} \\
 &\quad - \widehat{c}^2 \left(\widetilde{f}^2 \right)' \left[y - T \widehat{d} \right] + \widehat{c}^1 \widehat{c}^2 \phi_{21} + \left(\widehat{c}^2 \right)^2 \phi_{22}
 \end{aligned} \quad (4.17)$$

Concernant les paramètres D^r et a^r et les dérivées partielles associées de la vraisemblance (4.11) et (4.10), leurs estimations respectives sont solutions du système suivant :

$$\begin{cases} 0 = (T^r)'X^r - (T^r)'T^r\hat{D}^r - (T^r)'\tilde{f}^r(\hat{a}^r)' & (4.18a) \\ 0 = (X^r)'\tilde{f}^r - (\hat{D}^r)'(T^r)'\tilde{f}^r - \hat{a}^r(\tilde{f}^r)'\tilde{f}^r & (4.18b) \end{cases}$$

En notant $M_T = [(T^r)'T^r]^{-1}$, l'EMV \hat{D}^r sachant \hat{a}^r est donné par (4.18a) tel que :

$$\hat{D}^r = M_T \left[(T^r)'X^r - (T^r)'\tilde{f}^r(\hat{a}^r)' \right] \quad (4.19)$$

$$\Leftrightarrow (\hat{D}^r)' = \left[(X^r)'T^r - \hat{a}^r(\tilde{f}^r)'T^r \right] M_T \quad (4.20)$$

En injectant (4.20) dans (4.18b), nous déduisons l'expression de \hat{a}^r :

$$\hat{a}^r = \left[(X^r)'\tilde{f}^r - (X^r)'T^r M_T (T^r)'\tilde{f}^r \right] \left[(\tilde{f}^r)'\tilde{f}^r - (\tilde{f}^r)'T^r M_T (T^r)'\tilde{f}^r \right]^{-1} \quad (4.21)$$

Pour finir, sachant les estimations (4.19) et (4.21), les EMV des paramètres de variances $\hat{\sigma}_r^2$ sont :

$$\begin{aligned} \hat{\sigma}_r^2 &= \frac{1}{nq_r} \sum_{i=1}^n \left(x_i^r - (\hat{D}^r)'t_i^r \right)' \left(x_i^r - (\hat{D}^r)'t_i^r \right) \\ &\quad + \frac{1}{nq_r} \sum_{i=1}^n \left((\tilde{f}_i^r + \gamma_{rr})\hat{a}^r \right)' \hat{a}^r \\ &\quad - \frac{1}{nq_r} \sum_{i=1}^n \left(x_i^r - (\hat{D}^r)'t_i^r \right)' \hat{a}^r \tilde{f}_i^r \\ &\quad - \frac{1}{nq_r} \sum_{i=1}^n \tilde{f}_i^r \hat{a}^r' \left(x_i^r - (\hat{D}^r)'t_i^r \right) \end{aligned}$$

On résume les estimations ci-dessous dans l'ordre d'implémentation dans l'algorithme :

$$\begin{aligned}
 \widehat{c}^2 &= \left[M_3 \tilde{f}^2 + \phi_{22} - M_3 \tilde{f}^1 M_2 M_1 \tilde{f}^2 - M_3 \tilde{f}^1 M_2 \phi_{12} - \phi_{21} M_2 M_1 \tilde{f}^2 - \phi_{21} M_2 \phi_{12} \right]^{-1} \\
 &\quad \times \left[(\tilde{f}^2)' y - M_3 y - M_3 \tilde{f}^1 M_4 - \phi_{21} M_4 \right] \\
 \widehat{c}^1 &= M_2 (\tilde{f}^1)' y - M_2 M_1 y - M_2 M_1 \tilde{f}^2 \widehat{c}^2 - M_2 \phi_{12} \widehat{c}^2 \\
 \widehat{d} &= [T' T]^{-1} \left[T' y - T' \tilde{f}^1 \widehat{c}^1 - T' \tilde{f}^2 \widehat{c}^2 \right] \\
 \widehat{\sigma}_y^2 &= [y - T \widehat{d}]' [y - T \widehat{d}] - [y - T \widehat{d}]' \tilde{f}^1 \widehat{c}^1 - [y - T \widehat{d}]' \tilde{f}^2 \widehat{c}^2 \\
 &\quad - \widehat{c}^1 (\tilde{f}^1)' [y - T \widehat{d}] + (\widehat{c}^1)^2 \phi_{11} + \widehat{c}^1 \widehat{c}^2 \phi_{12} \\
 &\quad - \widehat{c}^2 (\tilde{f}^2)' [y - T \widehat{d}] + \widehat{c}^1 \widehat{c}^2 \phi_{21} + (\widehat{c}^2)^2 \phi_{22} \\
 \widehat{a}^r &= \left[(X^r)' \tilde{f}^r - (X^r)' T^r M_T (T^r)' \tilde{f}^r \right] \left[(\tilde{f}^r)' \tilde{f}^r - (\tilde{f}^r)' T^r M_T (T^r)' \tilde{f}^r \right]^{-1} \\
 \widehat{D}^r &= M_T \left[(T^r)' X^r - (T^r)' \tilde{f}^r (\widehat{a}^r)' \right] \\
 \widehat{\sigma}_r^2 &= \frac{1}{nq_r} \sum_{i=1}^n \left(x_i^r - (\widehat{D}^r)' t_i^r \right)' \left(x_i^r - (\widehat{D}^r)' t_i^r \right) + \frac{1}{nq_r} \sum_{i=1}^n \left((\tilde{f}_i^r + \gamma_{rr}) \widehat{a}^r' \widehat{a}^r \right) \\
 &\quad - \frac{1}{nq_r} \sum_{i=1}^n \left(x_i^r - (\widehat{D}^r)' t_i^r \right)' \widehat{a}^r \tilde{f}_i^r - \frac{1}{nq_r} \sum_{i=1}^n \tilde{f}_i^r \widehat{a}^r' \left(x_i^r - (\widehat{D}^r)' t_i^r \right)
 \end{aligned}$$

Initialisation et convergence de l'algorithme

D'un point de vue computationnel, l'algorithme nécessite une initialisation des paramètres ($\theta^{[0]}$) avant d'entamer les itérations. Pour cela, nous considérons dans un premier temps les équations de régression (4.2a) et (4.2b) pour proposer les valeurs initiales des paramètres respectifs de la manière suivante :

- $\forall r \in \{1, 2\}$, les coefficients de régression $D^{r[0]}$ sont obtenus par une régression linéaire multiple telle que $X^r = T^r D^r + \tilde{\varepsilon}_r$;
- $\forall r \in \{1, 2\}$, les valeurs spécifiques des facteurs $f^{r[0]}$ proposées correspondent à la première composante principale d'une ACP réalisée sur $X^r - T^r D^r$ pour le bloc de mesures r ;
- $\forall r \in \{1, 2\}$, les valeurs $a^{r[0]}$ et $\sigma_r^{2[0]}$ sont déduites d'une régression linéaire multiple telle que $X^r = T^r D^r + f^{r[0]} a^{r[0]} + \varepsilon_r^{[0]}$.

Enfin, il manque l'initialisation des paramètres de l'équation (4.2c). Sachant les initialisations précédentes, $d^{[0]}$, $c^{r[0]}$ et $\sigma_y^{2[0]}$ sont obtenus via une régression simple telle que $y = td + c^{1[0]} f^{1[0]} + c^{2[0]} f^{2[0]} + \varepsilon_y^{[0]}$.

Un autre aspect de l'algorithme qui n'a pas encore été discuté, est le critère de convergence. Nous avons opté dans la programmation pour un critère d'arrêt basé

sur la stabilité des paramètres, tel que :

$$\sum_{p=1}^{n_\theta} \frac{|\theta^{*[t+1]}(p) - \theta^{*[t]}(p)|}{\theta^{*[t+1]}(p)} < \epsilon \quad (4.22)$$

où θ^* est le vecteur qui contient les valeurs scalaires des paramètres, de dimension $n_\theta = 5 + q_y + \sum_{r=1}^2 q_r (1 + q_{Tr})$ (nombre total de paramètres) et $\epsilon = 10^{-3}$. Si (4.22) est vérifiée, alors les estimations du maximum de vraisemblance sont $\theta^{*[t+1]}$. Il peut être également envisagé un critère basé sur la stabilité de la (log-)vraisemblance. Dans notre cas, il s'agirait de la stabilité de la (log-)vraisemblance complète (4.3).

4.2.2 Analyse de sensibilité par simulations

Dans la section précédente, le SEM (4.1) a été proposé de telle manière qu'il respecte conceptuellement la construction des questionnaires EORTC de HRQoL. Des résultats numériques sur données simulées ont été proposés pour une généralisation de ce modèle, dans lequel un troisième facteur, associé à un nouveau bloc de mesure, est expliqué à la place de y (Bry et al., 2015). Les auteurs ont montré une stabilité du modèle pour l'estimation des paramètres et la reconstruction des facteurs. Une analyse de sensibilité a également été réalisée afin d'étudier la qualité d'estimation suivant le nombre de sujets pris en compte et le nombre de variables observées par blocs. L'influence du nombre de sujets affecte plus intensément l'estimation du modèle que celle du nombre de variables observées par bloc.

Le modèle (4.1) étant un appauvrissement du modèle proposé par (Bry et al., 2015), nous proposons une étude de simulation pour confirmer ces résultats, et étudier la capacité du SEM à reconstruire les facteurs et l'influence des covariables sur cette dernière. Ainsi, nous proposons dans un premier temps une analyse de sensibilité qui se focalise sur la reconstruction des facteurs en étudiant l'influence des différents éléments du modèles (valeur des paramètres, taille de l'échantillon ou encore nombre d'observations par bloc de mesure). Dans un second temps, nous verrons l'impact de la prise en compte ou non d'une covariable associée à un groupe de traitement. Ces simulations permettront ainsi de mettre en évidence la similarité entre un facteur reconstruit sans la connaissance de covariables et la somme des observations supposées le refléter. Les valeurs des paramètres utilisées pour simuler les données ont été choisies suivant les estimations obtenues sur les données réelles issues d'un essai clinique utilisant le questionnaire EORTC QLQ-C30. Pour chaque étude de simulation, $N = 500$ jeux de données ont été simulés.

Sensibilité de l'estimation des paramètres

La Figure 4.3 montre la sensibilité de l'estimation des paramètres suivant la taille de l'échantillon (nombre de sujets n) considérée. Les tailles d'échantillon testées dans les simulations recouvrent la majorité des différentes tailles d'effectifs rencontrées en pratique. Tous les paramètres ne sont pas représentés. Le nombre de sujets considérés influe sur la variabilité des estimations des paramètres, mais les estimations sont centrées sur les valeurs qui ont permis de générer les données. Nous pouvons avoir une bonne confiance quant à l'estimation des paramètres.

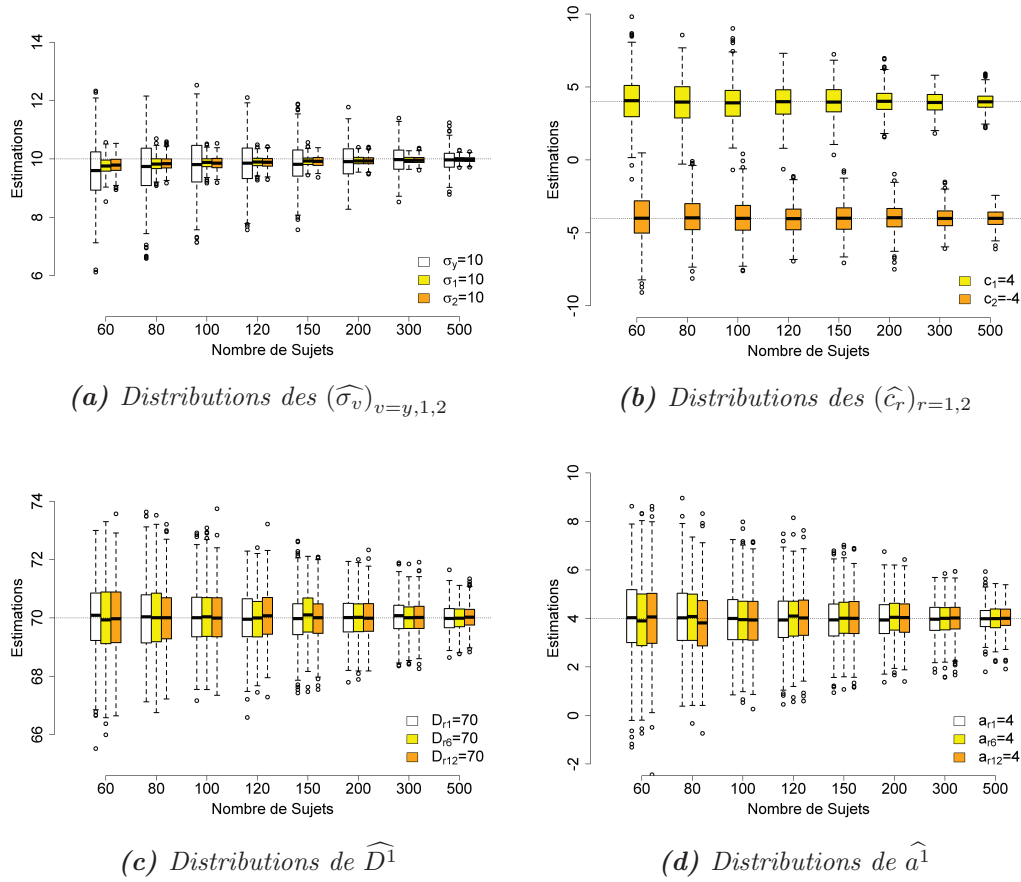


FIGURE 4.3: Sensibilité du modèle (4.1) en fonction de la taille de l'échantillon n . Les valeurs utilisées pour simuler les données sont : $(q_r = 12)_{r=1,2}$, $c_1 = -c_2 = 4$, $d = 80$, $\sigma_y = \sigma_r = 10$, $a_r = (4, 4, \dots, 4)$, $D^1 = (d_{11}, \dots, d_{1q_1})$ où $(d_{1s} = 70)_{s=1, \dots, q_1}$ et $D^2 = (d_{21}, \dots, d_{2q_2})$ où $(d_{2s} = 30)_{s=1, \dots, q_2}$.

Ensuite, la Figure 4.4 montre la sensibilité de l'estimation des paramètres du modèle suivant le nombre de variables $(q_r)_{r=1,2}$ par bloc. Les estimations sont stables quel que soit le nombre de variables observées par bloc. La qualité de l'estimation des paramètres semble être d'avantage influencée par le nombre de sujets que par le nombre de variables observées par bloc.

Sensibilité de la reconstruction des facteurs

Dans cette sous-section, nous étudions la sensibilité de la reconstruction des facteurs suivant les valeurs des paramètres du modèle (4.1), la taille de l'échantillon n et le nombre de variables observables par bloc. La Figure 4.5a montre une corrélation entre le facteur simulé (noté $(f^r)_{r=1,2}$) et le facteur estimé (noté $(\hat{f}^r)_{r=1,2}$) constante $\forall n \in \{60, \dots, 500\}$. Plus la taille de l'échantillon est grande, moins il y a de variabilité concernant la corrélation entre le facteur simulé et le facteur reconstruit des

4.2. Approche par modèles à équation structurelle

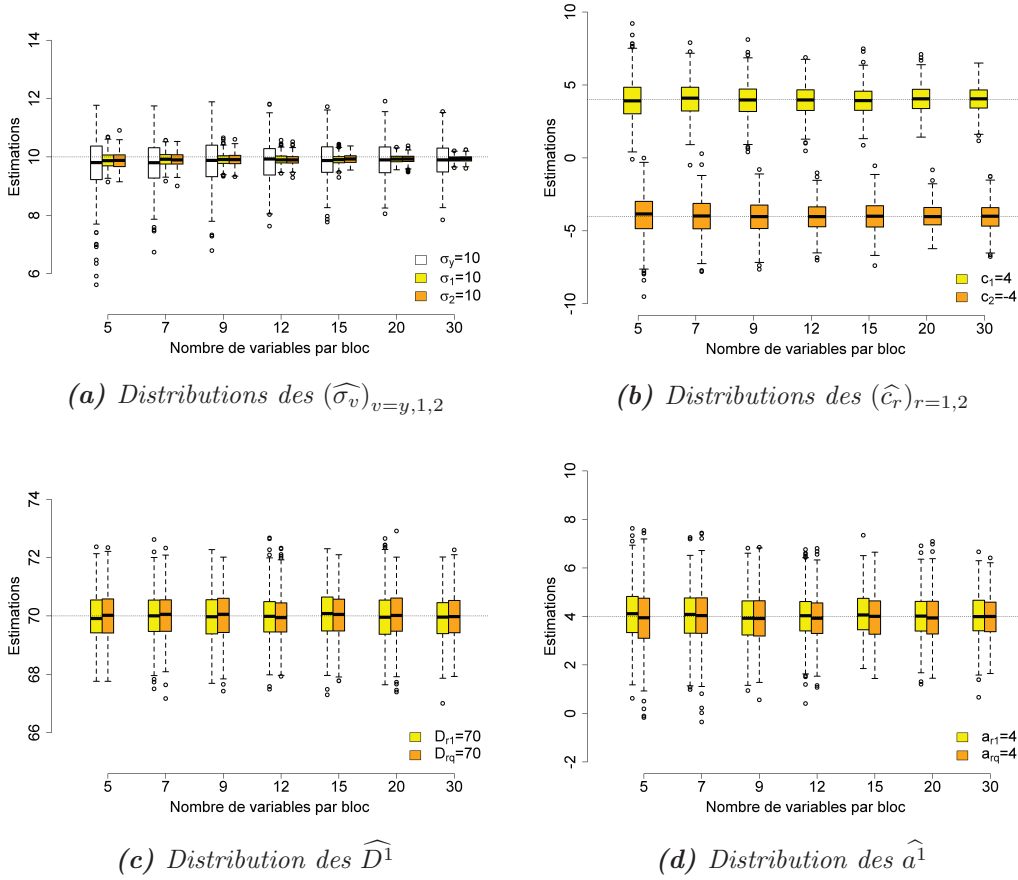


FIGURE 4.4: Sensibilité du modèle (4.1) en fonction du nombre d'observations (q_r) par bloc r ($r = 1, 2$). Les valeurs utilisées pour simuler les données sont : $n = 150$, $c_1 = -c_2 = 4$, $d = 80$, $\sigma_y = \sigma_r = 10$, $a_r = (4, 4, \dots, 4)$, $D^1 = (d_{11}, \dots, d_{1q_1})$ où $(d_{1s} = 70)_{s=1, \dots, q_1}$ et $D^2 = (d_{21}, \dots, d_{2q_2})$ où $(d_{2s} = 30)_{s=1, \dots, q_2}$.

$N = 500$ simulations. Le nombre $(q_r)_{r=1,2}$ d'observations par bloc de mesure influe sur la capacité du modèle à reconstruire le facteur $(f^r)_{r=1,2}$. En effet, la Figure 4.5b montre que, lorsque q_r augmente, la corrélation entre $(f^r)_{r=1,2}$ et $(\tilde{f}^r)_{r=1,2}$ augmente également. Ces résultats sont cohérents étant donné que plus il y a de variables pour refléter le facteur, plus il y a de l'information pour le reconstruire.

Les valeurs des paramètres ont bien entendu un impact sur la reconstruction des facteurs. En effet, les valeurs des paramètres $(a_r)_{r=1,2}$ sont importantes (voir Figure 4.6). Plus leurs coefficients sont grands, plus le facteur (source commune à l'ensemble des observations) a du poids pour expliquer les observations (Figure 4.6b). Le rapport entre la valeur des paramètres $(a_r)_{r=1,2}$ et les variances des erreurs $(\sigma_r^2)_{r=1,2}$ est logiquement important dans la reconstruction des facteurs. En effet, plus ce rapport est petit (beaucoup de bruit pour un petit signal), plus l'information des facteurs est "noyée". Ce rapport « signal sur bruit » a beaucoup moins d'influence concernant les paramètres $(c_r)_{r=1,2}$ et la variance σ_y de l'erreur de l'équation (4.2c). En effet, la

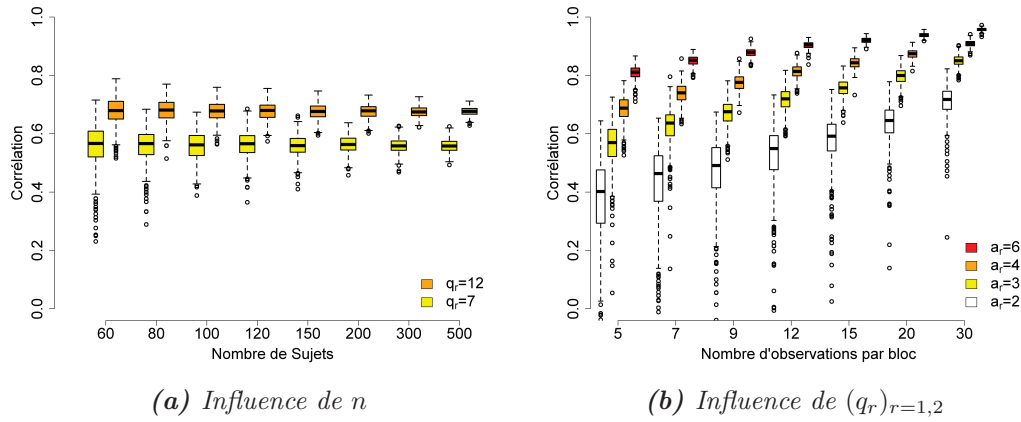
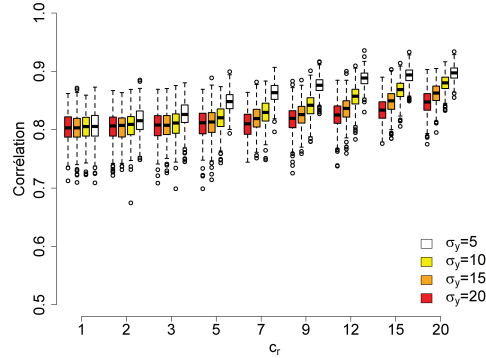


FIGURE 4.5: Corrélation entre les facteurs simulés pour générer les données et les facteurs reconstruits par le modèle (4.1), suivant la taille d'échantillon (n) et le nombre de variables observées par blocs (q_r). Les valeurs utilisées pour simuler les données sont : $\sigma_y = \sigma_1 = \sigma_2 = 10$, $c_1 = -c_2 = 4$, $d = 80$, $D^1 = (d_{11}, \dots, d_{1q_1})$ où $(d_{1s} = 70 + 2s)_{s=1, \dots, q_1}$, $D^2 = (d_{21}, \dots, d_{2q_2})$ où $(d_{2s} = 20 + 2s)_{s=1, \dots, q_2}$ et $a_r = (4, 4, \dots, 4)$. Pour la Figure 4.5a, $(q_r = 12)_{r=1,2}$ et pour la Figure 4.5b, $n = 150$.

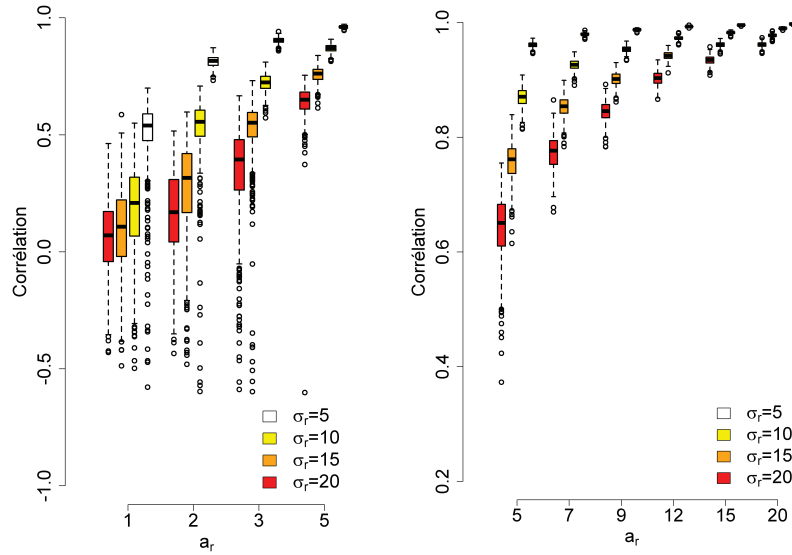
Figure 4.6a montre que l'influence de ces paramètres sur la reconstruction des facteurs est minime contrairement à celle des paramètres dans les équations de mesures (4.2a) et (4.2b). Ces résultats semblent tout à fait cohérents et valident indirectement la programmation. L'information portée par les variables des différents blocs de mesure est bien plus importante que celle de la variable réponse.

Influence d'un effet groupe sur les facteurs

La comparaison du niveau de HRQoL entre deux groupes de traitement est essentielle dans un essai clinique. Le SEM proposé dans ce chapitre permet de quantifier la différence en moyenne entre les scores de chaque groupe à chaque temps de mesure. Cependant, nous ne connaissons pas l'influence d'une différence entre deux groupes sur l'estimation des facteurs lorsque celle-ci n'est pas prise en compte en tant que variable explicative. En effet, si la covariable associée au groupe de traitement n'est pas prise en compte dans le modèle, l'estimation des facteurs est affectée par cette information. Afin d'étudier l'impact d'une éventuelle covariable, des simulations ont été réalisées avec des valeurs de paramètres similaires à celles utilisées précédemment. La nuance avec les simulations précédentes est dans la prise en compte de la variable explicative groupe dans les équations (4.2a), (4.2b) et (4.2c), qui se traduit notamment par $(q_{Tr})_{r=1,2} = q_t = 2$. Soit $g_i = 0$ si l'individu i appartient au groupe contrôle et $g_i = 1$ pour le groupe expérimental. Les premières colonnes de T et $(T^r)_{r=1,2}$ ne sont alors composées que de 1 et la seconde de 1 et 0 selon l'appartenance au groupe (le même nombre de patients est considéré par groupe). Les paramètres associées



(a) Influence de $(c_r)_{r=1,2}$



(b) Influence de $(a_r)_{r=1,2}$ (1)

(c) Influence de $(a_r)_{r=1,2}$ (2)

FIGURE 4.6: Corrélation entre les facteurs simulés pour générer les données et les facteurs reconstruits par le modèle (4.1). La Figure 4.6a montre l'influence des valeurs des paramètres c_r et σ_y , $\forall r \in \{1, 2\}$, avec $a_r = (4, 4, \dots, 4)$ et $\sigma_r = 10$. La Figure 4.6b montre l'influence des valeurs des paramètres a_r et σ_r , $\forall r \in \{1, 2\}$, avec $c_1 = -c_2 = 4$ et $\sigma_y = 10$. Les valeurs utilisées pour simuler les données sont : $(q_r = 12)_{r=1,2}$, $n = 150$, $d = 80$, $D^1 = (d_{11}, \dots, d_{1q_1})$ où $(d_{1s} = 70 + 2s)_{s=1, \dots, q_1}$ et $D^2 = (d_{21}, \dots, d_{2q_2})$ où $(d_{2s} = 20 + 2s)_{s=1, \dots, q_2}$.

sont alors notés tel que :

$$d = \begin{pmatrix} \mu_{0y} \\ \delta_y \end{pmatrix} ; \quad D^r = \begin{pmatrix} \mu_{01}^r \cdots \mu_{0q_r}^r \\ \delta_1^r \cdots \delta_{q_r}^r \end{pmatrix},$$

où μ_0 représente la moyenne du groupe $g = 0$ et $\delta = \mu_1 - \mu_0$ la différence en moyenne entre les deux groupes. Plusieurs différences entre les deux groupes ont été testées de manière uniforme sur l'ensemble des observations simulées : $\delta = \{0, 5, 10, 20, 30\}$. Ce

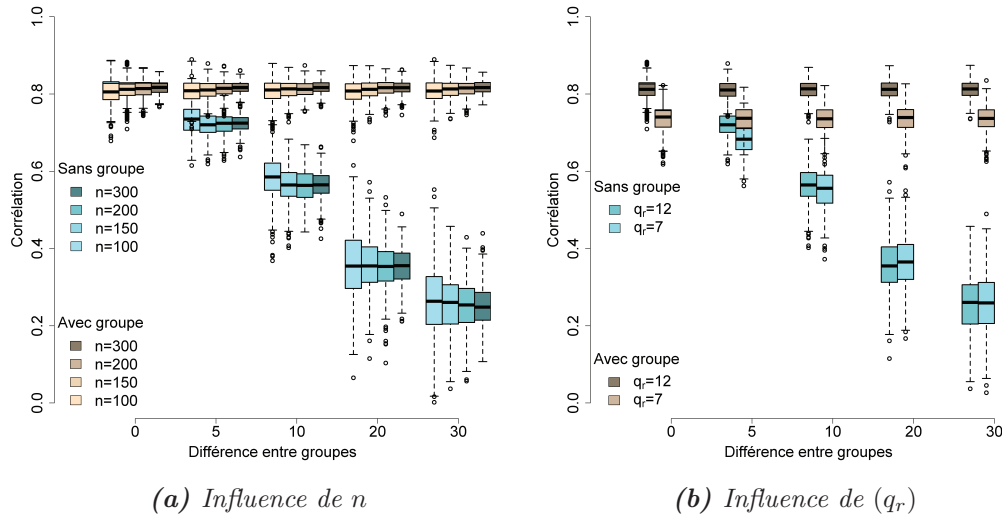


FIGURE 4.7: Corrélations entre les facteurs simulés f_r et les facteurs estimés par le SEM avec \tilde{f}_g^r ou sans \tilde{f}_g^r la considération de la covariable liée au groupe. Les différences testées sont $\delta = \{0, 5, 10, 20, 30\}$ et supposées uniforme sur l'ensemble des observations.

choix des valeurs est justifié par le fait que les MCID considérées dans les modèles de survie sont généralement de 5, 10 ou 20 points (sur une échelle de 0 à 100). En effet, [Osoba et al. \(1998\)](#) ont montré qu'un changement moyen de 5 à 10 points du score, correspond à un petit changement de HRQoL ; une différence moyenne de 10 à 20 points correspond à un changement modéré ; et enfin une différence de plus de 20 points correspond à fort changement. Cependant, une différence de 30 qui correspond à la différence entre deux modalités adjacentes pour une dimension uni-item à quatre modalités de réponse est également étudiée.

La Figure 4.7 montre la corrélation entre les facteurs simulés $(f^r)_{r=1,2}$ et les facteurs estimés en prenant en compte ou pas la variable associée au groupe (respectivement $(\tilde{f}_g^r)_{r=1,2}$ et $(\tilde{f}^r)_{r=1,2}$). Les influences du nombre de patients considérés et du nombre de variables dans les blocs de mesures sont également étudiées. Les conclusions sont identiques à la partie précédente : la variabilité de la distribution de la corrélation diminue en fonction de n et n'évolue pas en fonction de $(q^r)_{r=1,2}$. Cependant, la corrélation entre la vraie variable latente (facteur simulé) et celle estimée reste constante quelle que soit la différence entre les groupes pris en compte lorsque cette dernière est considérée dans le modèle. Enfin, plus la différence entre les groupes augmente, plus la corrélation diminue entre les facteurs simulés et les facteurs estimés sans la connaissance a priori du groupe.

La non prise en compte de la variable associée au groupe affecte également l'estimation des paramètres du modèles. L'estimation des paramètres d et $(D^r)_{r=1,2}$ change, ainsi que celle des paramètres directement liés aux facteurs $(c^r, a^r)_{r=1,2}$ (voir la Figure 4.8). Une part de l'information liée à l'effet groupe va être portée par le facteur et par conséquent modifier les estimations des paramètres associés. La diffé-

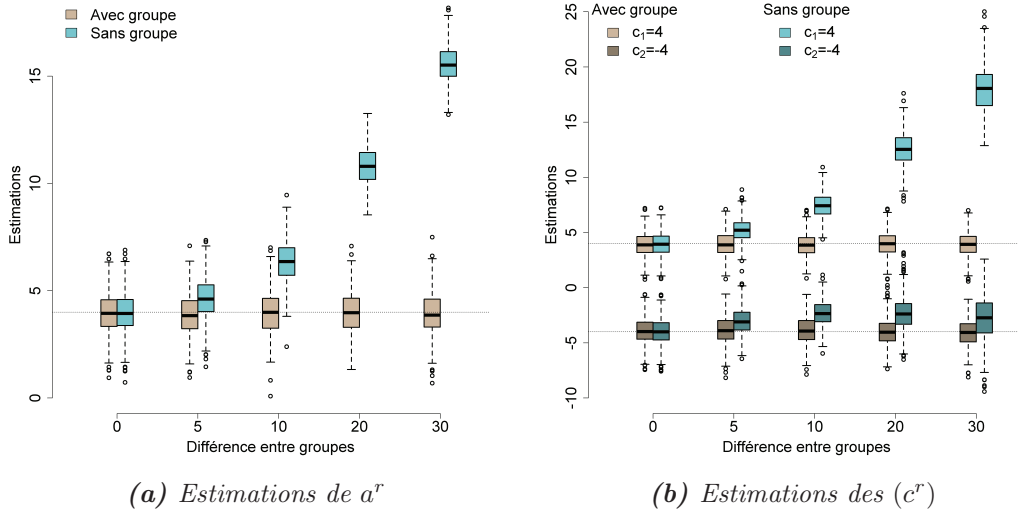


FIGURE 4.8: Estimation des paramètres a^r et c^r en considérant une différence uniforme sur l'ensemble des observations de $\delta = \{0, 5, 10, 20, 30\}$ entre deux groupes.

rence testée ($\delta \geq 0$) suggère que $\mu_1 \geq \mu_0$. Cela va se traduire par $\tilde{f}^r|_{g=1}$ ² supérieur en moyenne à $\tilde{f}^r|_{g=0}$, et $\tilde{f}^r_{g=1} < \tilde{f}^r|_{g=1}$ tout comme $\tilde{f}^r_{g=0} > \tilde{f}^r|_{g=0}$. Les paramètres (c^r, a^r)_{r=1,2} ont alors une tendance à augmenter et d'autant plus que la différence entre les groupes est grande.

Tous les résultats de simulations présentés jusqu'à maintenant montrent la robustesse du modèle avec les variations attendues dans les estimations des paramètres et des facteurs suivant la taille de l'échantillon et le nombre d'observation par bloc. Une des utilités du modèle réside également dans la possibilité de comparer la somme des observations par bloc et des facteurs sous-jacents du modèle. Cela permet de valider que le score³ est une bonne représentation du concept reflété par cet ensemble de variables. En effet, la Figure 4.9 montre que le « Score-somme » (SS) est très corrélé aux facteurs reconstruits par le SEM lorsque le groupe n'est pas compris dans le modèle ($\rho \approx 1$ quelque soit les valeurs de n, q_r et δ , Figures 4.9a et 4.9b). En revanche, plus la différence entre les deux groupes est grande, plus la corrélation entre SS et \tilde{f}^r est faible. Cela s'explique par le fait que l'influence de l'effet groupe sur SS et sur \tilde{f}^r ne permet plus de distinguer la vraie variable latente f sous-jacente aux observations. Son information est "noyée" par le signal de la covariable groupe et ne permet plus de capter la source individuelle qui renvoie une dépendance des mesures. En pratique, ces résultats sont intéressants pour l'analyse de la HRQoL. En effet, l'intérêt est de réduire les analyses et de pouvoir discriminer entre deux groupes suivant leur niveau de HRQoL. Sous cet objectif, les SS ou les facteurs associés aux différents blocs de mesure semblent être de bons candidats pour une analyse plus

2. $\tilde{f}^r|_{g=1}$: facteur estimé sans la connaissance de la variable groupe dans le SEM pour les patients du groupe expérimental ($g = 1$);

3. le score correspond à une somme d'un ensemble de variables. Il sera exprimé sur une échelle de 0 à 100 par souci d'homogénéité avec les autres scores du questionnaire de HRQoL;

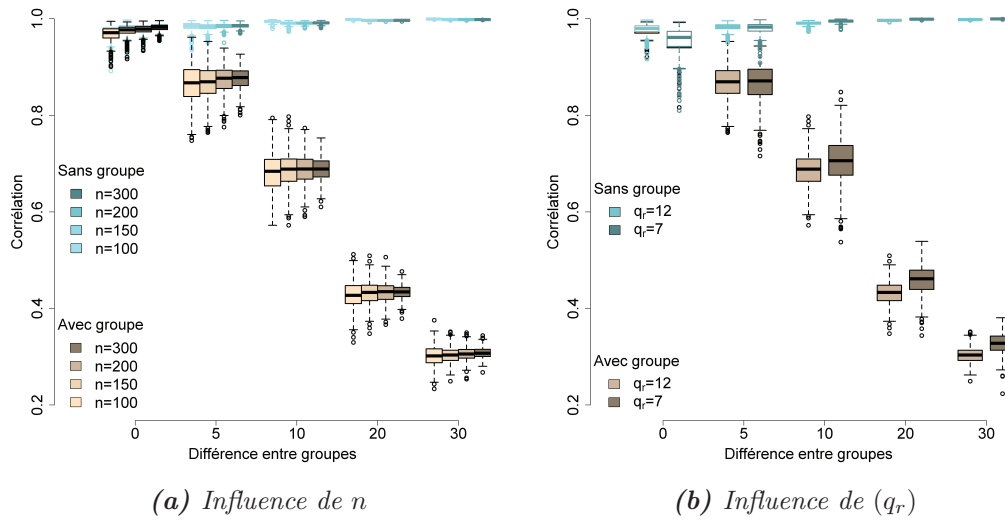


FIGURE 4.9: Corrélations entre le SS^r par bloc et les facteurs respectifs estimés avec \tilde{f}_g^r ou sans \tilde{f}_g^r la considération de la covariable liée au groupe. Les différences testées sont $\delta = 0, 5, 10, 20, 30$ et supposées uniforme sur l'ensemble des observations .

globale.

Dans le cas où une différence existe entre les deux groupes, l'information est portée par le SS^r mais également par \tilde{f}^r respectif. Par souci d'homogénéité, il est alors préférable de se ramener à la comparaison de \tilde{f}^r vs \tilde{f}_g^r qui sont à la même échelle, alors que le SS^r peut dépendre du calcul (somme exprimée sur une échelle de 0 à 100 ou simple somme mais dont l'échelle dépend alors du nombre de variables).

4.3 Application

En pratique, les données de HRQoL sont loin d'être normalement distribuées et nous avons proposé un modèle dont l'hypothèse de base est la normalité des données. Nous avons vu dans les chapitres précédents que faire cette hypothèse⁴ n'impacte pas les conclusions des analyses longitudinales. Sous l'hypothèse de normalité des scores de HRQoL nous avons appliqué le modèle (4.1) sur les données de HRQoL de l'essai de phase II CO-HO-RT (Azria et al., 2010) qui concernent $n = 149$ patientes atteintes d'un cancer du sein. L'objectif de cet essai est de comparer deux schémas d'administration du létrozole⁵ associé à une radiothérapie. Après l'intervention chirurgicale, les patientes pourront recevoir une chimiothérapie. Elles seront ensuite réparties de façon aléatoire en deux groupes. Les patientes du premier groupe ($n = 74$) recevront une hormonothérapie avec du létrozole quotidiennement pendant deux ans. Trois semaines après le début de ce traitement, les patientes débiteront également une radiothérapie (d'une durée de cinq semaines). Les patientes du deuxième groupe

4. hypothèse de « (grand) confort » ;

5. le létrozole inhibe l'enzyme *aromatase* qui est la principale source des œstrogènes dont dépend le développement tumoral dans le cancer du sein ;

($n = 75$) recevront les mêmes traitements mais la radiothérapie sera administrée avant le traitement par létrozole, ce dernier débutant trois semaines après la fin de la radiothérapie. Les patientes des deux groupes seront suivies tous les trois mois pendant deux ans.

Dans le contexte du cancer sein, un faible taux de données manquantes est observé⁶. Les mesures du module additionnel QLQ-BR23 (questionnaire spécifique au cancer du sein) complètent dans cet essai celles relevées avec le questionnaire QLQ-C30 (questionnaire générique à tous les cancers) et permettent de considérer un nombre plus élevé de dimensions et plus spécifiques à la localisation cancéreuse (voir Figure 4.1). Des analyses sur le critère de HRQoL ont déjà été réalisées de manière classique, c'est-à-dire dimension par dimension et de manière indépendante. Ces résultats n'ont montré aucune différence significative entre les deux groupes.

4.3.1 Résultats

Les résultats issus de cette approche par modèles à équations structurelles et à effets aléatoires sur les données de l'essai CO-OH-RT sont présentés dans l'article en section 4.5. Ces résultats n'ont pas apporté de nouvelles informations concernant l'analyse longitudinale de la HRQoL : aucune différence de HRQoL suivant les deux bras de traitement n'a été mise en évidence. Dans cet article, nous présentons les coefficients liés aux éléments $(D^r)_{r=1,2}$ et d du modèle (4.1). Les Tables 4.1 et 4.2 présentent respectivement les coefficients estimés des paramètres $(a^r)_{r=1,2}$ et $(\sigma_y, \sigma_1, \sigma_2, c_1, c_2)$. D'après les résultats de l'étude de simulation, l'estimation des facteurs est cohérente. En effet, les paramètres $(a^r)_{r=1,2}$ sont supérieurs à 10 en moyenne et les valeurs des écarts types des erreurs se situent en moyenne entre 15 et 20. De plus, la taille des effectifs par visite est en moyenne proche de 100 patientes et le nombre de variables observées par bloc est $q_1 = 7$ et $q_2 = 12$. Ces valeurs suggèrent une information suffisante portée par les facteurs estimés : forte corrélation (> 0.8) entre les facteurs estimés et la « vraie » variable latente (voir Figure 4.6).

Sur cet exemple concret, le SEM permet également de comparer l'information apportée par les facteurs et celle apportée par la somme des observations par bloc. La Figure 4.10 illustre la corrélation entre les facteurs $(\tilde{f}^r)_{r=1,2}$ et les sommes $(SS^r)_{r=1,2}$ des observations pour chaque bloc de mesure sur l'ensemble des visites. Les sommes des scores sont standardisées pour être exprimées sur une échelle de 0 à 100. La corrélation entre le facteur et la somme des observations pour chaque bloc de mesure est supérieure à 0.95 pour chaque bloc de mesure. Cela suggère que la somme des scores SS^r est une bonne représentation de la variable latente sous-jacente f^r , à condition qu'aucune variable explicative n'exerce une influence sur les mesures. Sans la prise en compte de variables explicatives, le facteur est aussi informatif que la somme des variables observées qui le reflète. Ce résultat est logique et découle de la conception du SEM. Il confirme également les résultats de l'étude de simulation.

Concernant l'estimation des facteurs, nous pouvons observer sur la Figure 4.10 que leur distribution n'est pas gaussienne. Cela est causé par le fait que les scores de

6. ce qui arrange étant donné que le programme ne prend pas (encore) en compte les données manquantes ;

Chapitre 4. Étude globale et longitudinale par modèles à équation structurelle et à effets aléatoires

TABLE 4.1: Estimations des paramètres $(a^r)_{r=1,2}$ du modèle (voir 4.2a) à chaque visite $t_v = (0, 3, 6, 12, 15, 18, 21, 24)$ (en mois). La covariable liée au groupe de traitement a été prise en compte dans chaque équation. La moyenne des paramètres est celle sur l'ensemble des paramètres pour une même visite.

t_v	r	Dimensions j												Moyenne
		1	2	3	4	5	6	7	8	9	10	11	12	
0	a_{j1}^1	9.6	13.5	15.0	11.3	11.6	10.4	24.2	-	-	-	-	-	13.6
	a_{j2}^2	16.8	4.9	16.5	13.2	13.8	8.7	12.9	5.2	7.2	14.8	7.2	9.1	10.9
3	a_{j1}^1	9.4	15.5	18.4	12.5	12.9	10.5	22.2	-	-	-	-	-	14.5
	a_{j2}^2	17.2	4.2	22.2	12.5	23.4	4.9	12.1	3.7	9.5	9.4	11.9	12.1	11.9
6	a_{j1}^1	13.9	14.8	16.1	14.9	16.5	11.0	17.4	-	-	-	-	-	14.9
	a_{j2}^2	18.1	2.5	22.7	15.5	16.1	9.3	9.2	6.2	8.9	11.0	10.2	13.5	11.9
12	a_{j1}^1	16.6	19.7	18.2	9.0	14.5	13.0	15.9	-	-	-	-	-	15.3
	a_{j2}^2	21.6	4.6	21.6	15.7	26.1	10.6	12.8	4.7	6.4	11.7	7.4	13.0	13.0
15	a_{j1}^1	12.0	19.2	18.9	14.3	16.6	12.9	21.9	-	-	-	-	-	16.6
	a_{j2}^2	19.6	3.7	19.5	17.7	26.4	7.4	4.3	4.3	2.3	8.9	8.3	8.9	10.9
18	a_{j1}^1	12.8	11.2	17.5	9.8	13.3	12.7	20.9	-	-	-	-	-	14.0
	a_{j2}^2	18.4	1.7	16.5	11.8	22.1	6.6	8.1	7.1	6.9	8.8	7.8	11.5	10.6
21	a_{j1}^1	11.2	15.7	18.0	14.4	13.9	12.8	19.3	-	-	-	-	-	15.1
	a_{j2}^2	16.7	3.8	17.3	15.5	25.9	6.2	7.8	4.6	6.9	10.2	10.3	12.5	11.5
24	a_{j1}^1	16.5	19.2	17.8	16.9	17.6	16.7	23.0	-	-	-	-	-	18.2
	a_{j2}^2	19.1	2.9	25.5	18.6	19.6	9.9	8.1	5.9	11.9	12.0	9.2	14.9	13.1

TABLE 4.2: Estimations des paramètres $(\sigma_m)_{m=y,1,2}$ et $(c_r)_{r=1,2}$ du modèle (4.1) pour chaque visite (en mois). La covariable liée au groupe de traitement a été prise en compte dans chaque bloc de mesure X^1 et X^2 . La moyenne des paramètres est celle sur l'ensemble des visites.

	Visites (en mois)								Moyenne
	0	3	6	12	15	18	21	24	
σ_y	16.8	14.7	14.1	16.2	15.8	16.4	16.4	16.1	15.8
σ_1	17.5	17.0	17.5	18.4	16.8	16.5	16.6	17.5	17.2
σ_2	17.7	18.3	18.1	18.4	16.5	17.2	16.3	17.2	17.5
c_1	6.5	10.3	8.0	3.6	7.6	7.6	5.4	11.4	7.6
c_2	-8.1	-7.9	-8.9	-13.8	-7.7	-5.4	-7.9	-4.7	-8.0

HRQoL pris en compte dans les blocs de mesures ne sont pas normalement distribués.

4.3.2 Compléments sur les données réelles

Un des intérêts non négligeable du SEM proposé est de comparer les différences de niveau de HRQoL entre les deux groupes de patients ponctuellement à chaque visite et pour chacune des dimensions HRQoL. Cependant, les distributions des scores entre dimensions sont hétérogènes et ne permettent pas cette comparaison via la simple étude des coefficients associés $\{D^r\}_{r=1,2}$. Dans le but de les rendre comparables, nous allons homogénéiser les variances pour chaque dimension de HRQoL. Un second intérêt à la stabilisation des variances est méthodologique. En effet, dans le SEM

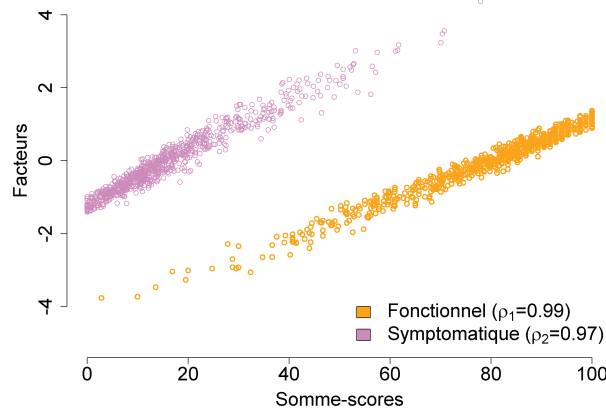


FIGURE 4.10: Scatterplots des facteurs fonctionnels (\tilde{f}_i^1) et des facteurs symptomatiques (\tilde{f}_i^2) en fonction respectivement des sommes des scores fonctionnels (SS_i^1) et des sommes des scores symptomatiques (SS_i^2). $(\rho_r)_{r=1,2}$ dénote la corrélation entre les deux variables.

implémenté, les variances des variables d'un même bloc de mesure sont supposées égales.

Afin de vérifier les conséquences de l'homogénéisation de la variabilité des données, nous considérons les données associées à la visite à 12 mois. Seules les données du bloc fonctionnel subissent la transformation, car une différence entre les groupes est visible uniquement pour ce bloc (voir \widehat{D}^1 article en section 4.5). Lorsque les variances sont homogènes, les résultats sont similaires sur ce jeu de donnée. En effet, la Table 4.3 montre que les estimations de δ ne changent quasiment pas. Cela est dû au fait que les variances des différents scores fonctionnels sont proches et que la différence entre les deux groupes est faible. Aux vues de ces résultats, les dimensions fonctionnelles physique, personnelle et la perception du futur semblent les plus impactées par l'effet groupe. En effet, une meilleure capacité fonctionnelle est visible pour le groupe ayant reçu son traitement de manière concomitante ($g_i = 0$). Par exemple, pour la dimension physique, le groupe ayant reçu son traitement de manière séquentielle ($g_i = 1$) a un score plus faible d'environ 5.6 points en moyenne.

4.4 Discussion

Le SEM proposé est motivé par la volonté de réduire le nombre d'analyses spécifiques pour l'étude du critère de HRQoL dans les essais cliniques. L'intérêt principal du SEM est d'étudier globalement la HRQoL à partir des liens structurels utilisés dans la modélisation. Les différents liens de ce modèle répondent à la conception des questionnaires EORTC. Pour cela, la HRQoL globale (GHS) est expliquée par l'intermédiaire de deux facteurs (caractérisant respectivement le statut fonctionnel et le statut symptomatique du patient) qui résument deux ensembles de variables

Chapitre 4. Étude globale et longitudinale par modèles à équation structurelle et à effets aléatoires

TABLE 4.3: Estimations des paramètres D^1 du modèle (voir 4.2a). La covariable liée au groupe de traitement a été prise en compte dans chaque équation du modèle.

Variances	Paramètres	Dimensions fonctionnelles						
		PF2	RF2	EF	CF	SF	BRBI	BRFU
Hétérogènes	μ_0	80.83	81.94	73.21	79.86	90.62	86.46	70.14
	δ	-5.74	-6.62	-3.65	2.19	-2.16	-3.12	-7.96
Variances scores fonctionnelles		558.21	731.17	593.44	460.27	424.55	622.03	762.74
Homogènes	μ_0	80.93	81.60	73.21	79.71	90.83	86.42	69.65
	δ	-5.91	-5.97	-3.65	2.49	-2.56	-3.05	-7.02
Variances scores fonctionnelles		593.26	593.26	593.26	593.26	593.26	593.26	593.26

observables : les scores fonctionnels et les scores symptomatiques. Une étude de simulation a permis de vérifier la programmation ainsi que les performances du modèle estimé par l'intermédiaire de l'algorithme EM. Des simulations complémentaires ont également mis en évidence l'influence des variables explicatives dans la reconstruction des facteurs. Dans l'article 4.5, nous avons montré, qu'en présence de variables explicatives, ce modèle permet à chaque temps de mesure d'évaluer leur impact sur le score de HRQoL pour chaque dimension considérée. Cela a l'avantage de donner une description globale sur l'ensemble des données, et spécifique à chacune des dimensions. De plus, concernant l'hypothèse des variances homogènes entre les variables d'un même bloc de mesure, homogénéiser les variances à travers un traitement au préalable permet d'une part de respecter l'hypothèse fixée dans le modèle sans pour autant impacter l'information des facteurs estimés. D'autre part, cela rend directement comparable l'impact des variables explicatives entre les variables du même bloc.

L'introduction de variables explicatives dans le modèle cloisonne leur information aux paramètres associés. En revanche, sans leur prise en compte, les facteurs portent leur information à la même hauteur que les score-sommes. Dans ce cas, l'utilisation des facteurs ou des score-sommes pour une analyse longitudinale concernant les statuts fonctionnel et symptomatique semble similaire. Or, si l'objectif est de discriminer l'évolution de la HRQoL entre deux bras de traitement et que des variables explicatives sont à disposition, l'utilisation des facteurs a un avantage certain comparée à celle des score-sommes. L'introduction dans le SEM des variables explicatives hormis celle liée à l'appartenance du bras de traitement, va permettre l'estimation des facteurs qui véhiculeront l'information de l'effet traitement et celle du réel facteur. Ainsi, leurs estimations ne seront pas "bruitées" par l'information des autres variables explicatives. En effet, les résultats ont permis de montrer qu'un effet global du groupe de traitement se répercute sur l'estimation des facteurs seulement si la covariable groupe n'est pas incluse dans le SEM. L'analyse longitudinale de la HRQoL peut alors être réduite à celles des deux facteurs et du GHS.

Hormis le fait de considérer les scores de HRQoL comme normalement distribués une limite au modèle proposé se situe sur les liens structurels. D'un point de vue méthodologique, les facteurs ont été considérés comme indépendants, en partie pour

4.5. Article : *EM algorithm estimation of a structural equation model for the longitudinal study of the quality of life*

faciliter l'approche. Or, une relation de dépendance existe entre le statut fonctionnel et le statut symptomatique. En effet, un patient ayant de forts symptômes va voir sa capacité fonctionnelle logiquement impactée. Cela se traduit en pratique par le fait que les deux facteurs estimés sont fortement corrélés. Les résultats présentés dans l'article 4.5 montrent une corrélation entre le facteur fonctionnel et le facteur symptomatique inférieure à -0.8 . Un lien structurel entre les deux facteurs représenterait mieux la dépendance des données analysées.

Enfin, comme mesure de la HRQoL globale des patients expliquée dans le modèle, nous avons utilisé la dimension relative au GHS du questionnaire générique EORTC QLQ-C30. Cette dimension est toutefois controversée dans la littérature et il est conseillé de l'utiliser avec prudence (Phillips et al., 2015). De manière longitudinale, le GHS est susceptible d'être en désaccord avec les autres dimensions. Cette dimension est sujette au phénomène de « Response Shift » (Schwartz and Sprangers, 1999). De plus, le fait qu'elle soit positionnée à la fin du questionnaire accentue le fait que les réponses soient de moins bonnes qualités. Plusieurs scores de substitution ont été proposés sur la base d'une somme de tous les scores du questionnaire au même titre que le score global du FACT-G. Par exemple, Nordin et al. (2001) proposent de considérer un score global exprimé dans l'intervalle $[-800; 600]$. Il correspond à la différence entre la somme des scores des différentes dimensions fonctionnelles et du GHS (où 100 est le meilleur score), et la somme des dimensions symptomatiques exceptée la dimension associée aux difficultés financières (où 0 est le meilleur score). Une donnée manquante est renvoyée si au moins un des scores d'une dimension est manquant. Plus récemment, Hinz et al. (2012) ont proposé un autre score global comme étant la moyenne des trente items du questionnaire, exprimée sur un score de 0 à 100 avec un inversement de l'échelle pour les items des dimensions fonctionnelles. Ce score a l'avantage de prendre en compte une certaine pondération entre les différentes dimensions, celles construites à partir de plusieurs items contribuent alors plus intensivement que les autres. Il paraît être une meilleure alternative au score proposé par Nordin et al. (2001). Cependant, considérer de tels scores dans le SEM que nous proposons n'est pas pertinent étant donné que les facteurs expliqueraient un score construit à partir des données qu'ils résument. L'utilisation du GHS est alors préférable dans notre modèle.

4.5 Article : *EM algorithm estimation of a structural equation model for the longitudinal study of the quality of life*

EM algorithm estimation of a structural equation model for the longitudinal study of the quality of life

Antoine Barbieri^{*,1,2,3}, Myriam Tami^{*,1,3}, Xavier Bry^{1,3}, David Azria²,
Sophie Gourgou², Caroline Mollevi² and Christian Lavergne^{3,4}

Abstract

The health-related quality of life data is measured through self-questionnaires filled up at different times. We focused on the oncology data reported through the EORTC questionnaires which decompose the health-related quality of life into several functioning dimensions, several symptomatic dimensions and the Global Health Status. The aim is to explain the latter, which represents the most general concept, through the other dimensions. First, a similar structural equation model is used at each time, in which the global health status is explained by two latent variables. Each latent variable is a factor which summarizes respectively the functional dimensions and the symptomatic dimensions. This is achieved through the maximization of the likelihood of each structural equation model using the EM algorithm, with the advantage to give an estimation of the subject-specific factors. Then, to consider the longitudinal aspect, the global health status variable and the two factors are concatenated for each visit. The global health status can be then explained by the two factors estimated in the first step and additional explanatory variables using a linear mixed model. This model takes into account the inter-subject variability via specific-subject random effects and other covariates such as the treatment.

Keywords: EM algorithm, Structural equation modeling, Longitudinal analysis, Mixed models, Health-related quality of life, Oncology data.

1 Introduction

The study of the quality of life concerns a lot of research domains such as social and medical sciences. The subjective, multidimensional and dynamic features of this concept makes it difficult to measure and analyze. Nowadays, the measure of the concept has become relatively standard through the use of specific psychometric questionnaires which take all the features into account. Indeed, this kind of questionnaire decomposes the concept into several (sub-)dimensions measured through a grouping of one or several objective questions, called items. The patient-reported outcomes are increasingly used to support decision making aiming at clinical benefits. In oncology, the health-related quality of life (HRQoL) endpoint is essentially to improve patients' care and better evaluate the impact of treatments on their everyday life (Fiteni et al., 2014). The HRQoL could be considered a primary endpoint in certain situations like geriatric or palliative situations. In Europe, the questionnaires are developed by

*Co-primary author: Antoine.Barbieri@umontpellier.fr & Myriam.Tami@umontpellier.fr

¹Université de Montpellier, Montpellier, France

²Institut régional du Cancer Montpellier (ICM) - Val d'Aurelle, Montpellier, France

³Institut Montpelliérain Alexander Grothendieck, Montpellier, France

⁴Université Paul-Valéry Montpellier 3, Montpellier, France

the European Organization for Research and Treatment of Cancer (EORTC). The standard one is the EORTC QLQ-C30 [Aaronson et al. \(1993\)](#) and it decomposes the HRQoL into five functional dimensions, nine symptomatic dimensions and the Global Health Status (GHS). In clinical trials, patients fill out the generic HRQoL questionnaire on every visit: inclusion, during the treatment and the follow-up. The repeated measurements over time give the longitudinal aspect of their HRQoL level for each dimension.

If the measurement tools are similar across the application fields, there is a strong heterogeneity as to the statistical model used for the longitudinal analysis ([Blanchin et al., 2011](#); [Anota et al., 2014](#); [Gorter et al., 2015](#)). The item responses are built from a Likert scale and then the analyses are performed on dimensions (multiple-item ordinal data). In literature, two of the key statistical approaches are: the item response theory (IRT) and the classical test theory (CTT) ([Boeck and Wilson, 2004](#); [De Ayala, 2009](#)). In the former approach, the IRT models take into account the raw data directly ([Rasch, 1961](#); [Samejima, 1968](#); [Masters, 1982](#); [Tutz, 1990](#)). The corresponding models are generalized linear mixed models for ordinal data with specific parametrization of the linear predictor. The main assumption is that there exists an unidimensional latent variable that contains all of the information which is common to each item ([Douglas, 1999](#)) (multidimensional responses). The distinction between the responses relative to different items is made through item parameters in the linear predictor. This kind of models has lot of success in the psychometric and social fields and its interest is growing in the medical domain. Not only does it offer the advantage of taking into account the raw data, but it also allows to further analyze every specific item ([Edelen and Reeve, 2007](#)). The second approach is based on score-study. The use of a scoring procedure by dimensions is required. The underlying philosophy is that the score is close to the real concept level up to some error. The use of a quantitative summary variable is common practice because so far, classical statistical methods for quantitative variables are more powerful and easier to implement and interpret than those conventionally applied to qualitative variables ([Grilli and Rampichini, 2011](#)). The associated classical model is the Linear Mixed Model (LMM) which allows to take into account the influence of both covariates and the variability within data through random effects. The IRT approach is scientifically more rigorous and informative than the score-study one ([Gorter et al., 2015](#); [Edelen and Reeve, 2007](#)), but the comparison studies did not show different conclusions across the different kind of mixed regression models used ([Anota et al., 2014](#)). In this paper, we focused on score-study for the HRQoL analysis in oncology medical research.

Nowadays, the HRQoL study is being carried out through an independent analysis on each HRQoL dimension. However, this leads to a problem of multiple testing and even more so if supplementary modules by cancer location are used adding other specific functional and symptomatic dimensions. Indeed, the influence of each covariate is tested for each HRQoL dimension separately. A way to get around that in the clinical trials is to take into account only some specific HRQoL dimensions and not all of them, the choice of the dimensions depending from pathology and their clinical pertinence for the trial. The GHS is always included in the chosen dimensions.

In questionnaire-data study in medical research, the use of models involving latent variables is common practice ([Bentler and Stein, 1992](#); [Rabe-Hesketh and Skrondal, 2008](#)). This began with IRT models and structural equation modeling has picked up momentum over time ([Lei and Wu, 2007](#); [Titman et al., 2013](#)), in particular for the HRQoL study ([Huang et al., 2010](#); [King-Kallimanis et al., 2011](#)). A structural equation model formalizes the dependence links of observed numeric variables through fewer unobserved ones, call latent variables. Every latent variable is assumed to be underlying a specific set of observed variables and summarize its information. Literature widely presents two competing families of methods that deal with structural equation models: factor-methods and component-methods. The first

family is based on maximum likelihood estimation and was first developed by Jöreskog (Joreskog, 1970) and implemented in the LISREL software. This method terms factors as latent variables merely assumed to be standardized normal and reflected by the observed variables. Besides, they base the estimation of their parameters (linear predictor coefficients and error-variances) on the structure of the covariance matrix of the data according to the model. But, they don't estimate the values of the factors. The component-methods family was initially developed by Wold and named PLS (for Partial Least Squares) (Wold, 1985, 1966; Noonan and Wold, 1977). This approach assumes that each latent variable is a component, i.e a linear combination of the corresponding observed variables. This assumption, stronger than the mere distribution assumption on factors, allows to easily estimate the latent variable. In many areas, it is of essence to be able to estimate the value of latent variables on subject units, since these values allow to efficiently analyze their disparities on a reduced number of dimensions. Therefore, we are interested in estimating these values but without the component constraint. So, we developed a new approach Bry et al. (2016) based on an EM-algorithm estimation of a structural equation model. In this approach, the structural equation model parameters and the factors values are estimated with only a normal-distribution assumption on factors.

In this work, we propose a global approach taking into account the multidimensional nature of data at each time and the longitudinal aspect induced by these repeated measures. The aim is to explain the GHS (representing the general concept of the HRQoL) using the information from all HRQoL functional and symptomatic dimensions as well as additional explanatory variables. To link all observations made at a given time, a structural equation model is used on data after a pre-treatment carrying out the scoring-procedure proposed by EORTC (Fayers et al., 2001). The EM algorithm (Dempster et al., 1977) is used to estimate the model parameters and factors. On the other hand, the longitudinal aspect is treated through the LMM in the same way as for the above-mentioned CTT approach. The latent variables estimated through the previous structural equation models at all times are taken into account in the same way as other observable variables which can influence the HRQoL evolution. The first section presents the methodology details of the first step, associated with the structural equation models. Then, the second section shows the LMM in this specific context. Finally, this new approach is demonstrated on real data from cancer clinical data.

2 A specific structural equation model for transversal way

A structural equation model has two types of equations, named measurement equations and structural equations. Each measurement equation relates a latent variable to the corresponding set of observed variables. A structural equation formalizes an assumed relationship between latent variables. In this part, the data feeding the structural model are taken at a specific time. The construct of the structural model proposed in this section results from the EORTC QLQ-C30 (Fayers et al., 2001) decomposition, which differentiates the functional dimension family, the symptomatic dimension family and the GHS.

This decomposition suggests a model built upon two factors f^1 and f^2 that quantify and summarize respectively the individual's overall functioning health (functional status) and the individual's overall symptomatic health (symptomatic status). Thus, f^1 and f^2 are respectively linked to the observed functional variables family X^1 and the observed symptomatic variables family X^2 , which respectively reflect the functional status and the symptomatic status. More precisely, X^1 and X^2 are quantitative observed variables obtained from the responses to the questionnaire after performing a scoring procedure. Besides,

f^1 and f^2 explain the quantitative observed variable y (GHS) obtained after the scoring procedure.

2.1 The model and notations

The data consists in blocks of observed variables describing the same n subjects. We consider the following data-matrices and denote:

$X^m = \{x_i^{j,m}\}; i \in \llbracket 1, n \rrbracket, j \in \llbracket 1, q_m \rrbracket, m \in \llbracket 1, 2 \rrbracket$ is the $n \times q_m$ matrix coding the m^{ieth} explanatory block of observed variables x^1, \dots, x^{q_m} . The value of variable $x_i^{j,m}$ for subject i is denoted $x_i^{j,m}$. Variable-blocks refer the corresponding matrices.

$y = \{y_i\}; i \in \llbracket 1, n \rrbracket$, is the n -length vector¹ coding the observed variable depending on factors f^1 and f^2 .

T (resp. T^1, T^2) refers to a $n \times r_T$ (resp. $n \times r_1, n \times r_2$) matrix of covariates.

We assume that:

- The n subjects, hence the rows of vector y and matrices X^1, X^2 are independent and are multi-variate normal vectors.

The structural equation model we handle here consists of three equations and is used to construct the factors \tilde{f}^1, \tilde{f}^2 for each visit. It is graphed on Figure 1.

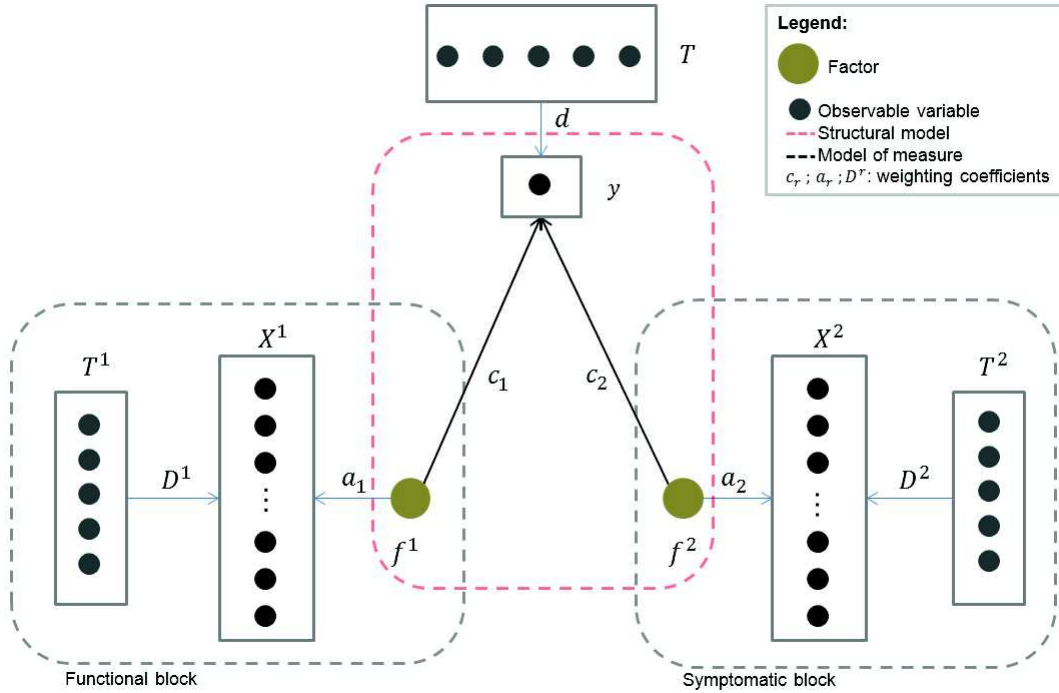


Figure 1: Diagram of the structural equation model considered from the QLQ-C30 decomposition

¹It is possible to easily generalize the n -length vector y to a matrix Y ($n \times q_Y$) of q_Y observed variables explained by the factors.

As formerly mentioned, for each equation of this model, each observed variable in a block is expressed as a linear combination of the corresponding factor (both factors for y), the covariates and some noise. Hence, the model:

$$\begin{cases} X^1 = T^1 D^1 + f^1 a^{1'} + \varepsilon^1 & (1a) \\ X^2 = T^2 D^2 + f^2 a^{2'} + \varepsilon^2 & (1b) \\ y = Td + f^1 c^1 + f^2 c^2 + \varepsilon^y & (1c) \end{cases}$$

The corresponding equation set, for a given subject i , reads:

$$\begin{cases} x_i^{1'} = t_i^{1'} D^1 + f_i^1 a^{1'} + \varepsilon_i^{1'} \\ x_i^{2'} = t_i^{2'} D^2 + f_i^2 a^{2'} + \varepsilon_i^{2'} \\ y_i' = t_i' d + f_i^1 c^1 + f_i^2 c^2 + \varepsilon_i^{y'} \end{cases} \quad (1)$$

where d (resp. D^m) is a $r_T \times 1$ (resp. $r_m \times q_m$) parameter matrix, a^m a $q_m \times 1$ parameter vector, and ε^y (resp. ε^m) an $n \times 1$ (resp. $n \times q_m$) measurement-error matrix.

We impose that the first column of T as well as of each T^m matrix is equal to the constant vector having all elements equal to 1. Thus, d and the first row of each D^m contains mean-parameters.

The main assumptions of this model are the following:

As far as distributions are concerned, we assume that:

- $\varepsilon_i^y \sim \mathcal{N}(0, \sigma_y^2)$;
- $\forall m \in \llbracket 1, 2 \rrbracket$: $\varepsilon_i^m \sim \mathcal{N}(0, \psi_m)$, where $\psi_m = \text{diag}(\sigma_m^2)$;
- ε^y and ε^m are independent $\forall m \in \llbracket 1, 2 \rrbracket$.

As to the factors, we assume that:

- $\forall m \in \llbracket 1, 2 \rrbracket$: $f^m \sim \mathcal{N}(0, I_n)$ with f^1, f^2 independent;
- In each block $\forall m \in \llbracket 1, 2 \rrbracket$ the observed variables X^m (e.g. $x_j^m, j \in \llbracket 1, q_p \rrbracket$) depend linearly on the block's factor (f^m) and a block of extra-covariates (e.g. T^m), conditional on which they are independent. The covariates are useful to test a treatment, typically;
- y depend linearly on the factors f^1 and f^2 and an extra-covariate vector T ;
- $\varepsilon^y, \varepsilon^m, f^m, \forall m \in \llbracket 1, 2 \rrbracket$ are mutually independent.

2.2 Estimation using the EM estimation method

The EM estimation method has been proposed by Bry et al (Bry et al., 2016) to estimate structural models. It is based on the EM likelihood-maximization algorithm (Dempster et al., 1977), which allows to estimate not only the coefficients of the model, but also the values of its factors at unit-level. The idea of this approach is to consider factors as missing data, which EM allows to estimate.

2.2.1 The complete log-likelihood function

Let $z = (y, X^1, X^2)$ be the observed variables and $h = (f^1, f^2)$ the factors. The EM algorithm is based on the log-likelihood associated with the complete data (z, h) . Let $p(z, h; \theta)$ denote the probability density of the complete data.

$$\begin{aligned} \mathcal{L}(\theta; z, h) = & -\frac{1}{2} \sum_{i=1}^n \{ \ln(\sigma_y^2) + q_1 \ln(\sigma_1^2) + q_2 \ln(\sigma_2^2) \\ & + \sigma_y^{-2} (y_i - t_i d - f_i^1 c^1 - f_i^2 c^2)^2 \\ & + \sigma_1^{-2} (x_i^1 - (D^1)' T_i^1 - f_i^1 a^1)' (x_i^1 - (D^1)' T_i^1 - f_i^1 a^1) \\ & + \sigma_2^{-2} (x_i^2 - (D^2)' T_i^2 - f_i^2 a^2)' (x_i^2 - (D^2)' T_i^2 - f_i^2 a^2) \\ & + (f_i^1)^2 + (f_i^2)^2 \} + \lambda \end{aligned}$$

Where λ is a constant and θ is the following K -dimensional set of parameters:

$$\theta = \{d, D^1, D^2, a^1, a^2, c^1, c^2, \sigma_y^2, \sigma_1^2, \sigma_2^2\}$$

Thus, we have:

$$K = 5 + r_T + \sum_{m=1}^2 q_m (r_m + 1)$$

2.2.2 Estimation of the structural equation model

To maximize this function, in the framework of the EM algorithm, we have to solve:

$$\mathbb{E}_z^h \left[\frac{\partial}{\partial \theta} \mathcal{L}(\theta; z, h) \right] = 0 \quad (2)$$

To make it, we need the derivatives of the log-likelihood function and the distribution $p_{z_i}^{h_i}$ of h_i conditional on z_i for each observation $i \in \llbracket 1, n \rrbracket$. Classically, the use of numeric methods are needed, but we deal with Gaussian distributions in our specific case and this causes $p_{z_i}^{h_i}$ to be an explicit conditional Gaussian distribution. Let us introduce the following notations:

$$p_{z_i}^{h_i} = \mathcal{N}(M_i = \begin{pmatrix} M_{1i} \\ M_{2i} \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix})$$

$$\begin{aligned} \widetilde{f}_i^1 &= \mathbb{E}_{z_i}^{h_i} [f_i^1] = M_{1i}; & \widetilde{\phi}_i^1 &= \mathbb{E}_{z_i}^{h_i} [(f_i^1)^2] = (\mathbb{E}_{z_i}^{h_i} [f_i^1])^2 + \mathbb{V}_{z_i}^{h_i} [f_i^1] = M_{1i}^2 + \sigma_{11} \\ \widetilde{f}_i^2 &= \mathbb{E}_{z_i}^{h_i} [f_i^2] = M_{2i}; & \widetilde{\phi}_i^2 &= \mathbb{E}_{z_i}^{h_i} [(f_i^2)^2] = (\mathbb{E}_{z_i}^{h_i} [f_i^2])^2 + \mathbb{V}_{z_i}^{h_i} [f_i^2] = M_{2i}^2 + \sigma_{22} \end{aligned}$$

Thus, when we follow the same procedure described in Bry et al (Bry et al., 2016). We obtain $\forall m \in \{1, 2\}$ the following system characterizing the solution of (2):

$$\begin{aligned}
\widehat{a}^m &= [\widetilde{f}^m x^m - \overline{x^m t^m} (\overline{t^m t^m})^{-1} \widetilde{f}^m t^m] [\widetilde{\phi}^m - \widetilde{f}^m t^m (\overline{t^m t^m})^{-1} \widetilde{f}^m t^m]^{-1} \\
\widehat{D}^{m'} &= [\overline{x^m t^m} - \widehat{a}^m (\widetilde{f}^m t^m)] [\overline{t^m t^m}]^{-1} \\
\widehat{c}^1 &= \frac{[\overline{y^t} (\overline{t^t})^{-1} \widetilde{f}^1 t - \widetilde{f}^1 y] [\widetilde{f}^2 t' (\overline{t^t})^{-1} \widetilde{f}^2 t - \widetilde{\phi}^2] - [\widetilde{f}^2 t' (\overline{t^t})^{-1} \widetilde{f}^1 t - (\sigma_{12} + \widetilde{f}^1 \widetilde{f}^2)] [\overline{y^t} (\overline{t^t})^{-1} \widetilde{f}^2 t - \widetilde{f}^2 y]}{[\widetilde{f}^1 t' (\overline{t^t})^{-1} \widetilde{f}^1 t - \widetilde{\phi}^1] [\widetilde{f}^2 t' (\overline{t^t})^{-1} \widetilde{f}^2 t - \widetilde{\phi}^2] - [\widetilde{f}^2 t' (\overline{t^t})^{-1} \widetilde{f}^1 t - (\sigma_{12} + \widetilde{f}^1 \widetilde{f}^2)] [\widetilde{f}^1 t' (\overline{t^t})^{-1} \widetilde{f}^2 t - (\sigma_{21} + \widetilde{f}^1 \widetilde{f}^2)]} \\
\widehat{c}^2 &= \frac{[\overline{y^t} (\overline{t^t})^{-1} \widetilde{f}^2 t - \widetilde{f}^2 y] [\widetilde{f}^1 t' (\overline{t^t})^{-1} \widetilde{f}^1 t - \widetilde{\phi}^1] - [\widetilde{f}^1 t' (\overline{t^t})^{-1} \widetilde{f}^2 t - (\sigma_{21} + \widetilde{f}^1 \widetilde{f}^2)] [\overline{y^t} (\overline{t^t})^{-1} \widetilde{f}^1 t - \widetilde{f}^1 y]}{[\widetilde{f}^1 t' (\overline{t^t})^{-1} \widetilde{f}^1 t - \widetilde{\phi}^1] [\widetilde{f}^2 t' (\overline{t^t})^{-1} \widetilde{f}^2 t - \widetilde{\phi}^2] - [\widetilde{f}^2 t' (\overline{t^t})^{-1} \widetilde{f}^1 t - (\sigma_{12} + \widetilde{f}^1 \widetilde{f}^2)] [\widetilde{f}^1 t' (\overline{t^t})^{-1} \widetilde{f}^2 t - (\sigma_{21} + \widetilde{f}^1 \widetilde{f}^2)]} \\
\widehat{d}^t &= [\overline{y^t} - \widehat{c}^1 \widetilde{f}^1 t' - \widehat{c}^2 \widetilde{f}^2 t'] [\overline{t^t}]^{-1} \\
\widehat{\sigma}_y^2 &= \frac{1}{n} \sum_{i=1}^n \{ (y_i - \widehat{d}^t t_i)^2 - 2(y_i - \widehat{d}^t t_i)(\widehat{c}^1 \widetilde{f}_i^1 + \widehat{c}^2 \widetilde{f}_i^2) + (\widehat{c}^1)^2 \widetilde{\phi}_i^1 + (\widehat{c}^2)^2 \widetilde{\phi}_i^2 + 2\widehat{c}^1 \widehat{c}^2 (\sigma_{12} + \widetilde{f}_i^1 \widetilde{f}_i^2) \} \\
\widehat{\sigma}_m^2 &= \frac{1}{n q_m} \sum_{i=1}^n \{ \|x_i^m - \widehat{D}^{m'} t_i^m\|^2 + \|\widehat{a}^m\|^2 \widetilde{\phi}_i^m - 2(x_i^m - \widehat{D}^{m'} t_i^m)' \widehat{a}^m \widetilde{f}_i^m \}
\end{aligned}$$

2.2.3 Algorithm

We index with $[s]$ any value currently obtained on the s^{ieth} iteration of the algorithm.

1. Initialization ² = choice of the initial parameter values $\theta^{[0]}$.
2. Current iteration $s \geq 1$, until stopping condition is met:
 - (a) **E-step:** with $\theta^{[s-1]}$,
 - i. Calculate explicitly distribution $p_{z_i}^{h_i}$ for each $i \in \llbracket 1, n \rrbracket$.
 - ii. Estimate the factor-values $\widetilde{f}^m^{[s]}$, $m \in \{1, 2\}$.
 - iii. Calculate $\widetilde{\phi}^m^{[s]}$, $m \in \{1, 2\}$.
 - (b) **M-step:**
 - i. Update θ to $\theta^{[s]}$ by injecting $\widetilde{f}^m^{[s]}$ and $\widetilde{\phi}^m^{[s]}$, $m \in \{1, 2\}$ into the formulas of (2).

²In the initialization step, $\forall m \in \{1, 2\}$ we propose to obtain $D^{m[0]}$ by multiple linear regression of X^m on T^m . Then, to initialize the others parameter values, we compute each approximated factor $\widetilde{f}^m^{[0]}$ as the first principal component of $X^m - T^m D^{m[0]}$. Then, we initialize a^m , σ_m^2 (resp. d , c^m , σ_y^2) through a multiple linear regression of $X^m - T^m D^{m[0]}$ on $\widetilde{f}^m^{[0]}$ (resp. of y on T , $\widetilde{f}^{1[0]}$ and $\widetilde{f}^{2[0]}$). In practice we use functions $lm()$ of R and $PCA()$ of the package *FactoMineR* (Husson et al., 2015) R-package.

3. We used the following stopping condition with the smallest possible ϵ :

$$\sum_{k=1}^K \frac{|\theta^{*[s+1]}[k] - \theta^{*[s]}[k]|}{\theta^{*[s+1]}[k]} < \epsilon \quad (3)$$

where θ^* is the K - dimensional vector containing the values of all scalar parameters in θ .

2.3 Tests and performances of the algorithm

In the case of a more general structural model, numerical results on simulated data have been presented (Bry et al., 2016). A data generation procedure was described and performed one hundred times, each time yielding a set of simulated data matrices (Y, X^1, X^2)³. Then for each simulated data, an estimation routine with a threshold value $\epsilon = 10^{-2}$ was run. The results showed that:

- Convergence was observed in almost all cases in less than five iterations.
- The estimations of the parameters θ and the factors were very close to their simulated values.

The quality of estimations was measured through the absolute relative deviation between each simulated scalar parameter in θ^* and its estimation, averaged over the hundred simulations, and through the averaged square correlations between simulated factors and their estimations.

Then a sensitivity analysis was performed to investigate how the quality of estimations could be affected by the number n of subjects and the number q of observed variables in each block. The results were that:

- The sample size n proved to have more impact on the quality of parameter estimation and factor reconstruction than the number of observed variables.
- The quality of factor reconstruction remained high for rather small values of n or q . We advise to use a minimal sample size of $n = 100$ to obtain really stable structural coefficients.

3 A linear mixed model for longitudinal analysis

As in the classical approach, the longitudinal step is performed via a LMM. The aim of the second step is to assess the HRQoL evolution over time, where the GSH is explained by the covariates and the estimated factors obtained in the first step. The response variable consists of repeated measurements on the same subjects over time. For this kind of analysis, the LMM methodology is classically used to take into account the dependence of data from the same subject and to estimate the inter-subject variability. The global HRQoL longitudinal measures y_{iv} at visit v for subject i are described by a standard linear mixed model (Anota et al., 2014; Rizopoulos, 2012):

$$y_{iv} = \alpha + \mathbf{x}'_{iv}\boldsymbol{\beta} + \mathbf{u}'_i\xi_i + \varepsilon_{iv}, \quad (4)$$

where

³Where Y is a $(n \times q_Y)$ matrix generalizing the n -length vector in the case of $q_Y \geq 1$ observed variable explained by the factors.

- α is the model intercept;
- β is the vector of fixed effects and \mathbf{x}_{iv} the known design vector containing $\tilde{f}_{iv} = (\tilde{f}_{iv}^1, \tilde{f}_{iv}^2)$ and others covariates;
- ξ_i is the vector of subject-specific random effects such that $\xi_i \sim \mathcal{N}(\mathbf{0}, \Sigma)$ where Σ is the covariance matrix, and \mathbf{u}_i is the known design vector;
- $\varepsilon_{iv} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is the error term.

In the HRQoL assessment, two individual random effects are considered: the intercept and the slope, as is classically done in longitudinal analyses on repeated measures by subject (Rizopoulos, 2012; Agresti, 2010). In the longitudinal analysis, the random intercept takes into account the variability of the HRQoL perception specific to each patient, while the random slope is associated with the variability between patients concerning the HRQoL evolution. For the fixed part of the model, the factors $\tilde{f}_{iv} = (\tilde{f}_{iv}^1, \tilde{f}_{iv}^2)$ estimated at each visiting time are considered to take into account the global symptomatic status and the global functional status of patients. It is thus possible to separate the effects of the two factors in explaining the global evolution of the HRQoL. Moreover, covariates are also considered so as to study their influence on this evolution. The explanatory variables can be the same as those used on the first step, but the interpretation differ between the two steps. In the first one, the covariate's influence at a given time is assessed, while in the second its influence on the evolution of the dependent variable is assessed.

4 Application

The implementation of the approach is performed on HRQoL data from CO-HO-RT clinical trial (Azria et al., 2010). This is a randomized phase II trial with one hundred and forty nine breast-cancer patients. The clinical trial aim was to compare two therapy programs with letrozole associated with a radiotherapy. After surgery and chemotherapy, patients from the first group (concomitant arm, $n = 74$) received some letrozole for two years and three weeks after treatment began they started radiotherapy for five weeks. Patients from the second group (sequential arm, $n = 75$) received the same treatments but the radiotherapy was administered before the treatment by letrozole. For both groups, the measurement times were every three months for two years.

At each time, the patients filled out the EORTC QLQ-C30 and the EORTC QLQ-BR23 questionnaires (Fayers et al., 2001). The latter questionnaire is specific to the breast-cancer and completes the generic EORTC QLQ-C30, and the acronym of its specific dimensions begin by "BR" for "breast". This complementary module adds eight specific HRQoL dimensions to the breast-cancer study: four dimensions among the functional dimensions and four dimensions among the symptomatic dimensions. The scoring procedure proposed by EORTC (Fayers et al., 2001) was used in this analysis. For each HRQoL dimension, a score was calculated as the average of item responses which compose the dimension, and expressed on scale ranging from 0 to 100. A score was considered missing data if more than half of the item responses was missing. In this application, three dimensions (sexual functioning, sexual enjoyment, upset by hair-loss) were not considered because of too much missing data (over fifty to the whole sample).

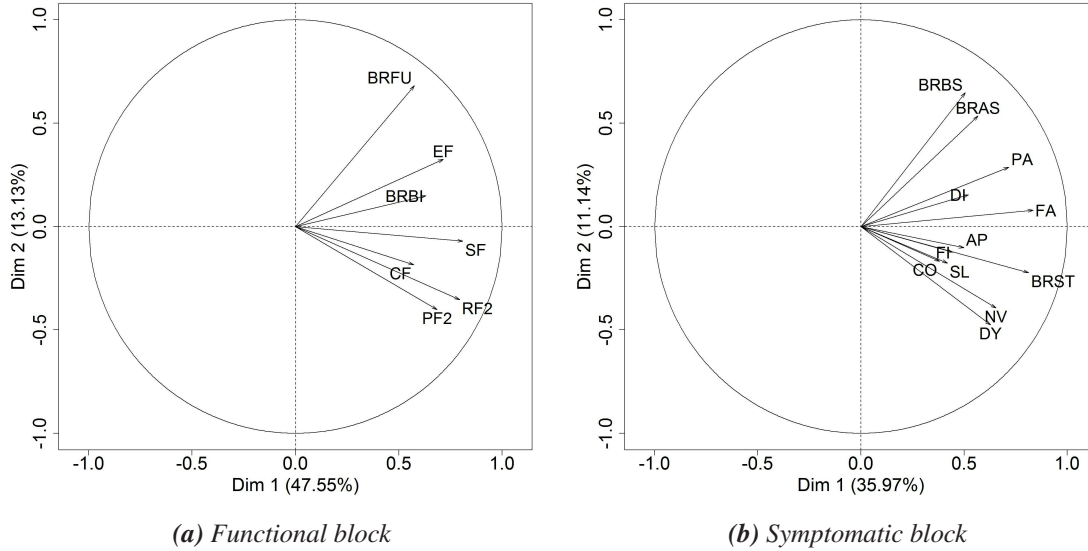


Figure 2: Correlation-scatterplots yielded by the PCAs of the X^1 and X^2 at the baseline.

4.1 Transversal analysis, reconstruction of factors

In this section, we aim at modeling the Global Health Status from the others variables on each visit, while reducing the dimension of the data. In this transversal step, only patients who have a measured score for each HRQoL dimension are considered in the structural equation model (1). The dependent variable y is the Global Health Status variable. Our model is built on a conceptual grouping of functional dimensions and symptomatic dimensions, suggested by the EORTC QLQ-C30 construction. For each visit, a Principal Component Analysis (PCA) of each dimension family illustrated the bundle organization and thus the grouping (Figure 2 at the baseline). On all visits, 121 different patients were considered and the number of patient taken into account at each visit was between 84 and 113 (Table 1). The number of iterations needed for the algorithm to converge did not exceed 64 for every measurement time. As for the convergence time, one second was always enough to estimate the model.

Table 1: Details associated with the structural model estimation for each measurement time (visit): sample size, number of iterations and time to convergence. A total of 121 patients have completed the questionnaires to have at least one score for each considered HRQoL dimension.

	Visit v (in month)							
	0	3	6	12	15	18	21	24
Sample size (n_v)	113	106	102	100	102	84	91	90
Number of iterations	34	50	58	45	56	59	64	62
Time (s)	0.42	0.79	0.83	0.71	0.58	0.80	0.50	0.49

The main aim of the first step is the reconstruction of the functional and symptomatic factors for each subject from their different scores. In the factor estimation, the group effect can be also taken into account in the structural model to explain the observed scores, so we took the group as covariate in

equations (1a), (1b) and (1c). The group-effect quantifies the score difference between the two groups for each dimension at the considered measurement times. The group covariate is defined so that value "0" corresponds to the group receiving the concurrent radiotherapy and "1" to that receiving the sequential radiotherapy. The first step is also very useful in the descriptive analysis of HRQoL scores. Indeed, each measurement equation allows to divide every HRQoL dimension into a group-effect and a group-deflated factor. The estimations of their parameters for each visit makes it possible, which can be well appreciated by clinicians. Indeed, the estimations give a short overview of covariates' effects or mean scores, and of the HRQoL evolution for each visit and specific HRQoL dimension. The transversal results may be a real benefit for clinicians to study the consistency of estimated effects with the clinical trail protocol. In our application for example, we may expect to a difference between the two groups at the visits 1 month and 3 months. This makes possible to retrieve a quick overview of studied data for a previous descriptive analysis.

For every measurement time, part of the results are showed in Tables 2, 3 and 4. They give respectively the estimations of vector d and matrices D^1 and D^2 related respectively to equations (1c), (1a) and (1b). Two estimations are distinguished. Let g denote the group-index, $(\mu_g)_{g=0,1}$ be the average scores of the two groups and $\delta = \mu_1 - \mu_0$. Regarding Table 2, the difference between the average scores of the two groups increased across time. Indeed, the GHS score of the concurrent group is constant around 70 while the GHS score of the sequential group ($\mu_1 = \mu_0 + \delta$) began to 70 and decreased from the visit at three months.

Table 2: For each measurement time, estimations of parameters p associated with vector d from equation (1c): intercept and treatment-group taken into account in each equation of model (1).

p	Visit v (in month)							
	0	3	6	12	15	18	21	24
μ_0	71.49	69.03	68.24	70.83	70.00	68.60	72.57	70.39
δ	0.98	-0.47	-1.74	-2.72	-1.41	-4.78	-5.90	-2.17

The results presented in Table 3 exhibit three general patterns across the visits. The first one is that the average score for functional dimension is over 60 and for several dimensions, around 80. On the scale ranging from 0 to 100, this suggests the patients had a relatively high perception of their functioning capacity. Secondly, if on the whole visits did not present a difference between the two groups, for some visits the difference could be clinically significant. Indeed, the visits which showed the greatest difference between groups as to the functional dimensions are those at 12 and 21 months. For both, the HRQoL level of patients receiving the sequential radiotherapy was uniformly lower than that of patients receiving the concurrent radiotherapy. For example, there is a difference of 5.48 points between the scores of the two groups for the physical functioning at 12 months, in favor of the concurrent group. When comparing the functional dimensions, the best perception appears to be that of the social functioning and the worst that associated with future perspective. Finally, the average scores did not appear to change the measurement times.

As for the symptomatic block, the results (Table 4) suggested similar conclusions as for the functional block. The observed symptomatic scores are low (between 0 and 45 on the 0-100 scale) that suggest a good level of HRQoL with few symptoms. The dimensions with the worst symptomatic level are insomnia, fatigue and pain. Globally, the scores seemed similar across groups whatever the visit, even if the visits at 3 months, 12 months and 21 months respectively presented scores relatively greater for

Table 3: For each measurement time, estimations of parameters p associated with matrix D^1 from equation (1a): intercept and treatment-group taken into account in each equation of model (1). X^1 corresponds to the functional dimension scores: physical (PF2), role (RF2), emotional (EF), cognitive (CF), social (SF), body image (BRBI) and future perspective (BRFU).

Visits	p	Functioning dimensions						
		PF2	RF2	EF	CF	SF	BRBI	BRFU
Baseline	μ_0	82.81	85.38	75.05	75.73	91.52	84.75	60.23
	δ	1.84	1.52	-1.49	2.84	0.44	2.31	1.08
3 months	μ_0	82.08	85.85	75.42	77.99	91.19	87.89	71.70
	δ	-1.29	-3.77	-1.10	1.57	0.00	-1.89	-5.66
6 months	μ_0	81.70	83.65	74.00	77.36	87.42	88.89	64.15
	δ	-0.88	1.05	0.37	1.55	-0.35	-2.32	3.20
12 months	μ_0	80.83	81.94	73.21	79.86	90.62	86.46	70.14
	δ	-5.74	-6.62	-3.65	2.19	-2.16	-3.12	-7.96
15 months	μ_0	82.80	81.00	74.56	79.67	90.00	89.50	68.00
	δ	0.95	-1.51	-4.52	0.14	-2.18	-6.65	-3.90
18 months	μ_0	84.19	83.33	70.54	78.68	91.86	89.34	70.54
	δ	1.18	-1.22	-0.83	3.03	-0.80	-6.41	-2.25
21 months	μ_0	85.83	82.29	78.30	80.90	90.62	88.19	65.97
	δ	-5.48	0.27	-9.31	-5.32	-3.03	-6.28	-3.18
24 months	μ_0	82.55	81.56	74.53	78.37	89.01	86.17	68.09
	δ	-0.46	2.55	-0.88	0.31	0.14	-2.84	-2.19

patient receiving sequential radiotherapy. On the whole, there was no difference between groups which could be explained by the administration method. A difference could be expected at three months after starting (end of radiotherapy), but the difference was not conclusive.

As previously mentioned, the main aim is the estimation of the two unknown factors reflected by symptomatic and functional dimensions, once the effect of covariates is removed. The question arises thus of the influence of covariates on the factor building. Figure 3 shows the correlation between the factors estimated with the group covariates $(\tilde{f}_T^m)_{m=1,2}$ and without $(\tilde{f}^m)_{m=1,2}$. According to the graphics, there is a low group effect for a few visits. Concerning the functional factor (Figure 3a), the visits at 12 months and 21 months exhibit a clear difference between groups. Indeed, $\tilde{f}_T^1 < \tilde{f}^1$ for the concurrent administration group (red), and $\tilde{f}_T^1 > \tilde{f}^1$ for the sequential administration group (blue). This suggests for these visits, the sequential administration group had a functional health status lower than the concurrent administration group. This confirms the estimation given in Table 3. However, no reason from the protocol allowed to explain the small difference for these specific visits. Concerning the measurement model associated with the symptomatic dimensions (Figure 3b), the estimation of the symptomatic factor is hardly affected by the group covariate, with the exception of the visit at 3, 12 and 18 months. Indeed, $\tilde{f}_T^2 > \tilde{f}^2$ for the concurrent administration group (red) for these measurement times, while for the sequential administration group $\tilde{f}_T^2 < \tilde{f}^2$. The reverse situation for the symptomatic health status is clinically consistent with the functional factor observations. The symptomatic status of the sequential administration group is higher than that of the concurrent administration group. The latter group seemed

Table 4: For each measurement time, estimations of parameters p associated with matrix D^2 from equation (1b): intercept and treatment-group taken into account in each equation of model (1). X^2 corresponds to the symptomatic dimension scores: fatigue (FA), nausea and vomiting (NV), pain (PA), dyspnea (DY), insomnia (SL), appetite loss (AP), constipation (CO), diarrhoea (DI), financial difficulty (FI), systemic therapy side effect (BRST), breast symptom (BRBS) and arm symptoms (BRAS).

Visits	p	Symptomatic dimensions											
		FA	NV	PA	DY	SL	AP	CO	DI	FI	BRST	BRBS	BRAS
Baseline	μ_0	24.95	3.80	17.84	14.62	31.58	8.77	17.54	8.19	4.09	19.26	15.84	16.96
	δ	0.64	0.66	1.51	-1.52	-1.82	-1.63	-1.47	-4.02	4.84	-1.38	-0.81	-2.77
3 months	μ_0	24.11	3.14	22.64	22.01	34.59	5.66	18.87	6.92	5.03	16.71	19.34	17.19
	δ	6.81	1.89	5.35	-1.26	4.40	0.63	-3.77	-1.26	3.77	2.61	2.15	4.82
6 months	μ_0	26.94	3.14	25.16	18.24	38.36	9.43	13.21	8.80	5.03	19.06	18.92	21.91
	δ	-0.52	1.28	2.05	2.17	-5.71	-3.31	5.16	-3.36	0.41	-2.58	-2.37	-6.49
12 months	μ_0	28.01	4.51	27.43	18.75	40.28	6.25	14.58	5.56	6.94	18.14	17.88	21.99
	δ	5.00	0.61	3.66	6.89	-1.18	5.29	4.65	3.42	-1.82	1.63	-0.20	-2.87
15 months	μ_0	26.78	4.67	26.67	18.00	38.67	8.67	14.00	8.67	3.33	18.45	13.72	16.33
	δ	3.56	-0.82	-2.95	2.51	1.72	-2.90	-3.10	-2.90	1.15	1.95	1.02	-0.31
18 months	μ_0	31.01	3.88	20.93	15.50	42.64	10.85	12.40	7.75	6.20	18.11	13.18	14.99
	δ	-2.69	-1.44	6.71	-0.06	-8.49	-6.79	3.04	-0.43	-4.58	-0.96	4.44	3.58
21 months	μ_0	21.18	3.13	25.35	15.28	38.19	4.86	12.50	8.33	4.86	17.87	14.29	16.20
	δ	8.92	2.69	-2.87	7.20	-0.21	2.89	2.23	-4.46	2.89	3.20	1.79	1.50
24 months	μ_0	22.70	3.55	23.76	16.31	32.62	7.09	12.06	7.80	9.22	18.22	13.18	14.78
	δ	1.85	-1.22	1.43	0.74	-0.07	-1.67	7.32	-2.38	-2.24	-0.30	-0.45	3.05

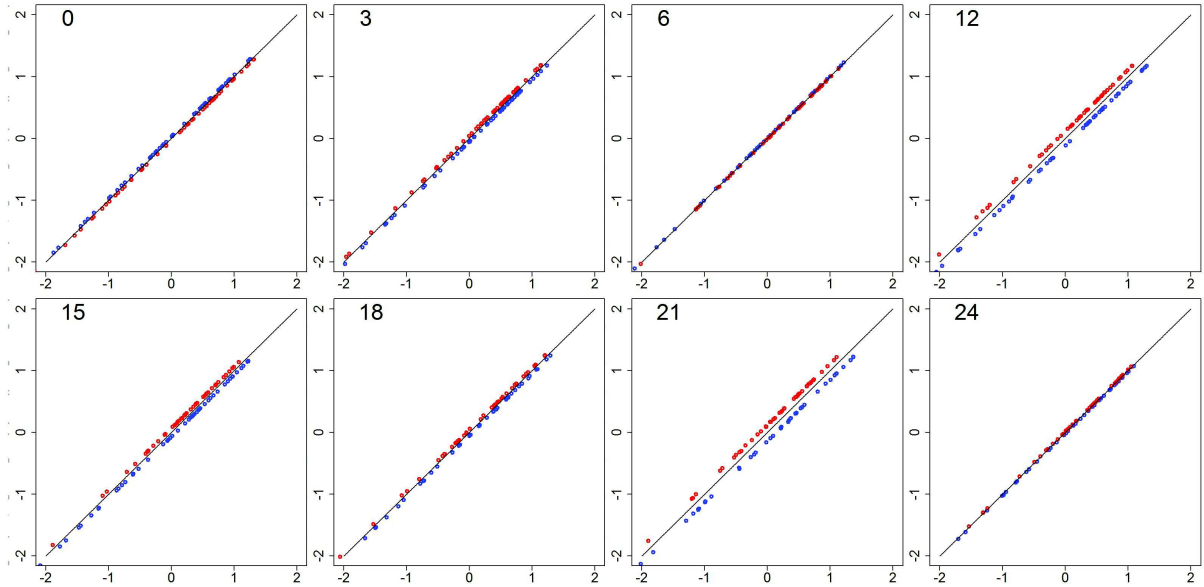
to have a low HRQoL given the symptomatic health status for the three visits.

The estimation of factors is impacted by the taking into account of covariates when modeling the dependent variable. When covariates are omitted, their effect on the HRQoL score is conveyed by the factors, the other variable(s) in T and the error term of the measurement models. According to Figure 3a, a lower functional capacity for the sequential administration group caused a global decreasing of the functional factor in this group when "group" covariate was not taken into account. By contrast, the functional factor for the concurrent administration group increased. When the covariate effect is not considered in the model, the estimated factors will be biased with respect to the true hidden factor.

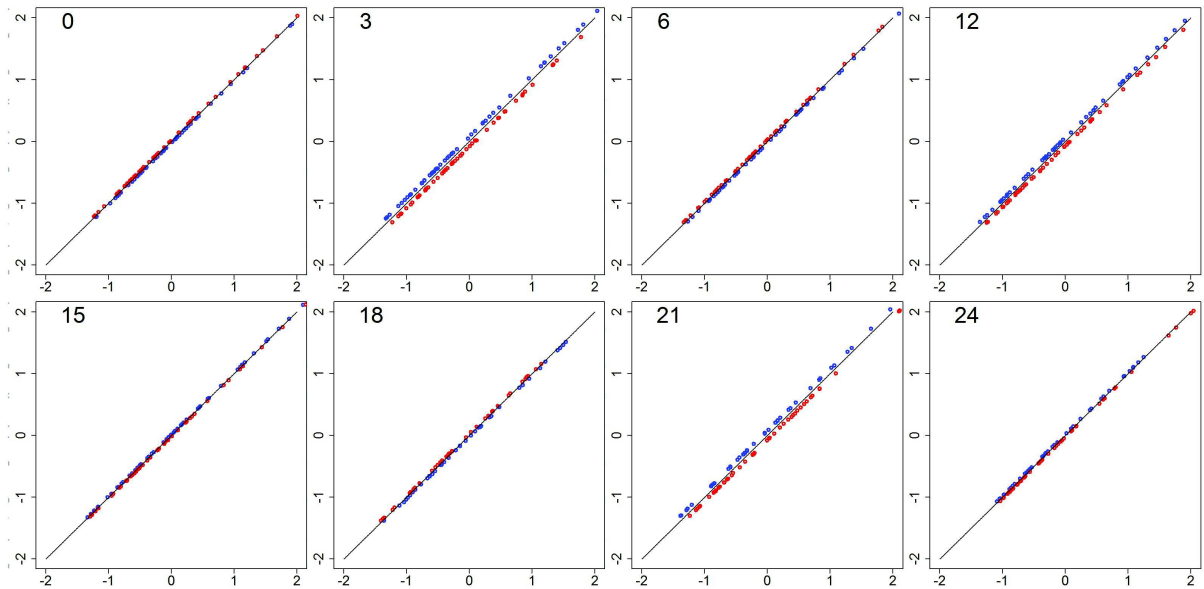
The transversal analysis brought some punctual information on the difference or similarity between the HRQoL at every measurement time across groups, information which was not visible in the classical longitudinal studies by dimension. The sequential administration group presented a better functional health status at the visits at 12 and 21 months after the inclusion, and a lower symptomatic health status three months after the inclusion than the concurrent administration group. Another strong advantage, the structural equation model approach via the algorithm EM is to summarize the symptomatic and functioning scores in two factors which will now be considered in the longitudinal step.

4.2 Longitudinal analysis

After considering the data in the transversely, we are interested in studying the global perception of HRQoL over time. To achieve that, the LMM is used to link the measurements of the same subject. Considering the GSH as the response variable y , the complete LMM (\mathcal{M}_c) considered for this specific application is:



(a) Functional factors (\tilde{f}^1)



(b) Symptomatic factors (\tilde{f}^2)

Figure 3: Scatterplot of estimated factors with (for x-axis) and without (for y-axis) taking into account the treatment group covariate in the both measurement blocks and the dependent variable y . The associated visit-date is mentioned on the top left corner of every plot. The concurrent treatment administration is plotted in red and the sequential one in blue.

$$\begin{aligned} \mathcal{M}_c : \quad y_{iv} = & \beta_0 + \beta_1 \tilde{f}_{iv}^1 + \beta_2 \tilde{f}_{iv}^2 + \beta_3 x_i + \beta_4 t_v \\ & + \beta_5 x_i t_v + \beta_6 \tilde{f}_{iv}^1 x_i + \beta_7 \tilde{f}_{iv}^2 x_i + \xi_{i0} + \xi_{i1} t_v + \varepsilon_{iv}, \end{aligned} \quad (5)$$

where

- $(\beta_j)_{j=0,\dots,7}$ are the fixed effects;
- x_i is the affiliation-indicator of the group: $x_i = 0$ if the subject i received the concurrent treatment, and $x_i = 1$ else (sequential administration);
- \tilde{f}_{iv}^1 and \tilde{f}_{iv}^2 are respectively the estimated functional and symptomatic factors for subject i ;
- t_v is the time (in month) elapsed since the inclusion (baseline, $t_0 = 0$) and the visit v .

No interaction between the time and factors was taken into account, since factors were estimated at every visit, and thus contain the visit-time information. To fit the model, the *lmer* function of the R package *lme4* was used (Bates et al., 2014). The Bayesian information criterion was used to select the model the most likely. The factors built while taking into account of the group variable were not considered. In the longitudinal analysis, they were not used because they seemed less relevant than the factors estimated without the knowledge of group. From empirical data, the model (\mathcal{M}_ℓ) found the most likely to explain the GSH was:

$$\mathcal{M}_\ell : \quad y_{iv} = \beta_0 + \beta_1 \tilde{f}_{iv}^1 + \beta_2 \tilde{f}_{iv}^2 + \xi_{i0} + \varepsilon_{iv} \quad (6)$$

As anticipated from previous transversal results, no evolution of the GHS over time was, found, nor any difference between groups. The time information included in the factor, if any, is sufficient. There is no difference between groups at inclusion, as expected, because the patients were randomized. Table 5 gives the estimation details of a few models (6). The functional factor seems to give twice as much information as the symptomatic factor to explain the GHS. The individual part conveyed by the random intercept is the most informative component. This suggests that the HRQoL depends strongly of the subject, and confirms the subjective specificity of the HRQoL endpoint. Moreover, if the two factors are not taken into account to explain the GHS, there is a consequent loss of information. Indeed, the two factors are rather strongly correlated (Figure 4, $\rho = -0.81$). Thus, the absence of one is to some extent compensated by the other. Figure 4 shows that each variable family is positively correlated with the corresponding factor.

5 Discussion

The new approach to study the HRQoL in oncology is original and suggests a new course of action. The proposed method is presented to study the HRQoL from a global point of view. The first strong point is the use of a structural equation model estimation at every measurement time. This allows the estimation of each parameter and subject-specific factor. A descriptive analysis can first be done on the estimated parameters at each time. This can give hints as to the influence of observable variables included in the model and allows to compare HRQoL dimension families or specific HRQoL dimensions.

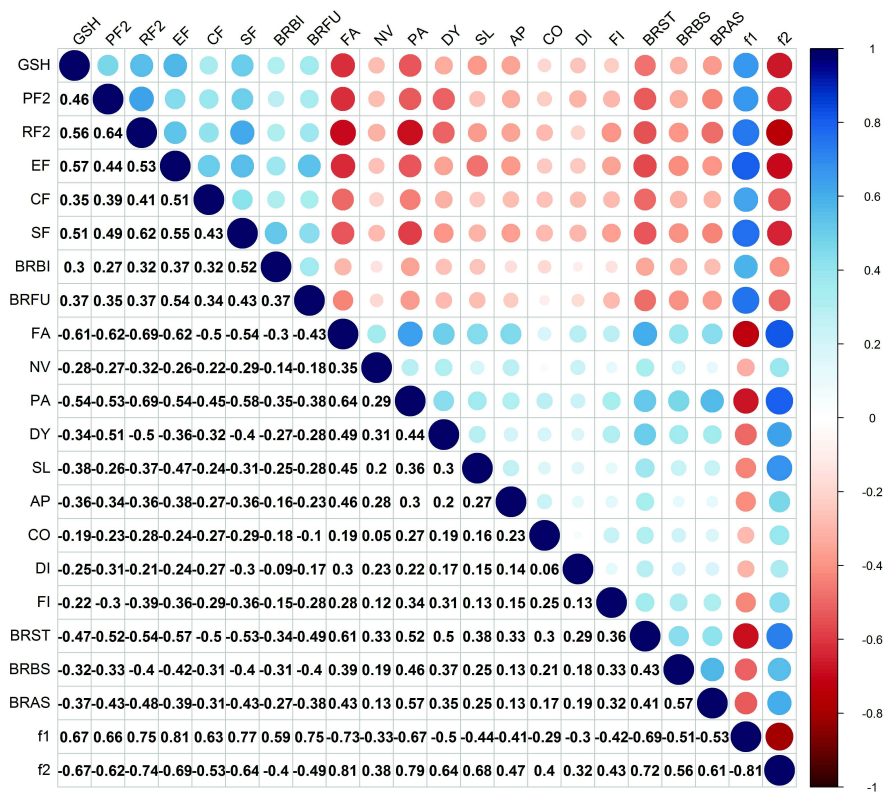


Figure 4: Correlation matrix between each HRQoL dimension score and the estimated factor concerning the CO-HO-RT clinical trial data.

Table 5: Details of the LMM estimation. The fixed effects, the variance of random intercept and the BIC associated with every model are given. γ is the difference between $BIC(\mathcal{M}_\ell)$ and $BIC(\mathcal{M}_j)_{j=1,\dots,4}$.

Models	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_{\xi_0}^2$	$\hat{\sigma}_\varepsilon^2$	BIC	γ
\mathcal{M}_ℓ	69.11	10.14	-7.61	78.77	160.40	6440.4	0
\mathcal{M}_1	68.97	15.89	×	92.91	168.14	6481.6	41.2
\mathcal{M}_2	69.07	×	-15.17	77.89	180.71	6516.2	75.8
\mathcal{M}_3	69.12	8.65	-8.38	×	236.83	6568.2	127.8
\mathcal{M}_4	68.14	×	×	225.64	245.65	6823.0	382.7

From these results, some clinical interpretations can emerge and provide information on how to perform longitudinal analysis. Moreover, the interpretation of coefficients in such a model is the same as in the classical regression equation for each equation regression of the model. One of the interests of the transversal step is to see the potentially different changes across the visits. The strongest asset of our approach is the estimation of latent factors. This allows to summarize the information of different variables into a single one for every group of variables measuring a given concept. In the proposed structural equation model, two factors summarizing the functional health status and the symptomatic health status are considered. After the first step, these factors are available for longitudinal analysis. The second and final step performs longitudinal analysis of the global HRQoL through the GHS study. In this article, the use of a classical LMM is proposed to assess the global HRQoL over time as a function of covariates and factors. The two-step approach performs a complete analysis of the HRQoL scores for clinical trial in oncology.

Longitudinal analyses of the functional and symptomatic parts may also be performed. Moreover, the influence of each dimension on the reflected factor was not presented in the results. To compare their influence, it is necessary to have homogeneous variances across variables of the same family (assumption of the structural equation model proposed). Then, the vector of coefficients $(a^m)_{m=1,2}$ represents the intensities of links between the dimensions of the family m and the associated factor $(f^m)_{m=1,2}$. Finally, we have considered but a single factor reflected by all dimensions in each family. Indeed, this seemed most appropriate for our data. But, for other data, we might imagine that symptomatic or functional dimensions reflect several uncorrelated factors.

To consider that the HRQoL score is normally distributed is questionable as it always is in clinical analyses of the HRQoL. Another consideration is questionable according to the literature about our approach: the use of the GSH to represent the global HRQoL (Phillips et al., 2015). The GHS is controversial and has to be used with caution because this HRQoL dimension is likely to be affected by the *Response Shift* phenomena (Sprangers and Schwartz, 1999; Schwartz and Sprangers, 1999). To remedy this situation, at least two substitution global scores are proposed in literature, built from all measurements of HRQoL dimensions. The first one is the difference between, on the one hand, the sum of functional dimension scores and the GSH score, and, on the other hand, the sum of symptomatic dimension scores except the financial difficulties (Nordin et al., 2001). This score belongs to the interval $[-800; 600]$ and is considered missing if any one of all HRQoL scores is missing. More recently, a global summary score expressed on a scale from 0 to 100 is proposed as the average of the twenty item responses of the questionnaire EORTC QLQ-C30, given the reverse permutation of response categories of functional item (Hinz et al., 2012). This one seems more relevant because such as calculation allows

to take into account some weighting of the dimensions according to their number of items. Indeed, the more items a dimension contains, the stronger its contribution. In our structural model, we chose to use the global health status score because its calculation is not derived from the other (symptomatic and functional) scores. Using a score which depends on the other scores in our model is not suitable because it would clearly ruin the exogeneity of the dependent score with respect to the explanatory ones.

Funding: Grant "INCa-DGOS-Inserm 6045".

References

- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., and de Haes, J. C. (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. *Journal of the National Cancer Institute*, 85(5):365–376.
- Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. John Wiley & Sons.
- Anota, A., Barbieri, A., Savina, M., Pam, A., Gourgou-Bourgade, S., Bonnetain, F., and Bascoul-Mollevi, C. (2014). Comparison of three longitudinal analysis models for the health-related quality of life in oncology: a simulation study. *Health and Quality of Life Outcomes*, 12.
- Azria, D., Belkacemi, Y., Romieu, G., Gourgou, S., Gutowski, M., Zaman, K., Moscardo, C. L., Lemanski, C., Coelho, M., Rosenstein, B., Fenoglietto, P., Crompton, N. E. A., and Ozsahin, M. (2010). Concurrent or sequential adjuvant letrozole and radiotherapy after conservative surgery for early-stage breast cancer (CO-HO-RT): a phase 2 randomised trial. *The Lancet. Oncology*, 11(3):258–265.
- Bacci, S., Bartolucci, F., and Gnaldi, M. (2012). A class of multidimensional latent class irt models for ordinal polytomous item responses. [arXiv:1201.4667 \[stat\]](https://arxiv.org/abs/1201.4667). arXiv: 1201.4667.
- Bates, D., Maechler, M., Bolker, B. M., and Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4. ArXiv e-print; submitted to *Journal of Statistical Software*.
- Bentler, P. and Stein, J. (1992). Structural equation models in medical research. *Statistical Methods in Medical Research*, 1(2):159–181.
- Blanchin, M., Hardouin, J.-B., Neel, T. L., Kubis, G., Blanchard, C., Mirallié, E., and Sébille, V. (2011). Comparison of CTT and Rasch-based approaches for the analysis of longitudinal Patient Reported Outcomes. *Statistics in Medicine*, 30(8):825–838.
- Boeck, P. d. and Wilson, M. (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. Springer, New York, 2004 edition edition.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley-Blackwell, New York, NY.
- Bry, X., Lavergne, C., and Tami, M. (2016). EM estimation of a structural equation model. [this paper is being written](#).

- De Ayala, R. J. (2009). The theory and practice of item response theory. Guilford Press, New York.
- 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, 39(1):1–38.
- Douglas, J. A. (1999). Item response models for longitudinal quality of life data in clinical trials. Statistics in Medicine, 18(21):2917–2931.
- Edelen, M. O. and Reeve, B. B. (2007). Applying item response theory (irt) modeling to questionnaire development, evaluation, and refinement. Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 16 Suppl 1:5–18.
- Fayers, P. M., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., and Bottomley, A. o. b. o. t. E. Q. o. L. G. (2001). EORTC QLQ-C30 Scoring Manual (3rd edition), volume Brussels: EORTC 2001.
- Fiteni, F., Westeel, V., Pivot, X., Borg, C., Vernerey, D., and Bonnetain, F. (2014). Endpoints in cancer clinical trials. Journal of Visceral Surgery, 151(1):17–22.
- Foulley, J.-L. (2002). Algorithme EM : théorie et application au modèle mixte. Journal de la société française de statistique, 143(3-4):57–109.
- Gorter, R., Fox, J.-P., and Twisk, J. W. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. BMC Medical Research Methodology, 15(1):55.
- Grilli, L. and Rampichini, C. (2011). Multilevel Models for Ordinal Data. In Kenett, R. S. and Salini, S., editors, Modern Analysis of Customer Surveys, pages 391–411. John Wiley & Sons, Ltd.
- Hinz, A., Einenkel, J., Briest, S., Stolzenburg, J.-U., Papsdorf, K., and Singer, S. (2012). Is it useful to calculate sum scores of the quality of life questionnaire EORTC QLQ-C30? European Journal of Cancer Care, 21(5):677–683.
- Huang, C.-C., Lien, H.-H., Tu, S.-H., Huang, C.-S., Jeng, J.-Y., Chao, H.-L., Sun, H.-L., and Chie, W.-C. (2010). Quality of life in taiwanese breast cancer survivors with breast-conserving therapy. Journal of the Formosan Medical Association = Taiwan Yi Zhi, 109(7):493–502.
- Husson, F., Josse, J., Le, S., and Mazet, J. (2015). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. R package version 1.29.
- Joreskog, K. G. (1970). A General Method for Analysis of Covariance Structures. Biometrika, 57(2):239–251.
- King-Kallimanis, B. L., Oort, F. J., Nolte, S., Schwartz, C. E., and Sprangers, M. A. G. (2011). Using structural equation modeling to detect response shift in performance and health-related quality of life scores of multiple sclerosis patients. Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation, 20(10):1527–1540.
- King-Kallimanis, B. L., ter Hoeven, C. L., de Haes, H. C., Smets, E. M., Koning, C. C. E., and Oort, F. J. (2012). Assessing measurement invariance of a health-related quality-of-life questionnaire in radiotherapy patients. Quality of Life Research, 21(10):1745–1753.

- Lei, P.-W. and Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. Educational Measurement: Issues and Practice, 26(3):33–43.
- Masters, G. (1982). A rasch model for partial credit scoring. Psychometrika, 42(2):149–174.
- Muthen, B. (1983). Latent variable structural equation modeling with categorical data. Journal of Econometrics, 22(1–2):43–65.
- Noonan, R. and Wold, H. (1977). NIPALS Path Modelling with Latent Variables. Scandinavian Journal of Educational Research, 21(1):33–61.
- Nordin, K., Steel, J., Hoffman, K., and Glimelius, B. (2001). Alternative methods of interpreting quality of life data in advanced gastrointestinal cancer patients. British Journal of Cancer, 85(9):1265–1272.
- Phillips, R., Gandhi, M., Cheung, Y. B., Findlay, M. P., Win, K. M., Hai, H. H., Yang, J. M., Lobo, R. R., Soo, K. C., and Chow, P. K. H. (2015). Summary scores captured changes in subjects' QoL as measured by the multiple scales of the EORTC QLQ-C30. Journal of Clinical Epidemiology.
- Rabe-Hesketh, S. and Skrondal, A. (2008). Classical latent variable models for medical research. Statistical Methods in Medical Research, 17(1):5–32.
- Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalized multilevel structural equation modeling. Psychometrika, 69(2):167–190.
- Rasch, G. (1961). On General Laws and the Meaning of Measurement in Psychology. The Regents of the University of California.
- Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. CRC Press.
- Samejima, F. (1968). Estimation of Latent Ability Using a Response Pattern of Graded Scores¹. ETS Research Bulletin Series, 1968(1):i–169.
- Schwartz, C. E. and Sprangers, M. A. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Social science & medicine, 48(11):1531–1548.
- Skrondal, A. and Rabe-Hesketh, S. (2004). Generalized latent variable models: multilevel, longitudinal, and structural equation models. Chapman and Hall/CRC, Boca Raton, FL; London.
- Sprangers, M. A. and Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. Social science & medicine (1982), 48(11):1507–1515.
- Titman, A. C., Lancaster, G. A., and Colver, A. F. (2013). Item response theory and structural equation modelling for ordinal data: Describing the relationship between KIDSCREEN and Life-H. Statistical Methods in Medical Research.
- Tutz, G. (1990). Sequential item response models with an ordered response. British Journal of Mathematical and Statistical Psychology, 43(1):39–55.

Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least squares. In Multivariate Analysis., pages 391–420. Academic Press.

Wold, H. (1985). Partial Least Squares. In Encyclopedia of Statistical Sciences, volume 6, pages 581–591. John Wiley & Sons, New York.

Étude longitudinale mono-dimensionnelle par un mélange de modèles mixtes

Nous faisons ici l’hypothèse que la population d’intérêt est hétérogène, et qu’il existe des sous-populations homogènes qui se distinguent de par leur trajectoire moyenne de HRQoL. Un premier objectif de ce chapitre est de distinguer ces sous-populations homogènes en utilisant des mélanges de modèles mixtes. Une fois les sous-populations homogènes (classes latentes) identifiées, l’appartenance à ces classes latentes sera expliquée par différentes observations (covariables) recueillies au cours de l’étude. Ce second objectif permettra de dégager des profils de patients associés aux différentes trajectoires latentes de HRQoL. Cette dernière partie est motivée par une étude psychosociale visant à expliquer l’évolution de la HRQoL de patientes atteintes d’un cancer du sein selon des covariables cliniques, sociodémographiques et enfin des traits de personnalité.

Sommaire

5.1	Introduction	142
5.2	Modèles de mélange	143
5.2.1	Approche IRT	146
5.2.2	Approche CTT pour données gaussiennes	147
5.2.3	Approche CTT pour données ordinales	148
5.3	Application	149
5.3.1	L’étude « Moral »	149
5.3.2	Classification	152
5.3.3	Étude de la classification <i>a posteriori</i>	159
5.4	Discussion	169
5.5	Compléments	172
5.5.1	Estimation par l’algorithme EM pour un mélange de LMM	172
5.5.2	Résultats additionnels de l’étude « Moral »	177

5.1 Introduction

Dans les chapitres précédents, nous avons proposé différentes modélisations statistiques pour répondre aux problèmes méthodologiques dans le cadre des essais cliniques en cancérologie pour l'analyse longitudinale de la HRQoL. Dans les chapitres 2 et 3, nous nous sommes intéressés aux difficultés engendrées par la nature ordinale multiple des données. Puis dans le chapitre 4, l'étude globale de la HRQoL a motivé la proposition d'une nouvelle modélisation. Cependant, une complexité supplémentaire dans les données de HRQoL n'a pas été prise en compte. En effet, une hétérogénéité non observée au sein de la population peut exister. Des sous-populations homogènes de patients peuvent se distinguer par les données observées et s'expliquer par un profil particulier non observé a priori. Arriver à discerner ces sous-populations pourrait alors permettre de mieux comprendre l'évolution de la HRQoL pour des profils particuliers de patients.

Dans cette dernière partie, nous faisons l'hypothèse que la population d'intérêt n'est pas homogène, et qu'il existe K sous-populations homogènes non contrôlées (Martinez et al., 2009) qui se distinguent par différents profils moyens de HRQoL. Dans ce chapitre, l'objectif est double :

- Déterminer les sous-populations homogènes ainsi que les trajectoires latentes de HRQoL qui les caractérisent ;
- Expliquer la classification obtenue par l'intermédiaire de variables explicatives/prédictives.

Lors du recrutement d'un essai clinique ou d'une étude sur une cohorte de patients, un tri qui est déjà effectué suivant les critères d'inclusion et d'exclusion pré-définis. Dans ce contexte, nous sommes en présence d'une population homogène particulière, répondant à des critères objectifs. Une étude sur ce type de données induit d'ores et déjà un trait commun de la population d'intérêt. Le premier objectif, la classification, revient à déterminer si dans un premier temps la population est homogène, et si elle ne l'est pas, opérer des regroupements en classes homogènes d'un ensemble de patient. Dans cette approche, nous ne souhaitons pas discriminer entre deux groupes de traitement, mais plutôt établir une partition des sous-populations homogènes (Lourme, 2011) selon des trajectoires de HRQoL similaires, afin de voir émerger certains patterns pouvant expliquer au mieux l'évolution de la HRQoL.

Le second objectif est d'expliquer l'appartenance à une classe latente en fonction des variables explicatives à disposition. Dans la plupart des études en cancérologie, les données de HRQoL sont accompagnées d'observations cliniques, biologiques et sociodémographiques. D'un point de vue pratique, cette approche suscite plusieurs intérêts dont le principal est l'amélioration de la prise en charge des patients. En effet, l'identification de sous-populations de patients présentant une détérioration de la HRQoL au cours du temps permettrait de développer des actions spécifiques afin de les soutenir et d'agir dans la mesure du possible contre cette détérioration.

Concernant la classification non supervisée des données, plusieurs approches sont proposées dans la littérature. Sans connaissance ou hypothèse particulière sur les données, les méthodes dites de partitionnement (classification hiérarchique, K means)

sont utilisées pour faire de la classification. Elles reposent sur des considérations géométriques visant à regrouper les observations les plus "proches" et ne font pas appel à de la modélisation statistique. Cependant, ces méthodes ne prennent pas en compte la variabilité induite dans les données, et la comparaison de partitions obtenues sur le critère d'homogénéité qui les caractérise est plus délicate. La notion d'homogénéité se traduisant ici par le fait que les données des individus d'une même sous-population sont issues d'une même distribution. Un des objectifs de la classification dans l'étude longitudinale de la HRQoL est la comparaison des différentes trajectoires latentes qui caractérisent les sous-populations homogènes obtenues. Pour l'étude de la HRQoL, nous avons des hypothèses quant à la modélisation statistique des trajectoires de HRQoL. Une approche probabiliste permettant de répondre à tous les objectifs de ce chapitre est l'approche par modèle de mélange. Dans la suite, nous présentons les modèles de mélange pour la classification de données, et plus précisément pour les données de HRQoL. Ils ont un avantage certain sur les autres méthodes de classification qui réside dans l'inférence statistique rendue possible par l'approche probabiliste.

Enfin, nous illustrerons les méthodes sur les données de l'étude MORAL concernant 132 patientes atteintes d'un cancer du sein. Les mesures de QdV sont issues de la collecte de l'auto-questionnaire EORTC QLQ-C30 à différentes visites au cours du traitement et du suivi. Le recueil des données incluant des variables sociodémographiques (âge, niveau d'études, situation familiale, ...), médicales (taille de la tumeur, grade, ...), ainsi que des données psychosociales (anxiété, dépression, soutien social, optimisme, ...) est réalisé à 6 visites (à la chirurgie puis à 1, 4, 7, 10 et 13 mois).

5.2 Modèles de mélange

Dans cette section, nous supposons l'existence de différents profils de patients (non observables) qui résultent des évolutions distinctes de la HRQoL. Les modèles de mélange apportent donc une réponse rigoureuse et précise. Un mélange fini est une combinaison linéaire de lois de probabilités $(f_k)_{k=1,\dots,K}$ donnée par

$$f(\bullet; \Upsilon) = \sum_{k=1}^K p_k f_k(\bullet; \lambda_k). \quad (5.1)$$

Les K lois de probabilités sont appelées les composantes du mélange et les coefficients p_k , appelés les proportions de mélange, sont strictement positifs pour chaque classe $(C_k)_{k=1,\dots,K}$ tels que $\sum_{k=1}^K p_k = 1$. Le paramètre du mélange est donné par $\Upsilon = (p_1, \dots, p_K, \lambda_1, \dots, \lambda_K)$, où λ_k ($k = 1, \dots, K$) représente les paramètres de la composante k . Les composantes du mélange peuvent être définies par de nombreuses lois de probabilités. Dans la majorité des cas, les lois appartiennent à une même famille et dépendent des données à représenter.

Considérons Y_1, \dots, Y_n un échantillon aléatoire de taille n où Y_i est une variable aléatoire de densité $f(y_i; \Upsilon)$. Dans le cadre du modèle de mélange (5.1), les données

sont considérées comme incomplètes car les composantes $(\mathcal{C}_k)_{k=1,\dots,K}$ auxquelles appartiennent les observations ne sont pas connues. Pour le formaliser, on associe à chaque Y_i une variable latente Z_i , indicateur d'appartenance à la composante \mathcal{C}_k pour l'unité statistique i de telle sorte que :

- Z_i est un vecteur binaire $Z_i = (Z_{i1}, \dots, Z_{iK})$ distribué suivant une loi multinomiale de paramètre (p_{i1}, \dots, p_{iK}) , où $Z_{ik} = 1$ si l'individu i appartient à la \mathcal{C}_k et 0 sinon ;
- $\forall k \in \{1, \dots, K\}$, $(Y_i | Z_{ik} = 1) \sim f_k(y_i; \lambda_k)$.

La vraisemblance des données complètes de Υ s'écrit dans ce contexte :

$$\mathcal{L}(\Upsilon; y, z) = \prod_{i=1}^n \prod_{k=1}^K [p_k f_k(y_i; \lambda_k)]^{z_{ik}}, \quad (5.2)$$

et la log-vraisemblance associée :

$$\ell(\Upsilon; y, z) = \sum_{i=1}^n \sum_{k=1}^K z_{ik} p_k + \sum_{i=1}^n \sum_{k=1}^K z_{ik} f_k(y_i; \lambda_k). \quad (5.3)$$

Dans le contexte paramétrique où l'échantillon est supposé provenir de (5.1), l'estimation de Υ à partir de la maximisation de (5.3), conduit à une partition probabiliste des individus suivant le calcul de la probabilité π_{ik} que l'individu i soit issu de la composante k . Cette probabilité d'appartenance à la classe \mathcal{C}_k est estimée suivant les données observées de l'individu i et les valeurs des paramètres telle que :

$$\pi_{ik} = \frac{p_k f_k(y_i; \lambda_k)}{\sum_{l=1}^K p_l f_l(y_i; \lambda_l)}. \quad (5.4)$$

En utilisant la méthode du Maximum A Posteriori (MAP), les observations de l'individu i sont affectées à la classe latente \mathcal{C}_k la plus probable *a posteriori*. Soit,

$$k = \operatorname{argmax}_{k \in \{1, \dots, K\}} (\pi_{ik}).$$

Dans la classification de données à l'aide de mélange, l'avantage est de bénéficier d'outils puissants et adaptés pour déterminer le nombre de groupes comme les critères de choix de modèle. Dans notre approche, plusieurs modèles \mathcal{M}_j ($j = 1, \dots, J$) sont en compétition de par le nombre J de classes latentes considérées. Parmi les plus connus, les critères généraux tels que le AIC (*Akaike information criterion*, (Akaike, 1973)) ou le BIC (*bayesian information criterion*, (Schwarz, 1978)) accordent de l'importance à l'adéquation du modèle aux données. Le critère proposé par Akaike est défini par :

$$AIC(\mathcal{M}_j) = -2\ell(\hat{\theta}_j; x) - 2p,$$

où ℓ est la log-vraisemblance du modèle \mathcal{M}_j et p désigne la taille du paramètre θ_j .

Le BIC est quant à lui défini par :

$$BIC(\mathcal{M}_j) = -2\ell(\hat{\theta}_j; x) - p \log(n). \quad (5.5)$$

Ces deux critères à minimiser sont basés sur deux termes : le terme de vraisemblance qui favorise le modèle le plus complexe, et le terme de pénalisation qui est une fonction croissante du nombre de paramètres. Dans la littérature, il a été montré que le BIC (pénalisation dépendante du nombre d'observations) a un bon comportement pour le choix du modèle et celui du nombre de composantes, tandis que l'AIC a tendance à surestimer le nombre de composantes (Keribin, 2000).

De plus, les deux critères jugent essentiellement l'adéquation du modèle aux données et non la qualité de la partition. Le BIC ne tient pas compte de la séparabilité des classes et détermine le nombre de groupes suivant le nombre le plus probable de sous-populations homogènes de l'échantillon étudié. La partition est donc sensible aux fluctuations d'échantillon, ce qui peut avoir tendance à surestimer le nombre de groupes dans des échantillons de petite taille.

Pour prendre en compte la qualité de la partition, il a été proposé d'ajouter une pénalisation prenant en compte la séparabilité des classes. Le critère ICL (*Integrated completed likelihood*, (Biernacki et al., 2000)) est une alternative robuste au BIC. Il correspond au BIC auquel est ajouté une pénalisation spécifique à la classification :

$$ICL(\mathcal{M}_j) = BIC(\mathcal{M}_j) - \sum_{i=1}^n \sum_{k=1}^K \hat{z}_{ik} \ln(\pi_{ik}) \quad (5.6)$$

où π_{ik} est la probabilité conditionnelle pour l'individu i d'appartenir à la classe k ($k = 1, \dots, K$) (voir équation (5.4)), et \hat{z}_{ik} est définie à partir de la plus grande des probabilités conditionnelles π_{ik} ($k = 1, \dots, K$) telle que :

$$\hat{z}_{ik} = 1 \Leftrightarrow \pi_{ik} \geq \pi_{ij} \quad (j = 1, \dots, K). \quad (5.7)$$

Cette pénalisation est d'autant plus négligeable que les groupes sont séparés ou peu mélangés. A contrario, cela peut entraîner également une sous-estimation de ce nombre de classes pour un mélange faiblement séparé.

Dans ce chapitre, nous présentons trois types de modèles de mélange afin de mettre en exergue les trajectoires latentes de chaque groupe de patients, et nous faisons l'hypothèse que ces trajectoires latentes ont une structure de LMM. Pour chaque approche, la trajectoire de HRQoL d'un individu i appartenant à \mathcal{C}_k est caractérisée par le processus latent défini par :

$$\theta_i(t_{iv})|_{z_{ik}=1} = x_{1i}(t_{iv})' \beta_k + x_{2i}(t_{iv})' \nu + u_i(t_{iv})' \xi_i, \quad (5.8)$$

où

- β_k le vecteur des q_1 effets fixes associé à \mathcal{C}_k , $x_{1i}(t_{iv})$ son vecteur design au temps t_{iv} ;
- ν le vecteur des q_2 effets fixes commun à toutes les classes latentes, $x_{2i}(t_{iv})$ son vecteur design au temps t_{iv} ;

- ξ_i le vecteur des q effets aléatoires individuels normalement distribués associé à \mathcal{C}_k , de moyenne nulle et de matrice de variance covariance Σ_k , et $u_i(t_{iv})$ son vecteur design associé au temps t_{iv} .

Sous l'hypothèse d'hétérogénéité de l'échantillon, il est intéressant d'assouplir la modélisation de ce processus en fonction du temps. Cette relation ne sera plus considérée comme linéaire, mais comme polynomiale pour permettre une évolution plus flexible au cours du temps de la HRQoL et une meilleure distinction des trajectoires des classes. Notons le profil moyen associé à la classe latente \mathcal{C}_k tel que :

$$\theta_k(t) = X_{1i}(t)\beta_k + X_{2i}(t)\nu. \quad (5.9)$$

Concernant les données de HRQoL, plusieurs distributions de probabilité peuvent être considérées en référence aux chapitres précédents. Si le score de HRQoL est considéré comme une bonne représentation du niveau de la HRQoL (CTT), nous pouvons supposer un mélange gaussien. Par contre, si nous voulons prendre en compte les données brutes directement issues du questionnaire, un mélange de modèles IRT sera alors proposé.

5.2.1 Approche IRT

Pour le mélange de modèles IRT, nous favorisons l'utilisation du modèle IRT (*cumulatif, gaussien, proportionnel*)¹ qui semble particulièrement approprié de par ses caractéristiques détaillées dans le chapitre 3. Remarquons que dans ce modèle de mélange, il ne semble pas pertinent de considérer des paramètres d'item spécifiques aux classes. Les paramètres d'item (les paramètres de difficulté et de discrimination) sont considérés communs à l'ensemble des K classes latentes. Nous aurons ainsi une similarité dans l'interprétation du processus latent sous jacent au modèle IRT et celui modélisé par le LMM.

Notons $y_{iv}^{(j)}$ l'observation associée à l'individu i au temps t_{iv} et à l'item j . Nous modélisons alors :

$$\begin{aligned} \Pr\left(Y_{iv}^{(j)} \leq m | \theta_i^*(t_{iv}), \delta_j, \alpha_j\right) &= \Phi(\alpha_j(\delta_{jm} - \theta_i^*(t_{iv}))) \\ &= \Phi(\eta_{ivjm}), \end{aligned} \quad (5.10)$$

avec :

- $m = 0, \dots, M_j$;
- $\theta_i^*(t_{iv})$ est le trait latent au temps t_{iv} ;
- α_j est le paramètre de discrimination associé à l'item j ;
- $\delta_j = (\delta_{j1}, \dots, \delta_{jM_j})$ le vecteur des paramètres de difficulté de l'item j à M_j paramètres, communs à l'ensemble des K classes du mélange;
- Φ la fonction de répartition de la loi normale centrée-réduite.

1. le modèle (*cumulatif, logistique, proportionnel*) aurait pu également être considéré;

La trajectoire latente définie en (5.8), sous-jacente au modèle est :

$$\theta_i(t_{iv}) = \theta_i^*(t_{iv}) + \varepsilon_{iv},$$

telle que $y_{iv}^{(j)} = m$ si $\theta_i(t_{iv}) \in [\delta_{jm}; \delta_{j,m+1}[$ où $\delta_{j0} = -\infty$ et $\delta_{j,M_j+1} = +\infty$.

La vraisemblance s'écrit alors :

$$\mathcal{L}(\Upsilon; y) = \prod_{i=1}^n \sum_{k=1}^K p_k L_i(\Upsilon), \quad (5.11)$$

où

$$L_i(\Upsilon) = \int \prod_{v=1}^{n_i} \prod_{j=1}^J \prod_{m=0}^{M_j} \left[\Phi\left(\eta_{ivjm}^{(k)}\right) - \Phi\left(\eta_{ivj,m-1}^{(k)}\right) \right]^{y_{ivm}^{(j)}} \phi(\xi_i | z_{ik} = 1) d\xi_i.$$

avec

- $y_{ivm}^{(j)} = 1$ si $y_{iv}^{(j)} = m$ pour $m = 0, \dots, M_j$;
- J est le nombre d'items considérés pour la dimension étudiée ;
- ϕ la fonction de répartition de la loi normale centrée réduite.

L'estimation de ce type de mélange est très couteuse en temps de calcul. Un outil pour l'estimation de ce mélange a été proposé par [Proust-Lima et al. \(2015a\)](#) (le programme *HETMIXSURV* en Fortran 90) . Une approche alternative à l'IRT est la CTT où le score est considéré comme gaussien. En effet, malgré les limites de cette dernière énoncées dans les chapitres précédents, nous avons vu que cette approche donne de bons résultats en pratique.

5.2.2 Approche CTT pour données gaussiennes

Dans le cas où le score de HRQoL est une bonne représentation du niveau HRQoL, nous pouvons nous orienter vers un mélange de LMM. Soit $\mathbf{y} = (y'_1, \dots, y'_n)'$ le vecteur des scores observés composé de n vecteurs associés à chaque individu. Le vecteur $(y_i)_{i=1, \dots, n}$ (de taille n_i) comprend l'ensemble des mesures répétées y_{iv} du patient i , où v est l'indice de la visite à laquelle la mesure est effectuée. Conditionnellement aux classes, le LMM appliqué aux mesures y_{iv} s'écrit de manière générale tel que :

$$Y_i | Z_{ik}=1 = X_{1i}\beta_k + X_{2i}\nu + U_i\xi_i + \varepsilon_i. \quad (5.12)$$

Dans ce cas linéaire, f_k représente la densité gaussienne associée à la classe \mathcal{C}_k d'espérance $X\beta_k + X_2\nu$ et de variance-covariance $\Gamma_k = U\Sigma_kU' + \tau^2 Id_{n_k}$, avec :

- $\xi_i | Z_{ik} = 1 \sim \mathcal{N}(\mathbf{0}, \Sigma_k)$ avec Σ_k^2 la matrice de covariance propre à la classe \mathcal{C}_k , U_i la matrice d'incidence associée ;
- $\varepsilon_i \sim \mathcal{N}_{n_i}(0, \tau^2 Id_{n_i})$ avec τ^2 la variance sur les erreurs de mesure supposée indépendante de la classification ;
- β_k est le vecteur des effets fixes de taille q_1 associé à \mathcal{C}_k ;
- ν est le vecteur des effets fixes de taille q_2 commun à l'ensemble des K classes ;

- X_{1i} est la matrice d'incidence de taille $(n_i \times q_1)$ associée aux q_1 effets fixes spécifique à \mathcal{C}_k et X_{2i} celle de taille $(n_i \times q_2)$ associée aux q_2 effets fixes commun aux K classes ;
- $\forall i$, les variables aléatoires ξ_i et ε_i sont supposées indépendantes ; $\forall i, j \in \{1, \dots, n\}^2$ et $i \neq j$, ξ_i et ξ_j , respectivement ε_i et ε_j , sont indépendants.

Par la suite, nous considérons le LMM le plus simple pour modéliser les trajectoires latentes de HRQoL :

$$y_{iv}|z_{ik}=1 = \beta_{k0} + t_{iv}\beta_{k1} + t_{iv}^2\beta_{k2} + t_{iv}^3\beta_{k3} + \xi_i + \varepsilon_{iv} . \quad (5.13)$$

Les différentes composantes se distinguent en premier lieu par la partie fixe qui permet de déterminer leur trajectoire moyenne sous forme polynomiale. Les variances de l'effet aléatoire et de l'erreur résiduelle peuvent être considérées communes entre les classes $(\mathcal{C}_k)_{k=1, \dots, K}$ ou propres à chacune d'elle pour plus de flexibilité.

En Annexe 5.5.1, nous présentons une méthode d'estimation des paramètres par l'algorithme *EM* (Martinez et al., 2009). Cette approche s'appuie sur l'hypothèse gaussienne des observations ; hypothèse qui peut être raisonnablement mise en cause pour les données de HRQoL.

Une modélisation de type IRT semble mieux adaptée que l'approche CTT pour données gaussiennes. Un compromis entre les deux approches décrites précédemment est de considérer le score comme une variable discrète.

5.2.3 Approche CTT pour données ordinales

Le score relatif à la somme des J items allant de 0 à M , où $M = \sum_{j=1}^J M_j$ est le score maximal, peut alors être modélisé par un modèle cumulatif. En effet, la variable réponse Y_{iv} est discrète à $M + 1$ modalités de réponse. Aucune distinction entre les différents items n'est alors prise en compte. Cela revient à la modélisation présentée dans l'équation (5.10) sans la considération de paramètres de discrimination étant donné que le score découle d'une seule échelle ($J = 1$ et $\alpha_1 = 1$). Les paramètres de difficulté sont quant à eux des paramètres de seuil permettant de distinguer les différents scores observables, tels que :

$$\{Y_{iv} = m\} \Leftrightarrow \theta_i(t_{iv}) \in [\delta_m; \delta_{m+1}[,$$

avec $m = 0, \dots, M$, $\delta_0 = -\infty$ et $\delta_{(M+1)} = +\infty$. Cette approche rentre dans le cadre de la CTT où les caractéristiques du score-somme sont mieux prises en compte dans la modélisation. L'avantage sur les modèles IRT est la disponibilité de programmes dans les logiciels courants pour modéliser ce type de réponse.

Quelle que soit la modélisation prise en compte, nous cherchons à définir les classes latentes suivant la trajectoire de HRQoL. La composante du mélange se distingue essentiellement par les paramètres de la trajectoire et non sur l'ensemble des paramètres du modèle.

5.3 Application

Pour ces derniers travaux de thèse, nous avons travaillé en collaboration avec Épidaure, le pôle de prévention de l'Institut régional du Cancer de Montpellier. Épidaure est dédié à la prévention des cancers et à l'éducation pour la santé. Il développe des actions de prévention pour la population générale (action sur le tabac, l'alcool, la nutrition et le soleil) et des programmes d'éducation thérapeutique pour le patient. Épidaure est un véritable laboratoire d'idées et d'échanges permettant de concevoir des programmes, de tester et de mettre au point des outils innovants d'éducation pour la santé. Ces innovations sont ensuite diffusées et mises au service de la population générale, des patients et de leur famille au niveau régional et au-delà. Trois axes de recherche y sont privilégiés :

1. La prévention primaire, qui a pour objectif la réduction des expositions à risque et la réduction du nombre de cancers ;
2. Le dépistage, où l'objectif est d'augmenter la participation au dépistage organisé en étudiant les freins et les leviers de la population ciblée et des médecins ;
3. L'étude des facteurs psycho-sociaux de la qualité de vie, qui a pour objet l'amélioration de la qualité de vie des patients atteints de cancer et de leurs familles, ainsi qu'une meilleure prise en compte de leurs besoins par les soignants.

5.3.1 L'étude « Moral »

L'étude « Moral » a été portée par Florence Cousson-Gélie (directrice scientifique du pôle de prévention Épidaure, ICM) et a été soutenue financièrement par l'INCa (Grant « INCa-DGOS-Inserm 6045 »). Cette étude a été réalisée afin de s'intéresser au parcours de la personne malade sans la réduire à son simple corps biologique, et vise à évaluer l'impact des facteurs psychologiques et sociaux sur la HRQoL chez des patientes atteintes de cancer du sein. De plus, elle évalue le processus dynamique de la relation entre l'individu et son environnement, c'est-à-dire comment les réactions cognitives, comportementales et émotionnelles peuvent influencer l'ajustement psychologique et l'évolution de la maladie chez des patientes atteintes d'un cancer. Une des motivations de cette étude a été de mettre en évidence comment la HRQoL et l'évolution de la maladie sont influencées par certaines variables de personnalité, certaines réactions vis-à-vis de la maladie et des traitements et la relation soignant-soigné.

L'étude longitudinale de la HRQoL est réalisée ici suivant l'approche par mélange de modèles mixtes, en supposant donc qu'il existe une structure latente des données. L'objectif est de faire émerger des profils de patientes avec une évolution différente de HRQoL, mais également de pouvoir interpréter, expliquer, décrire cette classification à l'aide de variables explicatives de natures diverses.

Données recueillies

L'étude « Moral » propose un panel de mesures complémentaires à celles de la HRQoL. Ces données ont été recueillies pour 132 patientes âgées de 30 à 70 ans à 6

Chapitre 5. Étude longitudinale mono-dimensionnelle par un mélange de modèles mixtes

TABLE 5.1: Résumé des différents outils de mesures psycho-sociales associés aux temps de suivi pour lesquels ils ont été utilisés. Les temps de suivi $T_0, T_1, T_2, T_3, T_4, T_5$ correspondent respectivement à la chirurgie, à 1 mois, 4, 7, 10 et 13 mois.

Questionnaires	Temps de suivi
Questionnaires EORTC QLQ-C30	$T_0 \rightarrow T_5$
Questionnaire HAD	$T_0 \rightarrow T_5$
Question de Soutien Social Perçu (SSQ-6)	$T_0 \rightarrow T_5$
Questionnaire QRI	$T_0 \rightarrow T_5$
Échelle LOT-R	T_0
Échelle NEO-PI R	T_0
Questionnaire Mental MAC (<i>Adjustment to Cancer Scale</i>)	$T_1 \rightarrow T_5$

visites : à la chirurgie (T_0) puis à 1, 4, 7, 10 et 13 mois (respectivement T_1, T_2, T_3, T_4 et T_5). Elles comprennent des variables sociodémographiques (âge, niveau d'étude, situation familiale, engagement professionnel), médicales (taille de la tumeur, envahissement ganglionnaire, type de chirurgie, chimiothérapie, radiothérapie) mesurées à la chirurgie, ainsi que des mesures psycho-sociales issues d'auto-questionnaire aux différents temps (voir Table 5.1).

Les données sociodémographiques (*) et médicales (★) que nous avons prises en comptes dans l'analyse sont :

- * Âge : inférieur à 50 ans, entre 50 et 60 ans, plus de 60 ans ;
- * Niveau d'étude : inférieur au Baccalauréat (Pas d'études, CAP-BEP) et supérieur ou égale au Baccalauréat (Baccalauréat, BAC+2, BAC+3 et plus) ;
- * Situation professionnelle : En activité, Sans emploi, Retraitée ;
- * Situation familiale : Séparée, Seule, Mariée ;
- * Nombre d'enfants $\{0, 1, 2, 3^+\}$ (qualitatif) ;
- ★ Grade histologique² : Grade 1 (pronostique favorable), Grade 2 (pronostique moyen), Grade 3 (évolution péjorative) ;
- ★ Type de chirurgie : Tumorectomie, Mastectomie ;
- ★ Taille de la tumeur : $< 15mm$ et $\geq 15mm$.

Les précisions sur les mesures psychosociales ainsi que les questionnaires associés sont les suivantes :

- L'anxiété et la dépression chez les patientes atteints de cancer sont évaluées par le questionnaire HAD (Zigmond and Snaith, 1983). Il est composé de 7 items permettant l'évaluation de l'anxiété et de 7 autres pour celle de la dépression. Ces deux regroupements d'items sont résumés respectivement en deux scores allant de 0 à 21. À partir des scores qui découlent de ces items, un score total est déduit. Certains auteurs proposent la catégorisation du statut anxieux (respectivement dépressif) des répondants (Bjelland et al., 2002) : 7 ou moins (absence de symptomatologie), 8 à 10 (symptomatologie douteuse), 11 et plus (symptomatologie certaine). Dans notre analyse, nous considérerons

2. Indicateur de pronostique en fonction de la taille de la tumeur, du nombre de ganglions et de la présence de métastases ;

deux classes pour l'anxiété (respectivement la dépression) : symptomatologie certaine pour un score supérieur ou égal à 9 (respectivement 8), et absence de symptomatologie pour un score inférieur à 9 (respectivement 8) ;

- La mesure de l'optimisme et du pessimisme des patientes à la chirurgie est réalisée par l'échelle LOT-R. Ces deux traits sont mesurés par l'intermédiaire d'un seul score construit sur 12 items (Scheier and Carver, 1985). Pour l'analyse, un score inférieur à 13 correspond au pessimisme et supérieur à 17 à l'optimisme (Kepka et al., 2013) ; Dans l'approche de la CTT, ce score présente une bonne représentation du niveau d'optimisme ou pessimisme de l'individu, où les deux aspects sont considérés aux deux extrémités d'un même continuum ;
- Le névrosisme (l'une des cinq caractéristiques de la personnalité avec l'ouverture, l'aspect consciencieux, extraversion, l'agréabilité) correspond à l'instabilité émotionnelle (prédisposition aux affects négatifs) et est mesuré au travers de l'échelle NEO-PI R. Deux classes sont considérées pour ce score : stabilité émotionnelle pour un score inférieur à la médiane (<102), instable sinon ;
- Parmi les questionnaires spécifiques au cancer, le dernier utilisé est le questionnaire Mental MAC qui permet d'évaluer dans quelle mesure un patient adopte l'une des cinq stratégies d'ajustement suivantes : esprit combattif, impuissance-désespoir, préoccupation et anxiété, fatalisme et évitement, et dénégation. À partir du manuel d'utilisation de ce questionnaire, une variable binaire est générée : j'adopte cette stratégie (oui) ou pas (non) (Cayrou et al., 2003) ;
- Le soutien social est mesuré par l'intermédiaire du questionnaire SSQ-6. Il est décomposé en deux dimensions comprenant 6 items chacun, qui permettent de construire un score de satisfaction du soutien social (de 0 à 36) et un autre lié à la disponibilité de ce soutien (score de 0-54) ;
- La qualité de la relation est mesurée par le questionnaire QRI (Pierce et al., 1991). Il comprend 25 items permettant la mesure de trois dimensions : conflit, intensité et disponibilité.

Données à expliquer

La HRQoL a été mesurée avec le questionnaire EORTC QLQ-C30 et également collectée au cours des 6 visites précédemment définies (Aaronson et al., 1993; Fayers et al., 2001). L'ensemble des dimensions de la HRQoL issues du questionnaire QLQ-C30 ne seront pas toutes analysées, seules les dimensions les plus pertinentes dans le contexte de l'étude « Moral » seront étudiées. Nous prendrons en compte le GHS qui est l'échelle du questionnaire représentant la qualité de vie globale. Ce score est cependant controversé dans la littérature car il serait sujet à un effet RS (Schwartz and Sprangers, 1999). De plus, du fait de son positionnement à la fin du questionnaire (items 29 et 30), Phillips et al. (2015) suggèrent que les réponses seraient de moins bonne qualité. Plusieurs scores de substitution ont été proposés. Par exemple, Nordin et al. (2001) proposent de considérer le score global compris dans l'intervalle $[-800; 600]$, ce score sera noté dans la suite « Nordin ». Ce score correspond à la différence entre la somme des scores des différentes dimensions fonctionnelles et du GHS (100 est le meilleur score) et la somme des dimensions symptomatiques exceptée la dimension associée aux difficultés financières (0 est le meilleur score). Dans le

cas où un score d'une dimension serait manquant, le score global le serait également.

Concernant les dimensions fonctionnelles, la dimension physique regroupant le plus grand nombre d'items du questionnaire (cinq items) et les dimensions émotionnelle et personnelle intéressantes à la vue des objectifs de l'étude Moral seront analysées. Concernant les dimensions symptomatiques, la fatigue (dimension symptomatique la plus conséquente), l'insomnie et la perte d'appétit qui sont les trois dimensions les plus liées à l'aspect psychosocial, ont également été choisies pour l'analyse. Par soucis de clarté dans le manuscrit, les résultats pour les dimensions fonctionnelles et le score global de Nordin sont présentés en section 5.5.2.

5.3.2 Classification

Dans cette première étape d'analyse des données, nous réalisons une classification de la population pour chacune des dimensions de HRQoL sélectionnées suivant les trajectoires de HRQoL. La variable temps associée aux différentes visites a été considérée comme une variable continue afin de modéliser les trajectoires sous forme polynomiale (ordre 3). Pour les dimensions comprenant plusieurs items (score global de Nordin, GHS, fonction physique, fonction émotionnelle, fatigue), l'approche par CTT pour données gaussiennes est privilégiée en considérant le score de HRQoL qui découle de la procédure de scoring proposée par l'EORTC (Fayers et al., 2001). Cela correspond à la modélisation (5.13) où les variances des variables aléatoires sont prises communes à l'ensemble de la population (après vérification que les prendre spécifiques aux classes n'améliore pas le modèle). En revanche, pour les dimensions à un seul item (insomnie, perte d'appétit), l'approche CTT pour données ordinales est préférée. Dans cette seconde approche, le processus latent $\theta_i(t_{iv})$ (équation (5.8)) est décomposé suivant la même structure que le score concernant les dimensions à plusieurs items. Nous profiterons de la classification établie sur les dimensions à un seul item pour initier une comparaison entre les deux approches CTT. Cette comparaison a pour objectif d'évaluer l'impact du score quand il est considéré comme une donnée continue gaussienne. Elle sera d'abord réalisée sur les données réelles puis sur des données simulées. Nous sommes conscients de l'absurdité de cette comparaison, où le mélange gaussien n'est forcément pas adapté pour des données ordinales, et que le mélange de GLMM pour données ordinales (*cumulatif, gaussien, proportionnel*) l'est particulièrement. Cela dit, la comparaison peut d'ores et déjà nous orienter sur les résultats potentiels de ce type de comparaison pour des dimensions constituées de plusieurs items. En effet, nous pouvons nous attendre à des différences sur le choix du nombre de classes, la classification des patients ou encore les trajectoires des classes latentes. Le choix K du nombre de classes latentes est effectué à l'aide des critères d'information BIC et ICL, associés respectivement aux équations (5.5) et (5.6). Les classifications suivant les deux approches CTT ont été réalisées sous le logiciel R à l'aide du package **lcmm** (Proust-Lima et al., 2015b).

Remarque : Le choix de K qui découle des critères BIC ou ICL peut conduire à des classes de faible effectif $\{n_k\}_{k=1,\dots,K}$. Sous l'objectif d'expliquer l'appartenance aux classes au travers d'un modèle de régression, la sous représentation de certaines

classes latentes va poser des problèmes en pratique. Pour éviter cela, nous considérons des règles pratiques complémentaires dans le choix de K :

- si $n_k < 5$ pour une classe $k \in \{1, \dots, K\}$, alors un nombre de classes $K - 1$ sera pris en compte ;
- si $5 \leq n_k < 10$ pour une classe $k \in \{1, \dots, K\}$, alors un nombre de classes $K - 1$ sera pris en compte si la trajectoire latente $\theta_k(t)$ ne renvoie pas à un niveau distinct des autres trajectoires latentes.

Approche CTT pour données gaussiennes

La Figure 5.1 retourne tous les éléments de la classification concernant la dimension liée au GHS. Les deux critères d'information (BIC et ICL) sont minimisés pour une valeur $K = 3$ (Figure 5.1a). Cependant, choisir $K = 3$ revient à considérer une classe latente constituée de 2 patientes, $K = 2$ est alors préféré. Ainsi, deux sous-populations homogènes d'effectifs relativement équivalents se distinguent suivant l'évolution de ce score (Figure 5.1b). Les deux trajectoires latentes associées s'opposent. La première (la classe D) renvoie à une détérioration de la HRQoL au cours du temps, et la seconde (la classe A) se caractérise par une amélioration de la HRQoL au cours du temps. La médiane (respectivement la moyenne) de la probabilité d'appartenir à la classe k ($k = 1, 2$) sachant les individus de la classe k est très proche de 1 (respectivement supérieure à 0.9).

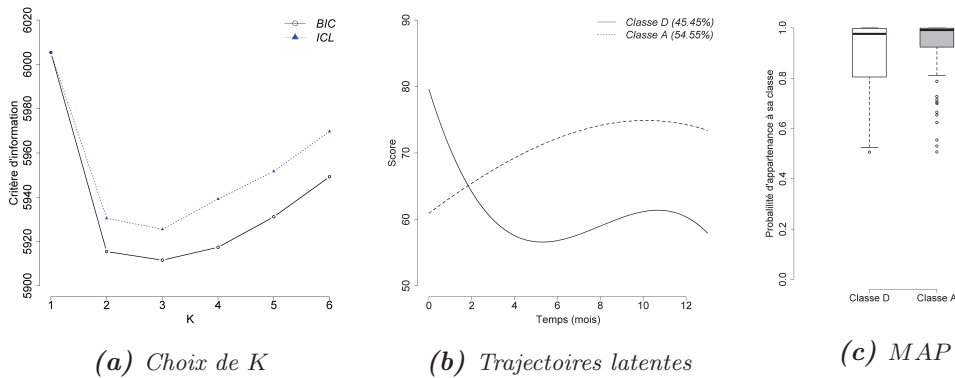


FIGURE 5.1: Classification pour le score du statut global de santé (GHS). Classe A (respectivement D) fait référence à la classe dont la HRQoL s'améliore (respectivement se détériore) au cours du temps.

Concernant la classification de la population d'intérêt pour la dimension de fatigue, c'est pour $K = 4$ que le BIC et l'ICL sont les petits (Figure 5.2a). Cependant, considérer quatre classes latentes revient à considérer une sous-population ayant un effectif constitué de trois patientes. Ne voulant pas traiter d'effectif inférieur à cinq individus (voire dix), nous considérons alors $K = 3$ sous populations homogènes pour la dimension de fatigue. Ces trois patientes caractérisées par une détérioration importante HRQoL (classe 4, Figure 5.2b) ont été réaffectées à la classe D dans la Figure 5.2c (anciennement classe 3, Figure 5.2b). Cela a donné lieu à un ajustement des trajectoires latentes des classes 1, 2 et 3 (Figure 5.2b) et une réaffectation de

certaines patientes. La Figure 5.2d illustre les probabilités *a posteriori* obtenues pour $K = 3$. La probabilité médiane d'appartenir à la classe d'appartenance est supérieure à 0.9. Le mélange met bien en évidence des population distinctes.

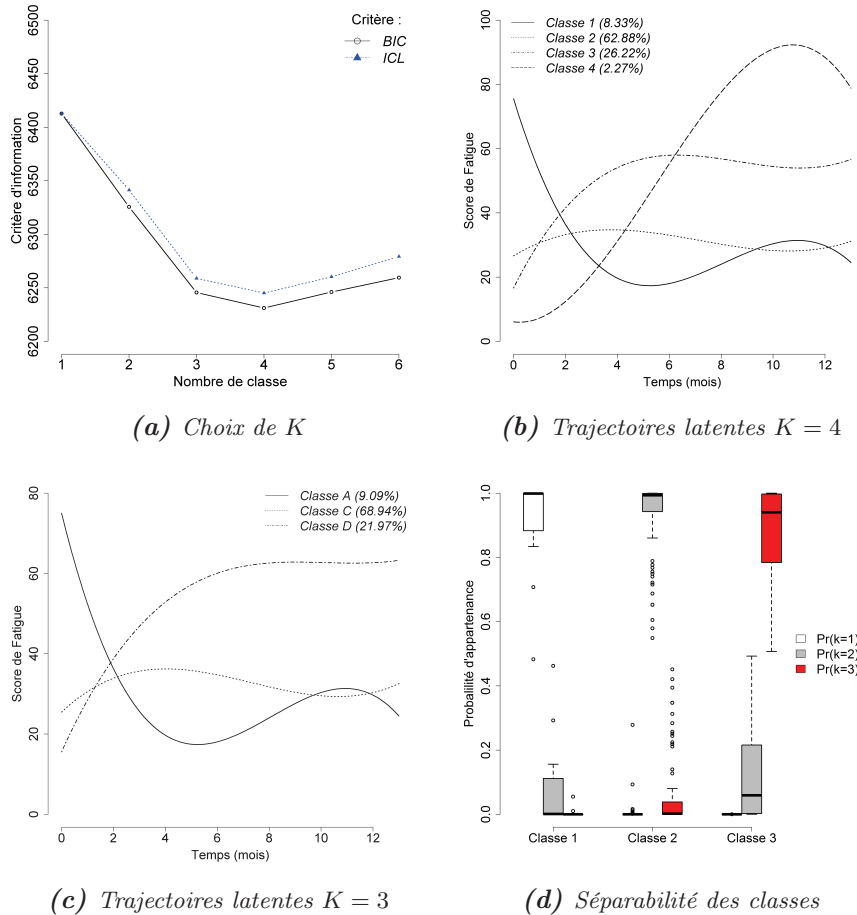


FIGURE 5.2: Classification relative à la dimension de fatigue. Classe A (respectivement C et D) fait référence à la classe dont la HRQoL s'Améliore (respectivement reste Constante et se Détérioré) au cours du temps.

Approche CTT pour données ordinales

Pour la classification liée à l'insomnie, $K = 4$ sous-populations homogènes se distinguent de par les trajectoires de HRQoL d'après les critères d'information (Figure 5.3a). Cependant, considérer quatre classes latentes revient à prendre en compte une quatrième classe composée de $n_4 = 6$ patientes. À la vue des information apportées par la Figure 5.3, nous préférons considérer trois sous-populations homogènes, avec des effectifs plus équilibrés. Pour le mélange de GLMM pour données ordinales, les différents paramètres de seuil permettant de discrétiser le continuum sont représentés par les droites bleues en pointillé (Figures 5.3b et 5.3c). En termes d'interprétation,

la classe C est caractérisée par une évolution Constante de la HRQoL (de l'insomnie). En effet, le processus latent θ_C^* diminue au cours du temps et $\theta_C^* < \delta_1$. Cela renvoie au fait que pour cette classe latente, il est plus probable d'observer un score égal à 0 (Figure 5.3b, axe de droite : $Y = 0$), et ce d'autant plus au cours du temps. Cette classe est donc caractérisée par niveau moyen constant d'insomnie quantifié par un score de 0. Ainsi, nous pouvons voir les catégories de réponse qui sont les plus probables suivant la trajectoire latente. Pour les deux autres classes latentes, le score moyen observé change au cours du temps de telle sorte que la classe A renvoie à une amélioration (diminution de l'insomnie) et que la classe D renvoie à une détérioration du niveau de HRQoL (augmentation de l'insomnie) au cours du temps. Le niveau moyen d'insomnie à la chirurgie est par contre meilleur dans la classe D que dans la classe A.

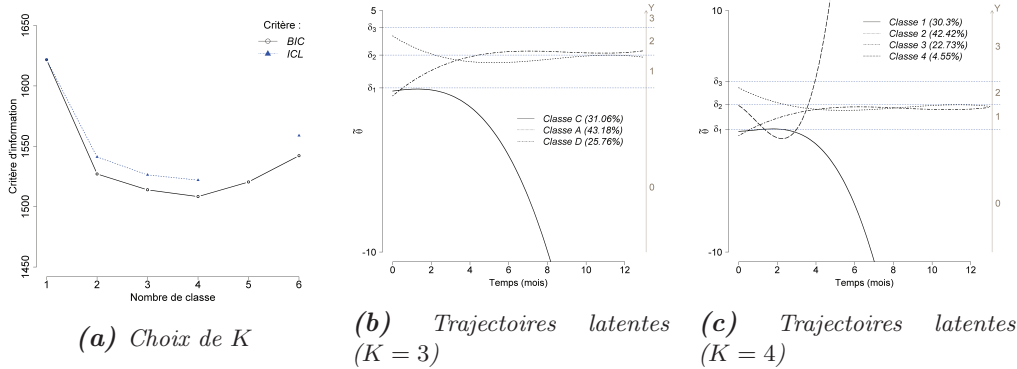


FIGURE 5.3: Classification relative à la dimension liée à l'insomnie (SL).

La Figure 5.4 présente la classification par mélange de GLMM pour données ordinales pour la dimension liée à la perte d'appétit. Il est intéressant de remarquer que le BIC et l'ICL sont en contradiction pour le choix du nombre de classes latentes K . En effet, le BIC suggère que la meilleure classification est celle qui considère $K = 3$ sous-populations homogènes alors que l'ICL $K = 2$ (Figure 5.4a). Ceci est dû au fait que lorsque trois classes latentes sont prises en compte, les probabilités d'appartenir à sa classe latente ne sont pas proches de 1 : les trajectoires ne sont pas suffisamment distinctes. Nous préconisons $K = 2$ afin d'étudier des sous-populations homogènes les plus distinctes possibles par rapport à leur trajectoire latente. Un argument supplémentaire au choix de $K = 2$ est la petitesse de l'effectif de la troisième classe $n_3 = 5$ lorsque que $K = 3$ (Figure 5.4c).

Bien que le mélange de LMM ne soit pas adapté pour une dimension à un seul item (avec quatre catégories de réponse), nous proposons dans cette partie de mettre en avant les différences et similitudes sur les classifications obtenues suivant les deux approches CTT pour la dimension AP (Figure 5.4). Le choix du nombre de classes est différent entre les deux approches. Pour le mélange de LMM (Figure 5.4d), $K = 6$ sous-populations homogènes sont le plus adaptées d'après les critères d'information (pour $K > 6$, certaines difficultés d'estimation sont rencontrées). Lorsque le score est considéré comme continu et gaussien, le nombre de classes latentes obtenu est

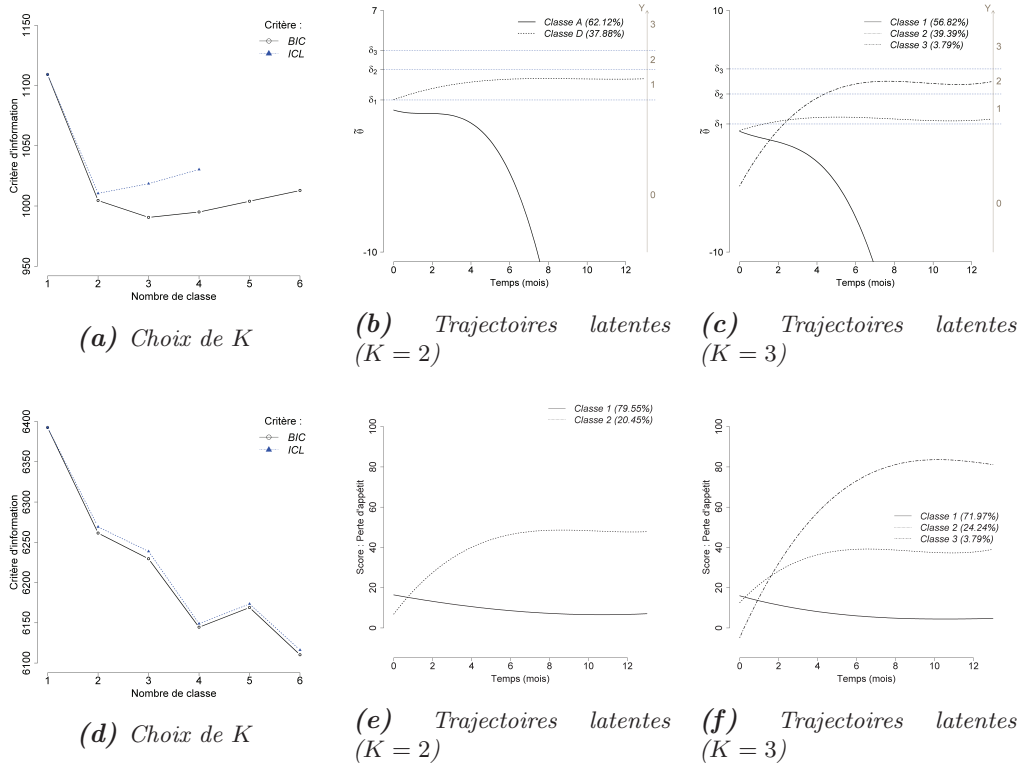


FIGURE 5.4: Classification relative à la dimension liée à la perte d'appétit (AP). Les figures 5.4a, 5.4b et 5.4c présentent les résultats de l'approche CTT pour données ordinales, et les figures 5.4d, 5.4e et 5.4f ceux de l'approche CTT pour données gaussiennes.

plus conséquent que lorsque que le score est considéré comme ordinal. Pour un K fixé ($K = 2, 3$), les trajectoires correspondantes nous renvoient une information identique pour les deux approches (Figures 5.4b et 5.4c pour l'approche pour données ordinales ; Figures 5.4e et 5.4f pour l'approche pour données gaussiennes). Par exemple, pour $K = 2$:

- Trajectoire associée à la classe 1 (ou A) : le niveau de HRQoL est bon à la baseline (faible perte d'appétit) et reste relativement constant au cours du temps ;
 - CTT pour données ordinales : $-\infty \leq \theta_1(t) < \delta_1$, ce qui laisse présager une majorité de réponse 0, un peu de réponse 1, un peu moins 2 et encore moins de 3 pour cette classe. Au cours du temps, les patients ont tendance à choisir des modalités plus petites ;
 - CTT pour données gaussiennes : niveau moyen de HRQoL aux alentours de 20 points sur le score (soit entre les valeurs du score 0 et 33 associées aux modalités 0 et 1), le score a tendance à diminuer au cours du temps ;
- Trajectoire associée à la classe 2 (ou D) :
 - CTT pour données ordinales : $\delta_1 \leq \theta_1(t) < \delta_2$, ce qui renvoie une modalité moyenne de 1. Au départ, les modalités 0 et 1 sont les plus observées pour

ensuite voir davantage la modalité 2 que la modalité 0 : détérioration du niveau de HRQoL (augmentation de la perte d'appétit) ;

- CTT pour données gaussiennes : bon niveau de HRQoL à la baseline (score proche de 0) qui se détériore au cours du temps avec une augmentation moyenne du score jusqu'à 50 (entre les valeurs 33 et 66 qui sont associées respectivement aux modalités 1 et 2).

Ces interprétations des trajectoires latentes suivant les deux approches résument bien la même évolution de la HRQoL.

En ce qui concerne l'attribution des patientes aux classes, les concordances entre ces deux approches CTT sont données par la Table 5.2. $\forall K = 2, 3$, l'attribution est similaire à plus de 80%, ce qui explique également les effectifs obtenus relativement proches suivant ces deux types de mélange, ainsi que la même information apportée par les trajectoires latentes. Le pourcentage d'attribution aux classes qui diffèrent est lié aux patientes qui doivent avoir une trajectoire entre les deux trajectoires moyennes. Il est intéressant de voir que pour $K = 3$, les deux approches considèrent la troisième classe latente regroupant les cinq mêmes patientes, caractérisées par une forte détérioration de la HRQoL.

TABLE 5.2: Tableaux de contingence entre les classifications obtenues suivant les deux approches CTT pour $K = 2, 3$ (résultats entre crochets pour $K = 2$). CTT¹ correspond à l'approche pour données gaussiennes et CTT² à l'approche pour données ordinales. Lorsque $K = 2$, 82% de la classification est similaire. Lorsque $K = 3$, 84% de la classification est similaire.

		CTT ¹		
		Classe 1	Classe 2	Classe 3
CTT ²	Classe 1	75 [82]	0 [0]	0
	Classe 2	20 [23]	32 [27]	0
	Classe 3	0	0	5

Pour confirmer les similarités obtenues sur ces deux approches pour classifier les patientes, nous proposons une étude de simulation.

Simulations pour comparer la classification des différents mélanges

Afin de confirmer les différences entre le mélange de LMM et le mélange de (*cumulatif, gaussien, proportionnel*), une étude de simulation basée sur les estimations obtenues sur le mélange lié à l'insomnie (trajectoires de la Figure 5.3b) a été effectuée. Pour cela, 500 jeux de données de 150 sujets ont été simulés. Nous considérons à chaque jeu de données trois sous-populations homogènes ($K = 3$) d'effectif similaire ($n_k = 50$ chacune). Les trajectoires individuelles sont définies par :

$$\theta_{ik}(t_{iv}) = \beta_{k0} + t_{iv}\beta_{k1} + t_{iv}^2\beta_{k2} + t_{iv}^3\beta_{k3} + \xi_i + \varepsilon_{iv}, \quad (5.14)$$

où $\xi_i \sim \mathcal{N}(0, \sigma_\xi^2)$, $\varepsilon_{iv} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ et les coefficients des paramètres utilisés pour simuler les données sont présentés dans la Table 5.3. Comme pour les données réelles, sept temps de mesure ont été pris en compte : à 0, 1, 3, 6, 9, 12 et 15 mois. Les

Chapitre 5. Étude longitudinale mono-dimensionnelle par un mélange de modèles mixtes

réponses simulées sont les réponses ordinales d'un item ($y_{iv} \in \{0, 1, 2, 3\}$). Elles sont déduites en fonction de $\theta_{ik}(t_{iv})$ et des coefficients des paramètres de difficulté (voir Table 5.3), telles que :

$$y_{ikv}^{(j)} = m \Leftrightarrow \theta_{ik}(t_{iv}) \in [\delta_m; \delta_{m+1}[,$$

où $m = 0, \dots, 3$, $\delta_0 = -\infty$ et $\delta_4 = +\infty$. Concernant le mélange de LMM, nous considérons le score de HRQoL calculé suivant les recommandations de l'EORTC : (moyenne des) réponses exprimées sur une échelle de 0 à 100. La trajectoire du score est alors modélisée suivant l'équation (5.13).

TABLE 5.3: Récapitulatif des coefficients qui ont permis de simuler les 500 jeux de données de 150 patients. Ces coefficients correspondent à ceux obtenus pour la classification liée à la dimension de l'insomnie.

k	n_k	Paramètres								
		β_0	β_1	β_2	β_3	σ_ξ^2	σ_ε^2	δ_1	δ_2	δ_3
1	50	0.000	0.163	-0.028	-0.018	1.4	1	0.2	2.2	4.0
2	50	3.426	-0.753	0.105	-0.004	1.4	1	0.2	2.2	4.0
3	50	-0.298	1.006	-0.117	0.004	1.4	1	0.2	2.2	4.0

La Table 5.4 retourne les résultats sur le choix du nombre de classes latentes K pour les deux mélanges considérés suivant le BIC et l'ICL. Concernant le mélange de CM (modèles (*cumulatif, gaussien, proportionnel*)), nous remarquons que $K = 1$ pour 5 des 500 cas simulés, autrement dit la population globale est homogène. Ce résultat n'est pas prendre en considération. En effet, pour ces 5 cas particuliers, le programme a rencontré des problèmes de convergence ce qui a engendré ces résultats. Cela reflète la difficulté d'estimation pour ce type de modèle. D'un point de vue général, nous retrouvons les hypothèses que nous avons émises selon les résultats sur données réelles. Concernant l'approche pour données ordinales, nous retrouvons bien les trois sous-populations homogènes simulées (494 fois sur 500, quel que soit le critère). Ce résultat est attendu étant donné que nous avons utilisé cette approche pour simuler les données. Par contre, le mélange de LMM a tendance à surestimer le nombre de sous-populations homogènes, bien que l'utilisation du critère ICL semble plus pertinente que celle du BIC pour ce mélange. Ces résultats de simulation confirment les différences mises en avant sur les données réelles entre ces deux approches pour une dimension à un item avec quatre modalités.

De plus, nous avons contrôlé les taux de bien-classés et mal-classés sachant $K = 3$. Nous trouvons pour la classification obtenue via le mélange de CM un taux de bien classés de 0.94, alors que pour la classification obtenue via le mélange de LMM, le taux de bien classés est de 0.71. Pour ces deux approches de classification, nous remarquons également que le taux concordance de la classification a posteriori est de 0.72 entre ces deux mélanges.

TABLE 5.4: Choix du nombre de classes latentes K suivant les critères BIC et ICL et sachant que $K = 3$ dans les $N = 500$ jeux de données simulés.

K	Mélange CM		Mélange de LMM	
	BIC	ICL	BIC	ICL
1	5	5	0	0
2	0	0	1	2
3	494	494	163	238
4	1	1	249	188
5	0	0	84	69
6	0	0	3	3

5.3.3 Étude de la classification *a posteriori*

Nous nous intéressons maintenant à déterminer le profil des patientes des différentes classes latentes obtenues. Cela revient à expliquer les classifications obtenues pour les quatre dimensions de HRQoL (GHS, fatigue, insomnie, perte d'appétit) suivant les variables médicales, sociodémographiques et psychosociales. Sous cet objectif, nous considérons ces variables à l'inclusion. Les données issues du questionnaire MAC étant recueillies qu'à partir de la première visite (à 1 mois), nous prenons en compte seulement les données de cette visite. L'appartenance à une classe est expliquée par l'intermédiaire d'un modèle de régression logistique multinomiale. Cependant, le nombre de variables explicatives à disposition étant conséquent, nous procédons tout d'abord à l'étude de la corrélation entre ces dernières et les classifications obtenues par un test d'indépendance du χ^2 . Ces tests permettront de cibler les variables les plus pertinentes afin expliquer l'appartenance aux sous-populations obtenues et faire émerger des profils de patientes.

(In)dépendance entre les classifications obtenues et données recueillies

Les Tables 5.5 et 5.6 renvoient les effectifs de chaque variable suivant les classes latentes obtenues pour la dimension du GHS. Parmi les trois types de variables explicatives, la taille de la tumeur ($p = .05^*$), le niveau d'étude ($p < .001^{***}$) et le fait d'être optimiste ou pas ($p = .05^*$, LOT-R) sont significativement informatifs de la classification.

Concernant la dimension de fatigue, l'appartenance aux classes latentes est significativement dépendante de la taille de la tumeur ($p = .02^*$, Table 5.7) ainsi que de la stabilité émotionnelle de la patiente ($p = .06$, névrosisme mesuré par l'échelle NE-PI-R, Table 5.8).

Les classes latentes obtenues pour la dimension liée à la perte d'appétit est quant à elle significativement dépendante du type de chirurgie ($p = .08$, Table 5.9) ainsi que de l'état anxieux de la patiente ($p < .001^{***}$, Table 5.10).

Pour finir, les différentes évolutions de la mesure de l'insomnie au cours du temps sont significativement dépendantes du type de chirurgie ($p = .04^*$, Table 5.11), de

Chapitre 5. Étude longitudinale mono-dimensionnelle par un mélange de modèles mixtes

la situation familiale ($p = .02^*$, Table 5.11), ainsi que de l'état anxieux ($p = .05^*$, Table 5.12) ou encore le fait d'adopter la stratégie d'adaptation : préoccupation et anxiété ($p = .01^*$, Table 5.12).

TABLE 5.5: Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score du GHS. Deux classes latentes sont considérées : classe A pour Amélioration de la HRQoL et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.

	Classe D ($n = 60$)	Classe A ($n = 72$)	P-value
Variabiles médicales			
Taille Tumorale (mm)			.05*
< 15	20 (36)	37 (55)	
≥ 15	36 (64)	30 (45)	
Type de Chirurgie			.65
Tumorectomie	41 (36)	53 (75)	
Mastectomie	18 (31)	18 (25)	
Grade			.36
1	15 (29)	16 (25)	
2	20 (39)	34 (52)	
3	16 (32)	15 (23)	
Variabiles sociodémographiques			
Âge (années)			.25
< 50	20 (33)	15 (21)	
[50; 60[24 (40)	32 (44)	
≥ 60	16 (27)	25 (35)	
Situation professionnelle			.99
En activité	27 (45)	33 (46)	
Sans emploi	8 (13)	9 (13)	
Retraitée	25 (42)	29 (41)	
Nombre d'enfants			.23
0 ou 1	14 (23)	25 (35)	
2	26 (44)	31 (43)	
≥ 3	20 (33)	16 (22)	
Situation familiale			.58
Mariée	46 (77)	51 (71)	
Séparée/Seule	14 (23)	21 (29)	
Étude			<.001***
<Bac	19 (32)	48 (67)	
≥Bac	41 (68)	24 (33)	

* : $p\text{-value} < .05$; ** : $p\text{-value} < .01$; *** : $p\text{-value} < .001$

TABLE 5.6: Résultats des tests d'indépendance entre les variables psychosociales et la classification de la population d'intérêt suivant le score du GHS. Deux classes latentes sont considérées : classe A pour Amélioration de la HRQoL et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.

	Classe D (<i>n</i> = 60)	Classe A (<i>n</i> = 72)	P-value
Variabes psychosociales			
Névrosisme (Stable < 102)			.44
Stable	28 (47)	39 (55)	
Instable	32 (53)	32 (45)	
HAD Anxiété (Anxieuse ≥ 9)			.99
Non Anxieuse	34 (57)	42 (58)	
Anxieuse	26 (43)	30 (42)	
HAD Dépression (Dépressive ≥ 8)			.29
Non dépressive	56 (93)	62 (86)	
Dépressive	4 (7)	10 (14)	
LOT-R (Optimiste > 17)			.05*
Non	18 (30)	35 (49)	
Oui	42 (70)	37 (51)	
Stratégies d'ajustement à la maladie (MAC)			
Esprit combatif			.80
Non	18 (31)	25 (35)	
Oui	40 (69)	47 (65)	
Impuissance, Désespoir			1
Non	46 (79)	57 (79)	
Oui	12 (21)	15 (21)	
Préoccupation Anxiété			.73
Non	42 (72)	49 (68)	
Oui	16 (28)	23 (32)	
Dénégation			.34
Non	33 (57)	48 (67)	
Oui	25 (43)	24 (33)	
Évitement			.65
Non	29 (50)	40 (56)	
Oui	29 (50)	32 (44)	

* : *p*value < .05 ; ** : *p*value < .01 ; *** : *p*value < .001

Chapitre 5. Étude longitudinale mono-dimensionnelle par un mélange de modèles mixtes

TABLE 5.7: Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score de fatigue. Trois classes latentes sont considérées : classe A pour Amélioration de la HRQoL, classe C pour HRQoL Constante et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.

	Classe A (<i>n</i> = 12)	Classe C (<i>n</i> = 91)	Classe D (<i>n</i> = 29)	P-value
Variabiles médicales				
Taille Tumorale (mm)				.02*
< 15	4 (36)	46 (55)	7 (25)	
≥ 15	7 (64)	38 (45)	21 (75)	
Type de Chirurgie				.39
Tumorectomie	7 (58)	64 (72)	23 (79)	
Mastectomie	5 (42)	25 (28)	6 (21)	
Grade				.45
1	1 (9)	24 (30)	6 (24)	
2	7 (64)	37 (46)	10 (40)	
3	3 (27)	19 (24)	9 (36)	
Variabiles sociodémographiques				
Âge (années)				.95
< 50	4 (33)	23 (25)	8 (28)	
[50; 60[5 (42)	38 (42)	13 (44)	
≥ 60	3 (25)	30 (33)	8 (28)	
Situation professionnelle				.87
En activité	6 (49)	43 (47)	11 (38)	
Sans emploi	1 (08)	12 (13)	4 (14)	
Retraitée	4 (33)	36 (40)	14 (48)	
Nombre d'enfants				.53
0 ou 1	4 (33)	27 (29)	8 (27)	
2	4 (33)	43 (47)	10 (34)	
≥ 3	4 (33)	21 (23)	11 (38)	
Situation familiale				.83
Mariée	8 (67)	67 (74)	22 (76)	
Seule/Séparée	4 (33)	24 (27)	7 (24)	
Étude				.86
< Bac	6 (50)	45 (49)	16 (55)	
≥ Bac	6 (50)	46 (51)	13 (45)	

* : $p\text{-value} < .05$; ** : $p\text{-value} < .01$; *** : $p\text{-value} < .001$

TABLE 5.8: Résultats des tests d'indépendance entre les variables psychosociales et la classification de la population d'intérêt suivant le score de fatigue. Trois classes latentes sont considérées : classe A pour Amélioration de la HRQoL, classe C pour HRQoL Constante et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.

	Classe A (<i>n</i> = 12)	Classe C (<i>n</i> = 91)	Classe D (<i>n</i> = 29)	P-value
Variabiles psychosociales				
Névrosisme (Stable<102)				.06
Stable	9 (75)	40 (44)	18 (62)	
Instable	3 (25)	50 (55)	11 (38)	
HAD Anxiété (Anxieuse≥ 9)				.28
Non Anxieuse	7 (58)	56 (62)	13 (45)	
Anxieuse	5 (42)	35 (38)	16 (55)	
HAD Dépression (Dépressive≥ 8)				.71
Non dépressive	11 (92)	80 (88)	27 (93)	
Dépressive	1 (8)	11 (12)	2 (7)	
LOT-R (Optimiste>17)				.88
Non	4 (33)	37 (41)	12 (41)	
Oui	8 (67)	54 (59)	17 (59)	
Stratégies d'ajustement à la maladie (MAC)				
Esprit combatif				.60
Non	5 (42)	27 (30)	11 (38)	
Oui	7 (58)	62 (70)	18 (62)	
Impuissance, Désespoir				.58
Non	10 (83)	72 (81)	21 (72)	
Oui	2 (17)	17 (19)	8 (28)	
Préoccupation Anxiété				.19
Non	6 (50)	66 (74)	19 (66)	
Oui	6 (50)	23 (26)	10 (34)	
Dénégation				.23
Non	10 (83)	52 (58)	19 (66)	
Oui	2 (17)	37 (42)	10 (34)	
Évitement				.81
Non	7 (58)	48 (54)	14 (48)	
Oui	5 (42)	41 (46)	15 (52)	

* : *p*value<.05; ** : *p*value<.01; *** : *p*value<.001

Chapitre 5. Étude longitudinale mono-dimensionnelle par un mélange de modèles mixtes

TABLE 5.9: Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score lié à la perte d'appétit. Deux classes latentes sont considérées : classe A pour Amélioration de la HRQoL et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.

	Classe D (n = 50)	Classe A (n = 82)	P-value
Variabiles médicales			
Taille Tumorale (mm)			.41
< 15	20 (41)	37 (50)	
≥ 15	29 (59)	37 (50)	
Type de Chirurgie			.08
Tumorectomie	41 (82)	53 (66)	
Mastectomie	9 (18)	27 (34)	
Grade			.25
1	13 (28)	18 (26)	
2	18 (38)	36 (52)	
3	16 (34)	15 (22)	
Variabiles sociodémographiques			
Âge (années)			.74
< 50	15 (30)	20 (24)	
[50; 60[21 (42)	35 (43)	
≥ 60	14 (28)	27 (33)	
Situation professionnelle			.66
En activité	21 (42)	39 (43)	
Sans emploi	8 (16)	9 (16)	
Retraitée	34 (40)	34 (41)	
Nombre d'enfants			.78
0 ou 1	17 (34)	22 (27)	
2	19 (38)	38 (46)	
≥ 3	14 (28)	14 (27)	
Situation familiale			.42
Mariée	36 (72)	67 (83)	
Seule/Séparée	14 (28)	15 (18)	
Étude			.97
< Bac	26 (52)	41 (50)	
≥ Bac	24 (48)	41 (50)	

* : pvalue<.05 ; ** : pvalue<.01 ; *** : pvalue<.001

TABLE 5.10: Résultats des tests d'indépendance entre les variables psychosociales et la classification de la population d'intérêt suivant le score lié à la perte d'appétit. Deux classes latentes sont considérées : classe A pour Amélioration de la HRQoL et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.

	Classe D (<i>n</i> = 50)	Classe A (<i>n</i> = 82)	P-value
Variabiles psychosociales			
Névrosisme (Stable<102)			.29
Stable	29 (58)	38 (46)	
Instable	21 (42)	43 (52)	
HAD Anxiété (Anxieuse≥ 9)			<.001***
Non Anxieuse	19 (38)	57 (70)	
Anxieuse	31 (62)	25 (30)	
HAD Dépression (Dépressive≥ 8)			.48
Non dépressive	43 (86)	75 (91)	
Dépressive	7 (14)	7 (9)	
LOT-R (Optimiste>17)			.99
Non	20 (40)	33 (40)	
Oui	30 (60)	49 (60)	
Stratégies d'ajustement à la maladie (MAC)			
Esprit combatif			.13
Non	21 (28)	22 (42)	
Oui	29 (72)	58 (58)	
Impuissance, Désespoir			.35
Non	37 (82)	66 (74)	
Oui	13 (18)	14 (26)	
Préoccupation Anxiété			.17
Non	31 (62)	60 (75)	
Oui	19 (38)	20 (25)	
Dénégation			.90
Non	32 (61)	49 (64)	
Oui	18 (39)	31 (36)	
Évitement			.99
Non	26 (52)	43 (54)	
Oui	24 (48)	37 (46)	

* : pvalue<.05; ** : pvalue<.01; *** : pvalue<.001

Chapitre 5. Étude longitudinale mono-dimensionnelle par un mélange de modèles mixtes

TABLE 5.11: Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score d'insomnie. Trois classes latentes sont considérées : classe A pour Amélioration de la HRQoL, classe C pour HRQoL Constante et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.

	Classe C (<i>n</i> = 41)	Classe D (<i>n</i> = 57)	Classe A (<i>n</i> = 34)	P-value
Variabiles médicales				
Taille Tumorale (mm)				.76
< 15	18 (51)	24 (44)	15 (45)	
≥ 15	17 (49)	31 (56)	18 (55)	
Type de Chirurgie				.04*
Tumorectomie	33 (82)	34 (61)	27 (79)	
Mastectomie	7 (18)	22 (39)	7 (21)	
Grade				.97
1	8 (26)	15 (28)	8 (15)	
2	16 (52)	24 (44)	14 (26)	
3	7 (23)	15 (28)	9 (17)	
Variabiles sociodémographiques				
Âge (années)				.67
< 50	8 (20)	18 (32)	9 (26)	
[50; 60[18 (44)	22 (38)	16 (47)	
≥ 60	15 (37)	17 (30)	9 (26)	
Situation professionnelle				.74
En activité	16 (39)	29 (52)	15 (44)	
Sans emploi	7 (17)	6 (11)	4 (12)	
Retraitée	18 (45)	21 (38)	15 (44)	
Nombre d'enfants				.36
0 ou 1	8 (20)	19 (34)	12 (36)	
2	21 (51)	25 (44)	11 (32)	
≥ 3	12 (29)	13 (23)	11 (32)	
Situation familiale				.02*
Mariée	27 (66)	49 (86)	21 (62)	
Seule/Séparée	14 (34)	8 (14)	13 (38)	
Étude				.58
< Bac	22 (65)	36 (76)	19 (56)	
≥ Bac	19 (55)	31 (91)	15 (44)	

* : $p\text{-value} < .05$; ** : $p\text{-value} < .01$; *** : $p\text{-value} < .001$

TABLE 5.12: Résultats des tests d'indépendance entre les variables explicatives médicales puis sociodémographiques, et la classification de la population d'intérêt suivant le score d'insomnie. Trois classes latentes sont considérées : classe A pour Amélioration de la HRQoL, classe C pour HRQoL Constante et classe D pour Détérioration. Pour chaque variable, les effectifs par modalité sont donnés ainsi que les pourcentages.

	Classe C (n = 41)	Classe D (n = 57)	Classe A (n = 34)	P-value
Variabiles psychosociales				
Névrosisme (Stable<102)				.53
Stable	18 (44)	31 (54)	18 (55)	
Instable	23 (56)	26 (46)	15 (45)	
HAD Anxiété (Anxieuse≥ 9)				.05*
Non Anxieuse	30 (73)	28 (49)	28 (53)	
Anxieuse	11 (27)	29 (51)	29 (47)	
HAD Dépression (Dépressive≥ 8)				.90
Non dépressive	36 (88)	51 (89)	31 (91)	
Dépressive	5 (12)	6 (11)	3 (9)	
LOT-R (Optimiste>17)				.52
Non	14 (34)	23 (40)	16 (47)	
Oui	27 (66)	34 (60)	18 (53)	
Stratégies d'ajustement à la maladie (MAC)				
Esprit combatif				.17
Non	12 (29)	23 (42)	8 (24)	
Oui	29 (71)	32 (58)	26(76)	
Impuissance, Désespoir				.28
Non	35 (85)	40 (73)	28 (82)	
Oui	6 (15)	15 (27)	6 (18)	
Préoccupation Anxiété				.01*
Non	32 (78)	31 (56)	28 (82)	
Oui	9 (22)	24 (44)	6 (18)	
Dénégation				.22
Non	30 (73)	32 (58)	19 (56)	
Oui	11 (27)	23 (42)	15 (44)	
Évitement				.74
Non	21 (51)	28 (51)	20 (59)	
Oui	20 (49)	27 (49)	14 (41)	

* : pvalue<.05; ** : pvalue<.01; *** : pvalue<.001

Explication de la classification

Nous avons identifié les variables qui sont significativement corrélées aux différentes classifications obtenues. Une régression logistique multinomiale multivariée nous permet de définir des profils de patientes pour chaque trajectoire des dimensions considérées.

La Table 5.13 montre l'influence des variables explicatives sur la classification obtenue concernant le GHS. Les scores du GHS permettent de distinguer deux sous-populations homogènes ayant des trajectoires opposées (Figure 5.1b). Il est bien plus probable que les patientes ayant un niveau d'étude supérieur ou égal au Baccalauréat appartiennent à la classe D qu'à la classe A ($OR = 4.30$ et $IC_{95\%} [1.95; 9.51]$), soit d'avoir un meilleur niveau de HRQoL au départ mais qui se détériore au cours du temps jusqu'à un niveau relativement faible. Au contraire, avoir un niveau inférieur au Baccalauréat suggérerait d'avoir un faible niveau de HRQoL au départ, mais qui s'améliore au cours du temps. L'optimisme des patientes aide également à expliquer la probabilité d'appartenir à la classe D plutôt qu'à la classe A. En effet, il y a plus de chance d'appartenir à la classe D plutôt qu'à la classe A si la patiente est optimiste ($OR = 2.34$ et $IC_{95\%} [1.04; 5.27]$). Ce résultat semble contradictoire mais peut s'expliquer par un meilleur niveau de HRQoL à la chirurgie pour la classe D par rapport à celle de la classe A. Pour finir, la taille tumorale semble moins impacter l'appartenance aux classes.

TABLE 5.13: Résultats associés au modèle multinomiale logistique concernant le statut global de santé (GHS). OR correspond à l'Odd Ratio, $IC_{95\%}$ son intervalle de confiance et p . la pvalue associée au test de rapport de vraisemblance relatif au paramètre testé.

Variables Explicatives	Classe D ($n = 60$)		Classe A ($n = 72$)		p .
	OR	$IC_{95\%}$	OR		
Niveau d'étude					<.001***
<Bac	1	-	1		
≥Bac	4.30	[1.95; 9.51]	1		
LOT-R (optimisme >17)					.04*
Non	1	-	1		
Oui	2.34	[1.04; 5.27]	1		
Taille tumorale (mm)					.05
<15	1	-	1		
≥15	2.19	[0.98; 4.87]	1		

* : pvalue<.05 ; ** : pvalue<.01 ; *** : pvalue<.001

La Table 5.14 montre l'influence des deux variables explicatives névrosisme et taille tumorale, sur la classification de la population d'intérêt par rapport aux scores de fatigue. Si la patiente est instable émotionnellement, il est plus probable qu'elle ait une évolution constante de la HRQoL (classe C) plutôt qu'une détérioration (classe D) ou une amélioration (classe A) de la HRQoL. Cependant, ce résultat est à la limite de la significativité. En revanche, la taille tumorale est statistiquement significative ($p = .008$) pour expliquer l'appartenance des classes. En effet, il est bien plus probable ($OR=4.07$; $IC_{95\%} = [1.56; 10.86]$) qu'une patiente ait une HRQoL qui se détériore au cours du temps plutôt qu'elle reste constante si la taille de sa tumeur excède 15 millimètres. La taille tumorale semble jouer sur l'évolution de la fatigue.

Si la tumeur est importante, cela suggère que la HRQoL est plus sujette à évoluer au cours du temps.

TABLE 5.14: Résultats associés au modèle de régression logistique multinomiale concernant la fatigue. OR correspond à l'Odd Ratio, $IC_{95\%}$ son intervalle de confiance et p la pvalue associée au test de rapport de vraisemblance relatif au paramètre testé.

Variables Explicatives	Classe A ($n = 12$)		Classe D ($n = 29$)		Classe C ($n = 91$)		p .
	OR	$IC_{95\%}$	OR	$IC_{95\%}$	OR		
Névrosisme (stable<102)							.05*
Stable	1		1		1		
Instable	0.26	[0.06 ; 1.07]	0.43	[0.17 ; 1.08]	1		
Taille tumorale (mm)							.008**
<15	1				1		
≥15	2.49	[0.66 ; 9.43]	4.07	[1.53 ; 10.86]	1		

* : $pvalue < .05$; ** : $pvalue < .01$; *** : $pvalue < .001$

La Table 5.15 montre que la classification pour l'insomnie est la plus riche en information. En effet, de nombreuses variables explicatives permettent d'expliquer l'appartenance aux différentes classes. Le type de chirurgie est à la limite de la significativité : il est plus probable ($OR=2.91$, $IC_{95\%} = [1.03; 8.26]$) que les patientes ayant subi une mastectomie aient une détérioration de leur HRQoL (augmentation de l'insomnie) au cours du temps plutôt qu'une évolution constante. De plus, il est beaucoup moins probable ($OR=0.24$; $IC_{95\%} = [0.08; 0.74]$) d'avoir une détérioration de la HRQoL plutôt qu'une évolution constante si la patiente n'est pas en mariée. Les deux dernières variables explicatives concernent l'anxiété. En effet, l'anxiété influence l'évolution de l'insomnie chez la patiente où il est plus probable que cette dernière évolue (amélioration et détérioration) si la patiente est anxieuse. La classe latente liée à un niveau constant de la HRQoL renvoie un niveau d'insomnie en moyenne plus bas pour les deux autres classe latentes. La stratégie d'adaptation à la maladie du type préoccupation anxieuse confirme ces résultats. Notamment, il est plus probable ($OR=2.66$; $IC_{95\%} = [0.99; 7.11]$) d'appartenir à la classe D plutôt qu'à la classe C si la patiente adopte cette stratégie.

Concernant la dimension liée à la perte d'appétit, la Table 5.16 confirme l'influence de l'anxiété et du type de chirurgie sur la classification obtenue. Il est beaucoup plus probable ($OR=4.29$; $IC_{95\%} = [1.97; 9.31]$) que la perte d'appétit augmente au cours du temps si la patiente est anxieuse. Le résultat sur le type de chirurgie est contradictoire avec celui de l'insomnie. En effet, il y a moins de chance que la patiente appartienne à la classe A plutôt qu'à la classe D si elle a eu une mastectomie plutôt qu'une tumorectomie ($OR=0.35$; $IC_{95\%} = [0.14; 0.88]$).

5.4 Discussion

L'analyse des données de l'étude « Moral » a permis de mettre évidence pour chaque dimension étudiée, des profils de patientes associés aux trajectoires latentes. Les résultats les plus intuitifs concernent les dimensions symptomatiques. En effet, l'état anxieux de la patiente est un indicateur de l'évolution des symptômes insomnie

Chapitre 5. Étude longitudinale mono-dimensionnelle par un mélange de modèles mixtes

TABLE 5.15: Résultats associés au modèle de régression logistique multinomiale concernant l'insomnie. OR correspond à l'Odd Ratio, $IC_{95\%}$ son intervalle de confiance et p . la p value associée au test de rapport de vraisemblance relatif au paramètre testé.

Variables Explicatives	Classe D (n=57)		Classe A (n=34)		Classe C (n=41)	
	OR	$IC_{95\%}$	OR	$IC_{95\%}$	OR	p .
Type de Chirurgie						.06
Tumorectomie	1		1		1	
Mastectomie	2.91	[1.03; 8.26]	1.11	[0.34; 3.64]	1	
Situation familiale						.01*
Mariée	1		1		1	
Seule/Séparée	0.24	[0.08; 0.74]	1.03	[0.39; 2.75]	1	
Anxiété (HAD)						.02*
Non anxieuse	1		1		1	
Anxieuse	3.44	[1.32; 8.98]	2.71	[0.99; 7.36]	1	
MAC : Préoccupation - Anxiété						.02*
Non	1		1		1	
Oui	2.66	[0.99; 7.11]	0.67	[0.21; 2.17]	1	

* : p value<.05; ** : p value<.01; *** : p value<.001

TABLE 5.16: Résultats associés au modèle de régression logistique multinomiale concernant la perte d'appétit. OR correspond à l'Odd Ratio, $IC_{95\%}$ son intervalle de confiance et p . la P value associée au test de rapport de vraisemblance relatif à la variable explicative.

Variables Explicatives	Classe D ($n = 50$)		Classe A ($n = 82$)		p .
	OR	$IC_{95\%}$	OR	$IC_{95\%}$	
Type de Chirurgie					<.001***
Tumorectomie	1	-	1		
Mastectomie	0.35	[0.14; 0.88]	1		
Anxiété (HAD)					.02*
Non anxieuse	1	-	1		
Anxieuse	4.29	[1.97; 9.31]	1		

* : p value<.05; ** : p value<.01; *** : p value<.001

et perte d'appétit. Il est significativement plus probable qu'une patiente anxieuse ait une détérioration de sa HRQoL au cours du temps concernant ces deux dimensions. L'anxiété et la dépression semblent également être deux facteurs influençant l'évolution de HRQoL. Les résultats présentés en section 5.5.2 (Table 5.18) montrent que les classifications obtenues pour les dimensions physique, émotionnelle, personnelle et pour le score global de Nordin sont également influencées. Une prise en charge adaptée des patientes présentant un état anxieux et/ou dépressif pourrait alors améliorer leur HRQoL. En revanche, l'optimisme est une variable de personnalité qui ressort peu, seule la classification du GHS est corrélée avec cette variable. Le résultat concernant l'influence du niveau d'étude pour expliquer la classification obtenue pour le GHS et la dimension physique est difficilement interprétable. En effet, il est plus probable pour ces deux dimensions que la HRQoL de la patiente se détériore au cours du temps si elle a un niveau d'étude supérieur au Baccalauréat. Ce résultat peut sembler contradictoire étant donné que les personnes les plus diplômées sont

généralement celles qui sont diagnostiquées les plus précocement.

Bien que les classifications obtenues ne sont pas expliquées suivant les mêmes covariables, il y a quand même une forte similitude entre les trajectoires de chaque dimension de HRQoL. La première est l'évolution constante de HRQoL pour la classe la plus représentée. La trajectoire caractérisée par un bon niveau de HRQoL, et constante au cours du temps, regroupe généralement la moitié de l'effectif total. Cela peut suggérer que c'est la réponse attendue³ du questionnaire qui est cochée la plupart du temps. Un autre phénomène récurrent sur les trajectoires latentes estimées est leur stabilisation à partir de 4 ou 5 mois. Plusieurs hypothèses peuvent être formulées sur ce phénomène. La première est d'ordre psychologique. En effet, un phénomène d'habituation (biais de mémoire) peut s'installer lorsque la patiente remplit de manière périodique le même questionnaire. La seconde suggère que les changements du niveau de HRQoL interviennent essentiellement dans les quelques mois (4 à 6 mois) suivant la chirurgie.

La classification des patients est réalisée par des mélanges de modèles mixtes. Dans cette approche, le critère d'homogénéité que nous avons utilisé est la trajectoire de HRQoL des patients. L'utilisation de mélange de modèles mixtes a permis d'obtenir les sous-populations homogènes ainsi que les trajectoires latentes qui les caractérisent. Une fois les classes latentes déterminées, l'appartenance à ces dernières est expliquée par un modèle de régression logistique (multinomiale) suivant les variables explicatives disponibles à la chirurgie.

Concernant la classification, le choix du modèle mixte pour le mélange a une incidence sur la classification obtenue. Nous en avons eu un aperçu pour une dimension à un item. En effet, l'utilisation d'un mélange de LMM a pour conséquence de favoriser un nombre K plus important que pour un mélange de modèles cumulatifs pour données ordinales (*cumulatif, gaussien, proportionnel*). Cependant, pour un même nombre de sous-populations homogènes considérées, l'information apportée par les deux modélisations concernant l'évolution de la HRQoL reste similaire. La concordance des sous-populations obtenues est également proche. Or, comme le montrent les simulations de la Table 5.4, ces deux dernières comparaisons ne semblent pas pertinentes en pratique étant donné que le choix de K diffère entre les deux approches. Il serait très intéressant de compléter cette comparaison pour des dimensions regroupant plus d'un item afin d'évaluer les discordances des approches. Celles-ci concerneraient le choix de K , la classification de la population d'intérêt ou encore l'information véhiculée par les trajectoires latentes. Dans l'idéal, il serait plus qu'intéressant d'étendre la comparaison aux mélanges de modèles IRT pour les données de HRQoL, et ainsi identifier l'impact d'analyser directement les données brutes plutôt que les scores.

3. au sens d'une réponse normale, classique ;

5.5 Compléments

5.5.1 Estimation par l'algorithme EM pour un mélange de LMM

L'algorithme *EM* ("*Espérance-Maximisation*") est un algorithme itératif construit sur deux étapes successives qui ont donné lieu à son appellation (Dempster et al., 1977). La première, l'étape "E" pour *Expectation*, revient à considérer l'espérance de la vraisemblance des données complètes sachant les données observées (y) et la valeur courante des paramètres, noté $\Upsilon^{[s]}$ à l'itération s , tel que :

$$\mathcal{Q}(\Upsilon, \Upsilon^{[s]}) = \mathbb{E} \left[\mathcal{L}(\Upsilon; y, h) | y, \Upsilon^{[s]} \right].$$

Les statistiques exhaustives sont alors remplacées par leur espérance conditionnelle. Prendre l'espérance revient à intégrer sur les données inconnues (h) afin de les faire disparaître du problème de résolution.

La seconde, l'étape "M" pour *Maximization*, revient en l'estimation des paramètres par la méthode du maximum de vraisemblance. Elle consiste à maximiser l'espérance conditionnelle obtenue à l'étape précédente en Υ , permettant la mise à jour des paramètres pour l'itération suivante tel que :

$$\Upsilon^{[s+1]} = \underset{\Upsilon}{\operatorname{argmax}} \mathcal{Q}(\Upsilon, \Upsilon^{[s]}). \quad (5.15)$$

Cet algorithme est computationnellement efficace. Une des particularités de ce dernier est que la vraisemblance croît d'une itération à l'autre. Les détails de cet algorithme concernant les modèles linéaires mixtes et le mélange de LMM sont présentés dans les points suivants. Ils sont inspirés des travaux de thèse de Catherine Trottier, Olivier Martin et Marie-José Martinez (Trottier, 1998; Martin, 2002; Martinez, 2006).

Estimation de LMM

Le calcul de l'espérance conditionnelle à la première étape s'effectue dans la majorité des cas avec des approximations par méthodes numériques. Dans le cas des modèles linéaires mixtes, les lois conditionnelles sont explicites. Par conséquent, nous pouvons remplacer directement les statistiques exhaustives par leur espérance conditionnelle pour l'étape de maximisation. Par souci de notation, nous nous plaçons dans le cas classique du LMM tel que :

$$Y = X\beta + U\xi + \varepsilon \quad (5.16)$$

avec :

- Y le vecteur aléatoire de taille $N = \sum_{i=1}^n n_i$ à expliquer
- ξ le vecteur d'effet aléatoire de taille q . Il se décompose en R parties (nombre d'effet aléatoire considéré) tel que $\xi = (\xi'_1, \dots, \xi'_R)'$. $\xi_r \sim \mathcal{N}_{q_r} (0, \sigma_r^2 A_r)$ vecteur modélisant un effet aléatoire de longueur q_r où $\sum_{r=1}^R q_r = q$ et A_r matrice carrée supposée connue. $\xi \sim \mathcal{N}_q \left(0, D = \operatorname{diag} \{ \sigma_r^2 A_r \}_{r=1, \dots, R} \right)$

- $U = \begin{bmatrix} U_1 \\ \vdots \\ U_R \end{bmatrix}$
- $\varepsilon \sim \mathcal{N}_n(0, \tau^2 V)$
- $\mathbb{E}(Y|\xi) = X\beta + U\xi$;
- $\text{Var}(Y|\xi) = \tau^2 V$;
- $\mathbb{E}(Y) = X\beta$;
- $\text{Var}(Y) = \tau^2 V + \sum_{r=1}^R \sigma_r U_r A_r U_r' = \Psi$.

Remarque : dans le cas des données longitudinales de HRQoL, si on considère l'effet aléatoire sur l'individu lorsque l'on a T mesures répétées avec un intercept aléatoire et une pente aléatoire sur la variable temps, nous avons $q_j = n$ et $q = 2n$. Si on considère un effet aléatoire individu à chaque temps, $q = Tn$.

Avec l'approche par l'algorithme EM, y est le vecteur observé et ξ celui des données considérées comme manquantes. La distribution des données complètes $x = [y, \xi_1, \dots, \xi_R]'$ est :

$$x = \begin{bmatrix} y \\ \xi \end{bmatrix} \sim \mathcal{N} \left(\mu = \begin{bmatrix} X\beta \\ \mathbf{0} \end{bmatrix}, \Sigma = \begin{bmatrix} \Psi & \{j \sigma_r^2 U_r A_r\}_{r=1, \dots, R} \\ \{c \sigma_r^2 A_r U_r'\}_{r=1, \dots, R} & D \end{bmatrix} \right)$$

On note $\{c m_r\}$ la matrice M obtenue en superposant en colonne les éléments m_r et $\{j m_r\}$ la matrice M obtenue en juxtaposant en ligne les éléments m_r . La log-vraisemblance complétée s'écrit alors :

$$\ell(\beta, \sigma, \tau; x) = -\frac{\ln(2\pi)}{2} \sum_{r=0}^R q_r - \frac{1}{2} \sum_{r=0}^R (q_r \ln(\sigma_r^2) + \ln|A_r|) - \frac{1}{2} \sum_{r=0}^R \frac{\xi_r' A_r \xi_r}{\sigma_r^2}$$

avec $q_0 = n$, $\xi_0 = \varepsilon = y - X\beta - \sum_{r=1}^R U_r \xi_r$, $\tau^2 = \sigma_0^2$ et $A_0 = V$. Dans le cas où tout est supposé connu, les dérivées suivant les paramètres à estimer sont :

$$\frac{\partial \ell}{\partial \beta} = X' V^{-1} (y - X\beta - U\xi), \quad (5.17)$$

$$\frac{\partial \ell}{\partial \sigma_r^2} = \frac{-q_r}{2\sigma_r^2} + \frac{\xi_r' A_r^{-1} \xi_r}{2\sigma_r^4}. \quad (5.18)$$

Par annulation de ces dérivées, cela implique les estimateurs du maximum de vraisemblance complète suivants :

$$\widehat{\beta} = [X' V^{-1} X]^{-1} X' V^{-1} (y - U\xi), \quad (5.19)$$

$$\widehat{\sigma^2} = \frac{\xi_r' A_r^{-1} \xi_r}{q_r}. \quad (5.20)$$

ξ n'étant pas connu, il est impossible d'utiliser les statistiques exhaustives $\xi_r' A_r^{-1} \xi_r$ et $y - U\xi$. L'idée avec l'algorithme EM est de la remplacer par l'espérance conditionnelle sachant les données observées. Dans ce cas, cette distribution conditionnelle $\xi|y \sim \mathcal{N}(m, \Sigma_{\xi|y})$ est explicite avec :

$$\begin{aligned} m &= \mathbf{0} + (\sigma_r^2 A_r U_r')_{r=1, \dots, R} \Psi^{-1} (y - X\beta) \\ \Sigma_{\xi|y} &= \text{diag} \{ \sigma_r^2 A_r \} - [\sigma_r^2 A_r U_r'] \Psi^{-1} [\sigma_r^2 U_r A_r] \end{aligned}$$

Cela implique dans l'estimation des paramètres le remplacement des statistiques exhaustives $(y - U\xi)$ et $(\xi_r' A_r^{-1} \xi_r)$ respectivement par :

$$\begin{aligned}
 \mathbb{E}(Y - U\xi|y) &= Y - U\mathbb{E}(\xi|y) \\
 &= y - U \sum_{r=1}^R \sigma_r^2 A_j U_j' \Psi^{-1} (y - X\beta) \\
 &= y - \sum_{r=1}^R \sigma_r^2 U_j A_j U_j' \Psi^{-1} (y - X\beta) \\
 &= y - (\Psi - \tau^2 V) \Psi^{-1} (y - X\beta) \\
 &= y - (\Psi \Psi^{-1} (y - X\beta) - \tau^2 V \Psi^{-1} (y - X\beta)) \\
 &= X\beta - \tau^2 V \Psi^{-1} (y - X\beta) \tag{5.21}
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{E}(\xi_r' A_r^{-1} \xi_r | y) &= \mathbb{E}(\xi_r | y)^2 + \text{Var}(\xi_r | y) \\
 &= \sigma_r^4 (y - X\beta)' \Psi^{-1} U_r A_r U_r' \Psi^{-1} (y - X\beta) \\
 &\quad + \text{tr}(\sigma_r^2 Id_{q_r} - \sigma_r^4 U_r' \Psi^{-1} U_r A_r). \tag{5.22}
 \end{aligned}$$

avec $r = 1, \dots, R$ et d'après la note 1 qui suit.

Note 1. Soit X un vecteur aléatoire de taille n d'espérance $\mathbb{E}(X) = \mu$ et de variance $\text{Var}(X) = \Sigma$, et M une matrice symétrique de taille n . L'espérance de la forme quadratique est alors donnée par :

$$\begin{aligned}
 \mathbb{E}(X' M X) &= \mathbb{E} \left(\sum_{i=1}^n \sum_{j=1}^n m_{ij} X_i X_j \right) \\
 &= \sum_{i=1}^n \sum_{j=1}^n m_{ij} \mathbb{E}(X_i X_j) \\
 &= \sum_{i=1}^n \sum_{j=1}^n m_{ij} (\sigma_{ij} + \mu_i \mu_j) \\
 &= \sum_{i=1}^n \sum_{j=1}^n m_{ij} \sigma_{ji} + \sum_{i=1}^n \sum_{j=1}^n m_{ij} \mu_i \mu_j \\
 &= \sum_{i=1}^n [M \Sigma]_{ii} + \mu' M \mu \\
 &= \text{tr}[M \Sigma] + \mu' M \mu
 \end{aligned}$$

L'algorithme EM est un outil performant pour estimer des LMM où la couche cachée est représentée par les effets aléatoires. Pour le mélange de LMM, une couche cachée supplémentaire est prise en compte, celle de l'appartenance aux classes.

Estimation d'un mélange de LMM

Dans un L3M, une donnée inconnue supplémentaire est considérée : l'appartenance à une classe du mélange. Soit $x = (y, \xi, z)$ les données complètes de densité :

$$\begin{aligned}
 f(x; \lambda, p) &= \prod_{i=1}^n f(y_i, \xi_i, z_i; \lambda, p) \\
 &= \prod_{i=1}^n \prod_{k=1}^K [p_k f_k(y_i, \xi_i | \lambda_k)]^{z_{ik}} \\
 &= \prod_{i=1}^n \prod_{k=1}^K [p_k f_k(y_i | \xi_i; \lambda_k) f_k(\xi_i; \lambda_k)]^{z_{ik}} \quad (5.23)
 \end{aligned}$$

De (5.23) découle la log-vraisemblance suivante :

$$\ell(\lambda, p; x) = \sum_{i=1}^n \sum_{k=1}^K z_{ik} \ln(p_k) + \sum_{i=1}^n \sum_{k=1}^K z_{ik} \ln[f_k(y_i | \xi_i)] + \sum_{i=1}^n \sum_{k=1}^K z_{ik} \ln[f_k(\xi_i)]$$

où lorsque un seul effet aléatoire individu est considéré :

$$\begin{aligned}
 \ln[f_k(y_i | \xi_i; \lambda_k)] &= -\frac{1}{2} \left[n_i \ln(2\pi) + n_i \ln(\tau^2) + \frac{(y_i - X_i \beta_k - u_i \xi_i)' (y_i - X_i \beta_k - u_i \xi_i)}{\tau^2} \right] \\
 \ln[f_k(\xi_i; \lambda_k)] &= -\frac{1}{2} \left[\ln(2\pi) + \ln(\sigma_k^2) + \frac{\xi_i^2}{\sigma_k^2} \right]
 \end{aligned}$$

La première étape (E) consiste à prendre l'espérance de $\ell(\lambda, p; x)$ par rapport à la distribution conditionnelle des données cachées $h = (\xi, z)$ sachant les observations y et la valeur courante des paramètres $(\lambda^{[s]}, p^{[s]})$. Soit,

$$\begin{aligned}
 \mathcal{Q}(\lambda, p | \lambda^{[s]}, p^{[s]}) &= \mathbb{E} \left[\ell(\lambda; y, h) | y, \lambda^{[s]}, p^{[s]} \right] \\
 &= \sum_{i=1}^n \sum_{k=1}^K \mathbb{E}_y^h(z_{ik}) \ln(p_k) + \sum_{i=1}^n \sum_{k=1}^K \mathbb{E}_y^h(z_{ik} \ln[f_k(y_i | \xi_i; \lambda_k)]) \\
 &\quad + \sum_{i=1}^n \sum_{k=1}^K \mathbb{E}_y^h(z_{ik} \ln[f_k(\xi_i; \lambda_k)]) \\
 &= \sum_{i=1}^n \sum_{k=1}^K \mathbb{E}_y^h(z_{ik}) \ln(p_k) - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^K \mathbb{E}_y^h(z_{ik}) \left[(n_i + 1) \ln(2\pi) + n_i \ln(\tau^2) + \ln(\sigma_k^2) \right] \\
 &\quad - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^K \mathbb{E}_y^h(z_{ik}) \left[\frac{\mathbb{E}_y^h(\varepsilon_i' \varepsilon_i)}{\tau^2} + \frac{\mathbb{E}_y^h(\xi_i^2)}{\sigma_k^2} \right] \quad (5.24)
 \end{aligned}$$

où $\varepsilon_i = (y_i - X_i \beta_k + U_i \xi_i)$ et $\mathbb{E}_y^h(\cdot)$ est l'espérance conditionnelle des effets aléatoires et des appartenances aux classes sachant les données observées y et la valeur courante des paramètres $(\lambda^{[s]}, p^{[s]})$. Cela engendre que :

- $\mathbb{E}_y^h(z_{ik}) = \Pr(z_{ik} = 1 | y_i, \lambda_k^{[s]}, p^{[s]})$ correspond à la probabilité d'appartenir à la classe \mathcal{C}_k pour l'individu i , sachant les données observées et les valeurs courantes des paramètres. Cette probabilité est modélisée de la manière suivante :

$$\mathbb{E}_y^h(z_{ik}) = \pi_{ik}^{[s]} = \frac{p_k^{[s]} f_k(y_i; \lambda_k^{[s]})}{\sum_{l=1}^K p_l^{[s]} f_l(y_i; \lambda_l^{[s]})} \quad (5.25)$$

- $\mathbb{E}_y^h(\xi_i) = \mathbb{E}(\xi_i | z_{ik}, y_i, \lambda_k^{[s]}) = \sigma_k^{2[s]} U_i' (\Gamma_{k,i}^{[s]})^{-1} (y_i - X_i \beta_k^{[s]})$
où $\Gamma_{k,i}^{[s]} = \tau^{2[s]} Id_{n_i} + \sigma_k^{2[s]} U_i U_i'$
- $\text{Var}_y^h(\xi_i) = \sigma_k^{2[s]} - \sigma_k^{4[s]} U_i' (\Gamma_{k,i}^{[s]})^{-1} U_i$
- $\mathbb{E}_y^h(\xi_i^2) = \mathbb{E}_y^h(\xi_i)' \mathbb{E}_y^h(\xi_i) + \text{tr}(\text{Var}_y^h(\xi_i))$
 $= \sigma_k^{4[s]} (y_i - X_i \beta_k^{[s]})' (\Gamma_{k,i}^{[s]})^{-1} U_i U_i' (\Gamma_{k,i}^{[s]})^{-1} (y_i - X_i \beta_k^{[s]}) + \sigma_k^{2[s]} - \sigma_k^{4[s]} \text{tr} \left((\Gamma_{k,i}^{[s]})^{-1} U_i U_i' \right)$
- $\mathbb{E}_y^h(\varepsilon_i' \varepsilon_i) = [y_i - X_i \beta_k - U_i \mathbb{E}_y^h(\xi_i)]' [y_i - X_i \beta_k - U_i \mathbb{E}_y^h(\xi_i)] + \text{tr}[U_i \text{Var}_y^h(\xi_i) U_i']$
 $= \tau^{4[s]} (y_i - X_i \beta_k^{[t]})' [\Gamma_{k,i}^{[s]}]^{-1} [\Gamma_{k,i}^{[s]}]^{-1} (y_i - X_i \beta_k^{[t]}) - n_i \tau^{2[s]} - \tau^{4[s]} \text{tr} \left([\Gamma_{k,i}^{[s]}]^{-1} \right)$

La seconde étape (M) consiste à maximiser l'expression (5.24) en (λ, p) . Pour cela, nous obtenons les estimations du maximum de vraisemblance suivant à partir de la dérivé de (5.24). Concernant le paramètre p , la contrainte $\sum_{k=1}^K p_k = 1$ est à prendre en compte. Pour cela, nous considérons le lagrangien suivant $\mathcal{L}a(p, \alpha) = \mathcal{Q}(\lambda, p | \lambda^{[s]}, p^{[s]}) - \alpha \left(\sum_{k=1}^K p_k - 1 \right)$, où α le multiplicateur de Lagrange. Ainsi, cela implique :

$$\begin{cases} \frac{\partial \mathcal{L}a(p, \alpha)}{\partial p_1} = 0 \\ \frac{\partial \mathcal{L}a(p, \alpha)}{\partial p_2} = 0 \\ \vdots \\ \frac{\partial \mathcal{L}a(p, \alpha)}{\partial p_K} = 0 \\ \frac{\partial \mathcal{L}a(p, \alpha)}{\partial \alpha} = 0 \end{cases} \Leftrightarrow \begin{cases} \frac{\sum_{i=1}^n \pi_{i1}^{[s]}}{p_1} - \alpha = 0 \\ \frac{\sum_{i=1}^n \pi_{i2}^{[s]}}{p_2} - \alpha = 0 \\ \vdots \\ \frac{\sum_{i=1}^n \pi_{iK}^{[s]}}{p_K} - \alpha = 0 \\ \sum_{k=1}^K p_k - 1 = 0 \end{cases} \Leftrightarrow \begin{cases} \alpha = \frac{\sum_{i=1}^n \pi_{i1}^{[s]}}{p_1} \\ p_1 = \frac{\sum_{i=1}^n \pi_{i1}^{[s]}}{\sum_{i=1}^n \sum_{k=1}^K \pi_{ik}^{[s]}} \\ p_2 = \frac{\sum_{i=1}^n \pi_{i2}^{[s]}}{\sum_{i=1}^n \sum_{k=1}^K \pi_{ik}^{[s]}} \\ \vdots \\ p_K = \frac{\sum_{i=1}^n \pi_{iK}^{[s]}}{\sum_{i=1}^n \sum_{k=1}^K \pi_{ik}^{[s]}} \end{cases}$$

D'après (5.25), $\sum_{k=1}^K \pi_{ik}^{[s]} = 1$, ce qui implique que $\alpha = n$. Donc,

$$p_k^{[s+1]} = \sum_{i=1}^n \frac{\pi_{ik}^{[s]}}{n}$$

Concernant les autres paramètres, la dérivation de $\mathcal{L}a(p, \alpha)$ (identique à dériver que

(5.23)) par rapport aux paramètres τ^2 , σ_k^2 et β_k donne respectivement :

$$\frac{\partial Q}{\partial \tau^2} = -\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^K \pi_{ik}^{[s]} \left(\frac{n_i}{\tau^2} - \frac{\mathbb{E}_y^h(\varepsilon'_i \varepsilon_i)}{\tau^4} \right) \quad (5.26)$$

$$\frac{\partial Q}{\partial \sigma_k^2} = -\frac{1}{2} \sum_{i=1}^n \pi_{ik}^{[s]} \left(\frac{1}{\sigma_k^2} - \frac{\mathbb{E}_y^h(\xi_i^2)}{\sigma_k^4} \right) \quad (5.27)$$

$$\frac{\partial Q}{\partial \beta_k} = \frac{1}{\tau^2} \sum_{i=1}^n \pi_{ik}^{[s]} X'_i \left[y_i - X_i \beta_k - U_i \mathbb{E}_y^h(\xi_i) \right] \quad (5.28)$$

Par annulation des dérivées (5.26), (5.27), (5.28) et considérant les quantités $\mathbb{E}_y^h(\xi_i)$, $\mathbb{E}_y^h(\varepsilon'_i \varepsilon_i)$ et $\mathbb{E}_y^h(\xi_i^2)$, les expressions des estimateurs du maximum de vraisemblance sont :

$$\begin{aligned} \tau^{2[s+1]} &= \frac{\sum_{i=1}^n \sum_{k=1}^K \pi_{ik}^{[s]} \mathbb{E}_y^h(\varepsilon'_i \varepsilon_i)}{n} \\ \sigma_k^{2[s+1]} &= \frac{\sum_{i=1}^n \pi_{ik}^{[s]} \mathbb{E}_y^h(\xi_i^2)}{\sum_{i=1}^n \pi_{ik}^{[s]}} \\ \beta_k^{[s+1]} &= \left[\sum_{i=1}^n \pi_{ik}^{[s]} X'_i X_i \right]^{-1} \sum_{i=1}^n \pi_{ik}^{[s]} \left[X'_i y_i - X'_i U_i \mathbb{E}_y^h(\xi_i) \right] \end{aligned}$$

5.5.2 Résultats additionnels de l'étude « Moral »

Trajectoires latentes

Explication de l'appartenance aux classes

TABLE 5.17: Pvalues des tests d'indépendance entre les variables médicales et sociodémographiques, et les classes latentes obtenues.

Dimensions	Médicales			Sociodémographiques				Nb Enf.
	Taille	Typ. Chir.	Grade	Age	Etude	Sit. prof.	Sit. Fam.	
Nordin	.31	.61	.007**	.70	.41	.44	.12	.85
PF	.67	.78	.87	.18	.001**	.83	.63	.27
RF	.003**	.31	.49	.37	.19	.73	.86	.36
EF	.29	.27	.30	.003**	.68	.09	.21	.72

* : pvalue<.05; ** : pvalue<.01; *** : pvalue<.001

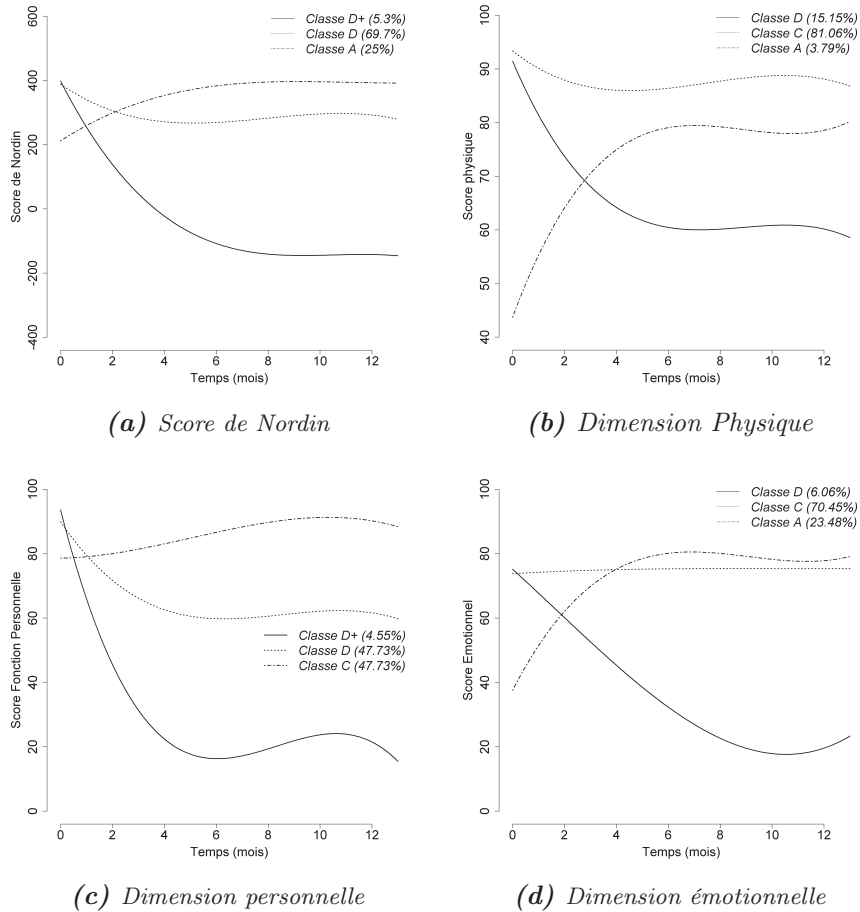


FIGURE 5.5: Trajectoires latentes issues de la classifications relatives aux dimensions suivantes : score global de Nordin, dimension physique (PF), dimension personnelle (RF) et dimension émotionnelle (EF).

TABLE 5.18: Pvalues des tests d'indépendance entre les variables psychosociales (à $t = 1$ pour celles du questionnaire MAC) et les classes latentes obtenues.

Dimensions	HAD		NEOP-R	Lot-R	MAC				
	Anx.	Dep.	Nevr	Opti.	Comb.	Imp.	Préo.	Déné.	Évit.
Nordin	.02*	.01*	.10	.48	.85	.35	.65	.20	.96
PF	.76	.09	.79	.13	.81	.88	.21	.97	.74
RF	.02*	.35	.90	.92	.05*	.59	.38	.30	.05
EF	.04*	.04	.15	.56	.09	.41	.08	.60	.36

* : pvalue<.05; ** : pvalue<.01; *** : pvalue<.001

TABLE 5.19: Résultats du modèle logistique multinomiale concernant le score global de Nordin.

Covariables	Classe D+		Classe D		Classe A		p
	n=7 (5%)	OR	n=92 (70%)	OR	n=33 (25%)	OR	
État dépressif (>7)							.12
Non	7(100%)	1	86(93%)	1	25(76%)	1	
Oui	0(0%)	0.0002	6(7%)	1	8(25%)	4.08	
État anxieux (>8)							.04*
Non	1(14%)	1	59(64%)	1	16(48%)	1	
Oui	6(86%)	8.30	33(36%)	1	17(52%)	1.48	
Grade							.005**
Grade 1	0(%)	1	27(29%)	1	4(12%)	1	
Grade 2	2(29%)	> 10 ⁻⁴	33(36%)	1	19(58%)	6.22 ¹	
Grade 3	5(71%)	> 10 ⁻⁴	17(18%)	1	9(27%)	5.12 ¹	
Névrosisme (Stable<102)							.21
Stable	5(100%)	1	41(98%)	1	21(97%)	1	
Instable	2(0%)	0.37	51(2%)	1	12(3%)	0.47	

* : pvalue<.05 ; ** : pvalue<.01 ; *** : pvalue<.001 ; ¹ : 1 ∉ IC_{95%}

TABLE 5.20: Estimations du modèle logistique multinomiale concernant la dimension physique (PF).

Covariables	Classe D		Classe C		Classe A		p
	n=20 (15%)	OR	n=107 (81%)	OR	n=5 (4%)	OR	
État dépressif							.25
Non	18(90%)	1	97(91%)	1	3(60%)	1	
Oui	2(10%)	1.70	10(9%)	1	2(40%)	5.47	
Niveau d'étude							< .001***
<Bac	3(15%)	1	60(56%)	1	4(80%)	1	
>=Bac	17(85%)	7.61 ¹	47(44%)	1	1(20%)	0.41	

* : pvalue<.05 ; ** : pvalue<.01 ; *** : pvalue<.001 ; ¹ : 1 ∉ IC_{95%}

TABLE 5.21: Estimations du modèle logistique multinomiale concernant la dimension émotionnelle (EF).

Covariables	Classe D		Classe C		Classe A		p
	n=8 (6%)	OR	n=93 (70%)	OR	n=31 (24%)	OR	
État Dépressif (>7)							.06
Non	8(100%)	1	86(92%)	1	24(77%)	1	
Oui	0(0%)	0.0008	7(8%)	1	7(23%)	3.87 ¹	
État anxieux (>8)							.15
Non	4(50%)	1	60(77%)	1	12(55%)	1	
Oui	4(50%)	1.73	33(23%)	1	19(45%)	2.37	
Âge							.005**
< 50	1(12.5%)	1	18(19%)	1	16(52%)	1	
[50; 60[6(75%)	2.55	42(45%)	1	8(26%)	.21 ¹	
≥ 60	1(12.2%)	.70	33(36%)	1	7(22%)	0.19 ¹	

* : pvalue<.05 ; ** : pvalue<.01 ; *** : pvalue<.001 ; ¹ : 1 ∉ IC_{95%}

TABLE 5.22: Estimations du modèle logistique multinomiale concernant la dimension personnelle (RF).

Covariables	Classe D-		Classe D		Classe C		<i>p</i>
	n=6 (4%)	OR	n=63 (48%)	OR	n=63 (48%)	OR	
Taille tumorale (mm)							.003**
< 15	4(67%)	1	17(30%)	1	36(60%)	1	
≥ 15	2(33%)	0.65	40(70%)	3.46 ¹	24(40%)	1	
État anxieux (>8)							.03*
Non	1(16%)	1	32(51%)	1	43(68%)	1	
Oui	5(83%)	10.43 ¹	31(49%)	1.80	20(32%)	1	
Esprit combatif							.11
Non	3(50%)	1	26(42%)	1	14(23%)	1	
Oui	3(50%)	0.36	36(58%)	0.42 ¹	48(77%)	1	

* : $p\text{-value} < .05$; ** : $p\text{-value} < .01$; *** : $p\text{-value} < .001$; ¹ : $1 \notin IC_{95\%}$

Troisième partie

Conclusions et perspectives

L'étude de la HRQoL est un objectif prioritaire des essais cliniques en cancérologie pour évaluer l'efficacité d'une prise en charge. La HRQoL constitue un critère de jugement clinique alternatif pertinent et disponible pour s'assurer de l'intérêt du traitement pour le patient et le système de santé. Elle est mesurée par l'intermédiaire d'auto-questionnaires qui la décomposent en plusieurs dimensions. Dans le contexte spécifique de la cancérologie, les questionnaires EORTC sont utilisés en Europe. Les données issues de ces questionnaires ont été à la base de nos travaux. Dans le questionnaire générique EORTC QLQ-C30, nous retrouvons les dimensions fonctionnelles, les dimensions symptomatiques et le GHS. Chacune des dimensions est mesurée par l'intermédiaire d'un ou plusieurs items construits selon une échelle de Likert. Suivant les recommandations de l'EORTC, des scores par dimension sont déduits des réponses aux items (moyenne des réponses exprimée de 0 à 100). L'analyse longitudinale de la HRQoL consiste à étudier indépendamment l'évolution des scores de HRQoL par dimension.

Dans la littérature, trois modélisations statistiques sont fortement associées à l'étude de ce type de données : celle définie par les modèles issus de l'IRT ([Masters, 1982](#)) ; celle définie par les SEM ([Rabe-Hesketh and Skrondal, 2008](#)) ; et enfin celle pour modèles mixtes étendus aux classes latentes ([Muthén and Shedden, 1999](#)). Dans ce travail, nous avons proposé des approches basées sur ces différentes modélisations pour l'analyse longitudinale de la HRQoL, ainsi que leur utilisation sur des données issues de plusieurs essais cliniques.

Dans la première partie de nos recherches, la HRQoL a été considérée comme un ensemble de dimensions indépendantes. Actuellement, les analyses longitudinales de ces données sont basées sur les scores de HRQoL et réalisées par des LMM ou des modèles de survie. Certaines dimensions étant uni-item, ces modélisations présentent des limites méthodologiques évidentes. Pour palier aux problèmes de l'analyse fondée sur ces scores, nous avons étudié les modèles IRT permettant de prendre en compte directement les données brutes. Ces modèles sont des GLMM pour données ordinales où leur particularité réside dans une paramétrisation spécifique du prédicteur linéaire. Bien que les réponses pour une dimension proviennent de plusieurs items différents, les modèles IRT modélisent une seule variable réponse. La considération de deux types de paramètres d'item dans le prédicteur linéaire permet un ajustement spécifique du modèle à chaque item. Le premier type de paramètre d'item sont les paramètres de difficulté. Ils correspondent aux paramètres de position (seuil) des GLMM pour données catégorielles et sont associés aux différentes catégories de réponse. Le second type de paramètre d'item sont les paramètres de discrimination. Ils permettent l'ajustement de la CdF à chacun des items considérés. Ces modèles regroupent presque tous les avantages qui peuvent être requis pour cette analyse. En plus des avantages qu'offrent les modèles mixtes classiques dans les analyses pour données longitudinales, les modèles IRT permettent de prendre en compte directement les données brutes. Cependant, un inconvénient réside dans la mise en œuvre et l'interprétation de ces modèles en clinique qui nécessitent une expertise particulière du statisticien. En pratique, ces modèles ne sont pas familiers ce qui engendre une réticence à leur utilisation, notamment dans le cadre d'un essai clinique en cancéro-

logie. Pour répondre à ce premier frein, une première initiative a été d'analyser un essai clinique selon un modèle IRT (chapitre 2) et de proposer une présentation des résultats intuitive et informative.

Dans un second temps, nous avons utilisé une nouvelle spécification des GLM pour données catégorielles proposée par [Peyhardi et al. \(2015\)](#) pour présenter les différents modèles IRT et choisir le plus adapté à notre problématique (chapitre 3). Le GRM, appartenant à la famille des modèles cumulatifs, semble le plus adapté pour l'analyse longitudinale de la HRQoL en cancérologie. D'abord, son interprétation à travers une variable latente sous-jacente au modèle facilite visuellement et conceptuellement l'interprétation des résultats. De plus, à condition que la CdF utilisée soit symétrique, ce modèle est invariant sous la permutation inverse. Si l'ordre des réponses aux items est renversé, cela ne modifie pas le modèle sous-jacent. Cette dernière propriété est intéressante car elle permet une interprétation plus intuitive et similaire à celle des résultats basés sur le score de HRQoL.

Outre la difficulté computationnelle de ces modèles, une perspective à ce travail est d'inclure les paramètres de discrimination dans les modèles afin d'étudier leur influence et leur utilité pour l'analyse spécifique des dimensions de HRQoL.

Une seconde étape de ce travail a été de considérer la HRQoL d'un point de vue global où les différentes dimensions sont considérées dans une seule et même analyse. En effet, ces dimensions sont inter-connectées et l'objectif est de quantifier les relations de dépendance. Pour cela, nous avons proposé dans le chapitre 4 un SEM dont les liens structurels découlent de la construction conceptuelle du questionnaire EORTC QLQ-C30. Les deux familles de dimensions (fonctionnelles et symptomatiques) caractérisent chacune un bloc de mesures dans lequel un facteur est reflété par les variables observées. Les deux facteurs définissent respectivement le statut fonctionnel et symptomatique du patient, et expliquent le GHS par des liens structurels. Ce modèle est estimé par l'algorithme EM. Un avantage certain à cette approche numérique est l'estimation des facteurs sujet-spécifiques.

À chaque temps de mesure, le SEM permet d'abord d'étudier le niveau de HRQoL ainsi que l'influence de variables explicatives sur chaque dimension de HRQoL. Les différents liens et estimations du SEM quantifient la dépendance des différentes mesures afin de mettre en évidence les dimensions qui impactent le plus sur la HRQoL globale. Une fois le SEM estimé à chaque temps de mesure, les estimations des facteurs sujets-spécifiques peuvent être introduites dans une analyse longitudinale expliquant le GHS. Cette analyse permet d'évaluer l'influence et l'interaction des facteurs avec d'autres variables explicatives au cours du temps. Une conséquence intéressante du SEM est de répondre à la problématique de la multiplicité des tests qui se présente lors de l'analyse indépendante des différentes dimensions du questionnaire.

Cependant, la conception du SEM peut être remise en cause. En effet, un poids similaire entre les dimensions a été considéré. Cependant, il est envisageable de prendre en compte différentes pondérations entre les dimensions de HRQoL, notamment suivant le nombre d'items qu'elles regroupent. De plus, un seul facteur a été supposé pour refléter les observations d'un bloc de mesure. Or, il se peut qu'une famille de di-

mension reflète non pas un, mais plusieurs facteurs. Une extension intéressante serait alors de ne pas considérer $(f^r)_{r=1,2}$ comme unidimensionnel, mais comme multidimensionnel. Par exemple, dans un bloc de mesure, un facteur pourrait représenter les dimensions générales issues du QLQ-C30, puis un second celles du QLQ-BR23. Outre ces extensions discutées, des limites sont tout de même à souligner concernant le SEM proposé. En effet, la conception du modèle ne suppose pas de lien (structurel) entre les mesures fonctionnelles et symptomatiques du patient. Il paraît toutefois logique que de forts symptômes impacteraient directement sur les capacités fonctionnelles du patient.

Dans la dernière partie de ce travail, nous avons émis l'hypothèse que la population d'intérêt n'est pas homogène et que des sous-populations homogènes se distinguent par des trajectoires de HRQoL différentes. L'objectif a été, d'une part, d'identifier des sous-populations homogènes caractérisées par des trajectoires latentes de HRQoL distinctes et, d'autre part, de déterminer des profils de patients pouvant expliquer au mieux l'évolution de la HRQoL. L'identification des patients présentant une détérioration de la HRQoL au cours du temps permettrait alors de développer des actions spécifiques afin d'améliorer leur prise en charge.

L'étape de classification a été réalisée par l'intermédiaire des mélanges de modèles mixtes. Pour cela, deux types de mélanges ont été considérés : le mélange de LMM classique appliqué sur les scores qui découlent de la procédure de scoring proposée par l'EORTC (Fayers et al., 2001), et le mélange de modèles cumulatifs appliqué sur la somme des réponses (score discret) aux items. Ces mélanges permettent d'établir les classes latentes ainsi que les trajectoires latentes. Ensuite, à l'aide d'un modèle de régression logistique (multinomiale), les variables explicatives (observations médicales et sociodémographiques) recueillies dans les études sont utilisées pour expliquer l'appartenance aux classes latentes et ainsi de définir des profils particuliers de patients suivant les trajectoires de HRQoL.

Une première comparaison entre les types de mélange utilisés a été initiée sur des données réelles et simulées. Elle a permis de mettre en avant les différences en termes de classification pour les dimensions composées d'un seul item. En effet, le mélange de LMM a tendance à surestimer le nombre de classes latentes par rapport au mélange de modèles cumulatifs. Une perspective à ce travail est d'approfondir la comparaison entre ces deux types de mélange aux dimensions multi-items. En effet, il serait intéressant d'étudier les différences concernant le choix du nombre de classes latentes, l'identification des sous-populations homogènes ou encore les trajectoires latentes. Pour conclure sur cette partie, étendre cette comparaison aux modèles IRT permettrait d'étudier l'intérêt de dissocier chaque item dans le modèle de régression pour données ordinales.

En conclusion, ce travail a permis de proposer de nouvelles stratégies d'analyse longitudinale de la HRQoL. Dans le contexte des essais cliniques, nous avons proposé dans un premier temps, de considérer directement les données brutes issues de la collecte des questionnaires ; puis dans un second temps, de proposer une analyse globale de la HRQoL, contrairement à l'analyse classique effectuée par dimension de

HRQoL et de manière indépendante. Enfin, hors du contexte des essais cliniques, l'analyse de l'hétérogénéité des données a permis d'une part, d'identifier des sous-populations homogènes définies par des trajectoires de HRQoL différentes, et d'autre part de définir des profils de patients associés à ces classes latentes grâce à des variables explicatives.

Enfin, une perspective pour l'analyse de ce critère dans les essais cliniques est la prise en compte de la nature des données manquantes. En effet, si l'état des patients ne permet pas de remplir le questionnaire à une (des) visite(s), il serait nécessaire d'inclure cette information. Le décès comme la progression apportent une information quant à la nature des données manquantes et ont l'avantage d'être recueillies dans les essais cliniques. Contrairement aux modèles TJDD, où des modèles à risque compétitif sont proposés dans cette optique, la prise en compte de cette information dans les modèles de régression n'est pas aussi aisée méthodologiquement.

Bibliographie

- (1993). Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 2(2) :153–159.
- 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., Rofe, 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.
- Agresti, A. (2002). Categorical Data Analysis. John Wiley & Sons.
- Agresti, A. (2007). An Introduction to Categorical Data Analysis. John Wiley & Sons.
- Agresti, A. (2010). Analysis of Ordinal Categorical Data. John Wiley & Sons.
- Akaike, H. (1973). {Information theory and an extension of the maximum likelihood principle}. pages 267–281. Akadémiai Kiado.
- Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43(4) :561–573.
- Anota, A. (2014). Analyse longitudinale de la qualité de vie relative à la santé en cancérologie. PhD thesis, Université de Franche-Comté.
- Anota, A., Barbieri, A., Savina, M., Pam, A., Gourgou-Bourgade, S., Bonnetain, F., and Bascoul-Mollevi, C. (2014a). Comparison of three longitudinal analysis models for the health-related quality of life in oncology : a simulation study. Health and Quality of Life Outcomes, 12.
- Anota, A., Bascoul-Mollevi, C., Conroy, T., Guillemin, F., Velten, M., Jolly, D., Mercier, M., Causeret, S., Cuisenier, J., Graesslin, O., Hamidou, Z., and Bonnetain, F. (2014b). Item response theory and factor analysis as a mean to characterize occurrence of response shift in a longitudinal quality of life study in breast cancer patients. Health and Quality of Life Outcomes, 12 :32.
- Anota, A., Hamidou, Z., Paget-Bailly, S., Chibaudel, B., Bascoul-Mollevi, C., Auquier, P., Westeel, V., Fiteni, F., Borg, C., and Bonnetain, F. (2013). Time to health-related quality of life score deterioration as a modality of longitudinal

Bibliographie

- analysis for health-related quality of life studies in oncology : do we need RE-CIST for quality of life to achieve standardization? Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation.
- Azria, D., Belkacemi, Y., Romieu, G., Gourgou, S., Gutowski, M., Zaman, K., Moscardo, C. L., Lemanski, C., Coelho, M., Rosenstein, B., Fenoglio, P., Crompton, N. E. A., and Ozsahin, M. (2010). Concurrent or sequential adjuvant letrozole and radiotherapy after conservative surgery for early-stage breast cancer (CO-HO-RT) : a phase 2 randomised trial. The Lancet. Oncology, 11(3) :258–265.
- Bacci, S., Bartolucci, F., and Gnaldi, M. (2012). A class of Multidimensional Latent Class IRT models for ordinal polytomous item responses. arXiv :1201.4667 [stat]. arXiv : 1201.4667.
- Barbieri, A., Anota, A., Conroy, T., Gourgou-Bourgade, S., Juzyna, B., Bonnetain, F., Lavergne, C., and Bascoul-Molle, C. (2015). Applying the Longitudinal Model from Item Response Theory to Assess the 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.
- Beck, F., Guilbert, A., and Gautier, A. (2010). Baromètre santé 2010, attitudes et comportements de santé. Edition INPES, Paris, 12(1).
- Beitz, J., Gnecco, C., and Justice, R. (1996). Quality-of-life end points in cancer clinical trials : the U.S. Food and Drug Administration perspective. Journal of the National Cancer Institute. Monographs, (20) :7–9.
- Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. IEEE Trans. Pattern Anal. Mach. Intell., 22(7) :719–725.
- Bjelland, I., Dahl, A. A., Haug, T. T., and Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale. An updated literature review. Journal of Psychosomatic Research, 52(2) :69–77.
- Blanchin, M., Hardouin, J.-B., Guillemin, F., Falissard, B., and Sébille, V. (2013). Power and sample size determination for the group comparison of patient-reported outcomes with Rasch family models. PloS One, 8(2) :e57279.
- Blanchin, M., Hardouin, J.-B., Neel, T. L., Kubis, G., Blanchard, C., Mirallié, E., and Sébille, V. (2011). Comparison of CTT and Rasch-based approaches for the analysis of longitudinal Patient Reported Outcomes. Statistics in Medicine, 30(8) :825–838.
- Bock, I. d., Hardouin, J.-B., Blanchin, M., Neel, T. L., Kubis, G., Bonnaud-Antignac, A., Dantan, t., and Sébille, V. (2013). Rasch-family models are more valuable than score-based approaches for analysing longitudinal patient-reported outcomes with missing data. Statistical Methods in Medical Research, page 0962280213515570.

- Boeck, P. d. and Wilson, M. (2004). Explanatory Item Response Models : A Generalized Linear and Nonlinear Approach. Springer, New York.
- Bonnetain, F. (2010). Health related quality of life and endpoints in oncology. Cancer Radiothérapie : Journal De La Société Française De Radiothérapie Oncologique, 14(6-7) :515–518.
- Bonnetain, F., Dahan, L., Maillard, E., Ychou, M., Mitry, E., Hammel, P., Legoux, J.-L., Rougier, P., Bedenne, L., and Seitz, J.-F. (2010). Time until definitive quality of life score deterioration as a means of longitudinal analysis for treatment trials in patients with metastatic pancreatic adenocarcinoma. 46 :2753–2762.
- Bradburn, N. M., Sudman, S., and Blair, E. (1979). Improving interview method and questionnaire design. Jossey-Bass.
- Brady, M. J., Cella, D. F., Mo, F., Bonomi, A. E., Tulsky, D. S., Lloyd, S. R., Deasy, S., Cobleigh, M., and Shiimoto, G. (1997). Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology, 15(3) :974–986.
- Bry, X., Lavergne, C., and Tami, M. (2015). EM estimation of a Structural Equation Model. arXiv :1510.00137 [math, stat]. arXiv : 1510.00137.
- Cayrou, S., Dickès, P., Gauvain-Piquard, A., and Rogé, B. (2003). The mental adjustment to cancer (MAC) scale : French replication and assessment of positive and negative adjustment dimensions. Psycho-Oncology, 12(1) :8–23.
- Cella, D. F., Tulsky, D. S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., Silberman, M., Yellen, S. B., Winicour, P., and Brannon, J. (1993). The Functional Assessment of Cancer Therapy scale : development and validation of the general measure. Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology, 11(3) :570–579.
- Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Gourgou-Bourgade, S., de la Fouchardière, C., Bennouna, J., Bachet, J.-B., Khemissa-Akouz, F., Péré-Vergé, D., Delbaldo, C., Assenat, E., Chauffert, B., Michel, P., Montoto-Grillot, C., and Ducreux, M. (2011). FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. The New England journal of medicine, 364(19) :1817–1825.
- De Ayala, R. J. (2009). The theory and practice of item response theory. Guilford Press.
- 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, 39(1) :1–38.
- Douglas, J. A. (1999). Item response models for longitudinal quality of life data in clinical trials. Statistics in Medicine, 18(21) :2917–2931.

Bibliographie

- Edelen, M. O. and Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 16 Suppl 1 :5–18.
- Fahrmeir, L. and Tutz, G. (2001). Multivariate Statistical Modelling Based on Generalized Linear Models. Springer.
- Fayers, P. M., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., and Bottomley, A. o. b. o. t. E. Q. o. L. G. (2001). EORTC QLQ-C30 Scoring Manual (3rd edition), volume Brussels : EORTC 2001.
- Fayers, P. M. and Machin, D. (2007). Quality of life : the assessment, analysis and interpretation of patient-reported outcomes. John Wiley & Sons, Chichester, England ; Hoboken, NJ.
- 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.
- Fiteni, F., Pam, A., Anota, A., Vernerey, D., Paget-Bailly, S., Westeel, V., and Bonnetain, F. (2015). Health-related quality-of-life as co-primary endpoint in randomized clinical trials in oncology. Expert Review of Anticancer Therapy, 15(8) :885–891.
- Fiteni, F., Westeel, V., Pivot, X., Borg, C., Vernerey, D., and Bonnetain, F. (2014). Endpoints in cancer clinical trials. Journal of Visceral Surgery, 151(1) :17–22.
- Food and Administration, D. (2007). Clinical trial endpoints for the approval of cancer drugs and biologics. Guidance for Industry, pages 1–22.
- Fox, J.-P. (2010). Bayesian Item Response Modeling : Theory and Applications. Springer, New York, NY.
- Glas, C. A. W., Geerlings, H., van de Laar, M. A. F. J., and Taal, E. (2009). Analysis of longitudinal randomized clinical trials using item response models. Contemporary Clinical Trials, 30(2) :158–170.
- Gorter, R., Fox, J.-P., and Twisk, J. W. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. BMC Medical Research Methodology, 15(1) :55.
- Gotay, C., Korn, E., McCabe, M., Moore, T., and Cheson, B. (1992). Building quality of life assessment into cancer treatment studies. Oncology (Williston Park, N.Y.), 6(6) :25–8 ; discussion 30–2, 37.
- Gourgou-Bourgade, S., Bascoul-Mollevis, C., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Boige, V., Bérille, J., and Conroy, T. (2012). Impact of FOLFIRINOX Compared With Gemcitabine on Quality of Life in Patients With Metastatic Pancreatic Cancer : Results From

- the PRODIGE 4/ACCORD 11 Randomized Trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology.
- Guilleux, A., Blanchin, M., Vanier, A., Guillemin, F., Falissard, B., Schwartz, C. E., Hardouin, J.-B., and Sébille, V. (2015). RespOnse Shift ALgorithm in Item response theory (ROSALI) for response shift detection with missing data in longitudinal patient-reported outcome studies. Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 24(3) :553–564.
- Hamidou, Z., Dabakuyo, T. S., and Bonnetain, F. (2011a). Impact of response shift on longitudinal quality-of-life assessment in cancer clinical trials. Expert Review of Pharmacoeconomics & Outcomes Research, 11(5) :549–559.
- Hamidou, Z., Dabakuyo, T. S., Mercier, M., Fraisse, J., Causeret, S., Tixier, H., Padeano, M.-M., Loustalot, C., Cuisenier, J., Sauzedde, J.-M., Smail, M., Combier, J.-P., Chevillote, P., Rosburger, C., Arveux, P., and Bonnetain, F. (2011b). Time to Deterioration in Quality of Life Score as a Modality of Longitudinal Analysis in Patients with Breast Cancer. The Oncologist, 16(10) :1458–1468.
- Hardouin, J.-B., Audureau, E., Leplège, A., and Coste, J. (2012). Spatio-temporal Rasch analysis of quality of life outcomes in the French general population. Measurement invariance and group comparisons. BMC Medical Research Methodology, 12(1) :182.
- Hardouin, J.-B., Blanchin, M., Feddag, M.-L., Néel, T. L., Perrot, B., and Sébille, V. (2015). Power and sample size determination for group comparison of patient-reported outcomes using polytomous Rasch models. Statistics in Medicine, pages n/a–n/a.
- Hedeker, D. and Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. Biometrics, 50(4) :933–944.
- Hinz, A., Einenkel, J., Briest, S., Stolzenburg, J.-U., Papsdorf, K., and Singer, S. (2012). Is it useful to calculate sum scores of the quality of life questionnaire EORTC QLQ-C30? European Journal of Cancer Care, 21(5) :677–683.
- Howard, G. S., Dailey, P. R., and Gulanick, N. A. (1979). The Feasibility of Informed Pretests in Attenuating Response-Shift Bias. Applied Psychological Measurement, 3(4) :481–494.
- Huber, C., Limnios, N., Mesbah, M., and Nikulin, M. (2013). Mathematical Methods in Survival Analysis, Reliability and Quality of Life. John Wiley & Sons.
- Institute, S. A. S. (2011). SAS/STAT 9.3 User’s Guide : Mixed Modeling (Book Excerpt). SAS Institute.
- Jaeschke, R., Singer, J., and Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. Controlled Clinical Trials, 10(4) :407–415.

Bibliographie

- Jafari, P., Bagheri, Z., Ayatollahi, S. M., and Soltani, Z. (2012). Using Rasch rating scale model to reassess the psychometric properties of the Persian version of the PedsQLTM 4.0 Generic Core Scales in school children. Health and Quality of Life Outcomes, 10(1) :27.
- Joreskog, K. G. (1970). A General Method for Analysis of Covariance Structures. Biometrika, 57(2) :239–251.
- Kepka, S., Baumann, C., Anota, A., Buron, G., Spitz, E., Auquier, P., Guillemin, F., and Mercier, M. (2013). The relationship between traits optimism and anxiety and health-related quality of life in patients hospitalized for chronic diseases : data from the SATISQOL study. Health and Quality of Life Outcomes, 11 :134.
- Keribin, C. (2000). Consistent Estimation of the Order of Mixture Models. Sankhya : The Indian Journal of Statistics, Series A (1961-2002), 62(1) :49–66.
- Linden, W. J. v. d. and Hambleton, R. K. (1997). Handbook of Modern Item Response Theory. Springer Science & Business Media.
- Little, R. J. A. and Rubin, D. B. (2002). Statistical Analysis with Missing Data. Wiley-Blackwell, 2nd edition edition.
- Liu, L. C. and Hedeker, D. (2006). A mixed-effects regression model for longitudinal multivariate ordinal data. Biometrics, 62(1) :261–268.
- Lord, F. M. (1980). Applications of Item Response Theory to Practical Testing Problems. Routledge.
- Lourme, A. (2011). Contribution à la classification par modèles de mélange et classification simultanée d'échantillons d'origines multiples. PhD thesis, Lille 1.
- Martin, O. B. (2002). Approches statistiques pour l'analyse de données des puces à ADN. Grenoble 1.
- Martinez, M.-J. (2006). Modèles linéaires généralisés à effets aléatoires : contributions aux choix de modèle et au modèle de mélange. PhD thesis, Université Montpellier 2.
- Martinez, M. J., Lavergne, C., and Trottier, C. (2009). A mixture model-based approach to the clustering of exponential repeated data. Journal of Multivariate Analysis, 100(9) :1938–1951.
- Masters, G. (1982). A rasch model for partial credit scoring. Psychometrika, 42(2) :149–174.
- McCullagh, P. (1980). Regression models for ordinal data (with discussion). Journal of the Royal Statistical Society, Series B, 42 :109–142.
- McDowell, I. (2006). Measuring Health : A guide to rating scales and questionnaires. OUP USA, 3 edition.

- Muraki, E. (1992). A Generalized Partial Credit Model : Application of an EM Algorithm. Applied Psychological Measurement, 16(2) :159–176.
- Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics, 55(2) :463–469.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. Journal of the Royal Statistical Society. Series A (General), 135(3) :370–384.
- Noonan, R. and Wold, H. (1977). NIPALS Path Modelling with Latent Variables. Scandinavian Journal of Educational Research, 21(1) :33–61.
- Nordin, K., Steel, J., Hoffman, K., and Glimelius, B. (2001). Alternative methods of interpreting quality of life data in advanced gastrointestinal cancer patients. British Journal of Cancer, 85(9) :1265–1272.
- Osoba, D., Rodrigues, G., Myles, J., Zee, B., and Pater, J. (1998). Interpreting the significance of changes in health-related quality-of-life scores. Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology, 16(1) :139–144.
- 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.
- Petersen, M. A., Groenvold, M., Aaronson, N., Blazeby, J., Brandberg, Y., de Graeff, A., Fayers, P., Hammerlid, E., Sprangers, M., Velikova, G., Bjorner, J. B., and for the European Organisation for Research and Treatment of Cancer Quality of Life Group (2006). Item response theory was used to shorten EORTC QLQ-C30 scales for use in palliative care. Journal of Clinical Epidemiology, 59(1) :36–44.
- Peyhardi, J., Trottier, C., and Guédon, Y. (2015). A new specification of generalized linear models for categorical responses. Biometrika, page asv042.
- Phillips, R., Gandhi, M., Cheung, Y. B., Findlay, M. P., Win, K. M., Hai, H. H., Yang, J. M., Lobo, R. R., Soo, K. C., and Chow, P. K. H. (2015). Summary scores captured changes in subjects' QoL as measured by the multiple scales of the EORTC QLQ-C30. Journal of Clinical Epidemiology.
- Pierce, G. R., Sarason, I. G., and Sarason, B. R. (1991). General and relationship-based perceptions of social support : are two constructs better than one? Journal of Personality and Social Psychology, 61(6) :1028–1039.
- Proust-Lima, C., Dartigues, J.-F., and Jacqmin-Gadda, H. (2015a). Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death : a latent process and latent class approach. Statistics in Medicine, pages n/a–n/a.

Bibliographie

- Proust-Lima, C., Philipps, V., and Liqueur, B. (2015b). Estimation of extended mixed models using latent classes and latent processes : the R package lamm. arXiv :1503.00890 [stat]. arXiv : 1503.00890.
- Rabe-Hesketh, S. and Skrondal, A. (2008). Classical latent variable models for medical research. Statistical Methods in Medical Research, 17(1) :5–32.
- Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalized multilevel structural equation modeling. Psychometrika, 69(2) :167–190.
- Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Danmarks Paedagogiske Institut.
- Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data : With Applications in R. CRC Press.
- Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. Psychometrika, 34(1) :100–114.
- Sébille, V., Blanchin, M., Guillemin, F., Falissard, B., and Hardouin, J.-B. (2014). A simple ratio-based approach for power and sample size determination for 2-group comparison using Rasch models. BMC medical research methodology, 14 :87.
- Schaake, W., de Groot, M., Krijnen, W. P., Langendijk, J. A., and van den Bergh, A. C. M. (2013). Quality of life among prostate cancer patients : a prospective longitudinal population-based study. Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology, 108(2) :299–305.
- Scheier, M. F. and Carver, C. S. (1985). Optimism, coping, and health : assessment and implications of generalized outcome expectancies. Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association, 4(3) :219–247.
- Schwartz, C. E. and Sprangers, M. A. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Social science & medicine, 48(11) :1531–1548.
- Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6(2) :461–464.
- Titman, A. C., Lancaster, G. A., and Colver, A. F. (2013). Item response theory and structural equation modelling for ordinal data : Describing the relationship between KIDSCREEN and Life-H. Statistical Methods in Medical Research.
- Trottier, C. (1998). Estimation dans les modeles lineaires generalises a effets aleatoires. PhD thesis, INP GRENOBLE.
- Tutz, G. (1990). Sequential item response models with an ordered response. British Journal of Mathematical and Statistical Psychology, 43(1) :39–55.

- van Nispen, R. M. A., Knol, D. L., Langelaan, M., de Boer, M. R., Terwee, C. B., and van Rens, G. H. M. B. (2007). Applying multilevel item response theory to vision-related quality of life in Dutch visually impaired elderly. Optometry and Vision Science : Official Publication of the American Academy of Optometry, 84(8) :710–720.
- Vanier, A., Leplège, A., Hardouin, J.-B., Sébille, V., and Falissard, B. (2015). Semantic primes theory may be helpful in designing questionnaires such as to prevent response shift. Journal of Clinical Epidemiology, 68(6) :646–654.
- Verhagen, J. and Fox, J.-P. (2013). Longitudinal measurement in health-related surveys. A Bayesian joint growth model for multivariate ordinal responses. Statistics in Medicine, 32(17) :2988–3005.
- Walters, S. J. (2009). Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation : A Practical Guide to Analysis and Interpretation. Wiley-Blackwell, 1 edition edition.
- Ware, J. E. and Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical Care, 30(6) :473–483.
- Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least squares. In Multivariate Analysis., pages 391–420. Academic Press.
- Wold, H. (1985). Partial Least Squares. In Encyclopedia of Statistical Sciences, volume 6, pages 581–591. John Wiley & Sons, New York.
- Wright, B. D. and Masters, G. N. (1982). Rating Scale Analysis. Pluribus Press.
- Wu, Y., Amonkar, M. M., Sherrill, B. H., O’Shaughnessy, J., Ellis, C., Baselga, J., Blackwell, K. L., and Burstein, H. J. (2011). Impact of lapatinib plus trastuzumab versus single-agent lapatinib on quality of life of patients with trastuzumab-refractory HER2+ metastatic breast cancer. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO, 22(12) :2582–2590.
- Yellen, S. B., Cella, D. F., Webster, K., Blendowski, C., and Kaplan, E. (1997). Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. Journal of Pain and Symptom Management, 13(2) :63–74.
- Zigmond, A. S. and Snaith, R. P. (1983). The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica, 67(6) :361–370.

Annexes

Liste des abréviations

ASCO	: <i>American Society of Clinical Oncology</i>
AUC	: <i>Area under Curve</i> (aire sous la courbe)
BIC	: <i>Bayesian information criteria</i>
CdF	: <i>Cumulative distribution Function</i> (fonction de répartition)
CTT	: <i>Classical Test Theory</i> (théorie classique de tests)
DMCI	: Différence Minimale Clinique Importante
EORTC	: <i>European Organisation for Research and Treatment of Cancer</i>
FACT	: <i>Functional Assesement of Cancer</i>
FDA	: <i>Food and Drug Administration</i>
GHS	: <i>Global Health Status</i> (Statut global de santé)
GLMM	: <i>Generalized Linear Mixed Model</i> (modèle linéaire généralisé mixte)
GRM	: <i>Graded Response Model</i>
HRQoL	: <i>Health-Related Quality of Life</i> (qualité de vie relative à la santé)
ICL	: <i>Intergreted Completed Likelihood</i>
IRT	: <i>Item Response Theory</i> (théorie de réponse à l'item)
LMM	: <i>Linear Mixed Model</i> (modèle linéaire mixte)
MAR	: <i>Missing At Random</i> (données manquantes aléatoires)
MAP	: Maximum a posteriori
MCAR	: <i>Missing Not At Random</i> (données manquantes non aléatoires)
MCAR	: <i>Missing Completely At Random</i> (données manquantes complètement aléatoires)
OMS	: Organisation Mondiale de la Santé
OS	: <i>Overall Survival</i> (survie globale)
PCM	: <i>Partial Credit Model</i>
PRO	: <i>Patient Reported Outcomes</i> (mesures rapporter par les patients)
QdV	: Qualité de Vie
RS	: <i>Response Shift</i>
SEM	: <i>Structural Equation Model</i> (modèle à équations structurelles)
SRM	: <i>Sequential Rasch Model</i>
TJD(D)	: Temps Jusqu'à Détérioration (Définitive)
WHOQOL	: <i>World Health Organization Quality of Life</i>
2PL	: <i>2 Parameter logistic</i>

Questionnaire EORTC QLQ-C30



EORTC QLQ-C30 (version 3)

Nous nous intéressons à vous et à votre santé. Répondez vous-même à toutes les questions en entourant le chiffre qui correspond le mieux à votre situation. Il n'y a pas de "bonne" ou de "mauvaise" réponse. Ces informations sont strictement confidentielles.

Merci de préciser:

Vos initiales:

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

Date de naissance (jour/mois/année):

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

La date d'aujourd'hui (jour/mois/année):

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

	Pas du tout	Un peu	Assez	Beaucoup
1. Avez-vous des difficultés à faire certains efforts physiques pénibles comme porter un sac à provision chargé ou une valise?	1	2	3	4
2. Avez-vous des difficultés à faire une <u>longue</u> promenade?	1	2	3	4
3. Avez-vous des difficultés à faire un <u>petit</u> tour dehors?	1	2	3	4
4. Etes-vous obligée de rester au lit ou dans un fauteuil pendant la journée?	1	2	3	4
5. Avez-vous besoin d'aide pour manger, vous habiller, faire votre toilette ou aller aux toilettes?	1	2	3	4

Au cours de la semaine passée:

	Pas du tout	Un peu	Assez	Beaucoup
6. Avez-vous été gênée pour faire votre travail ou vos activités de tous les jours?	1	2	3	4
7. Avez-vous été gênée dans vos activités de loisirs?	1	2	3	4
8. Avez-vous eu le souffle court?	1	2	3	4
9. Avez-vous ressenti de la douleur?	1	2	3	4
10. Avez-vous eu besoin de repos?	1	2	3	4
11. Avez-vous eu des difficultés pour dormir?	1	2	3	4
12. Vous êtes-vous sentie faible?	1	2	3	4
13. Avez-vous manqué d'appétit?	1	2	3	4
14. Avez-vous eu des nausées (mal au coeur)?	1	2	3	4
15. Avez-vous vomi?	1	2	3	4

Passez à la page suivante S.V.P.

Au cours de la semaine passée:

	Pas du tout	Un peu	Assez	Beaucoup
16. Avez-vous été constipée?	1	2	3	4
17. Avez-vous eu de la diarrhée?	1	2	3	4
18. Etiez-vous fatiguée?	1	2	3	4
19. Des douleurs ont-elles perturbé vos activités quotidiennes?	1	2	3	4
20. Avez-vous eu des difficultés à vous concentrer sur certaines choses par exemple pour lire le journal ou regarder la télévision?	1	2	3	4
21. Vous êtes-vous sentie tendue?	1	2	3	4
22. Vous êtes-vous fait du souci?	1	2	3	4
23. Vous êtes-vous sentie irritable?	1	2	3	4
24. Vous êtes-vous sentie déprimée?	1	2	3	4
25. Avez-vous eu des difficultés pour vous souvenir de certaines choses?	1	2	3	4
26. Votre état physique ou votre traitement médical vous ont-ils gênée dans votre vie <u>familiale</u> ?	1	2	3	4
27. Votre état physique ou votre traitement médical vous ont-ils gênée dans vos activités <u>sociales</u> (par exemple, sortir avec des amis, aller au cinéma...)?	1	2	3	4
28. Votre état physique ou votre traitement médical vous ont-ils causé des problèmes financiers?	1	2	3	4

Pour les questions suivantes, veuillez répondre en entourant le chiffre entre 1 et 7 qui s'applique le mieux à votre situation

29. Comment évalueriez-vous votre état de santé au cours de la semaine passée?

1 2 3 4 5 6 7

Très mauvais

Excellent

30. Comment évalueriez-vous l'ensemble de votre qualité de vie au cours de la semaine passée?

1 2 3 4 5 6 7

Très mauvais

Excellent

Productions scientifiques

Publications

- Anota A, Barbieri A, Savina M, Pam A, Gourgou-Bourgade S, Bonnetain F, Bascoul-Mollevi C. *Comparison of three longitudinal analysis models for the health-related quality of life in oncology : a simulation study*. Health and Quality of Life Outcomes. 2014 Dec 31;12 :192. doi : 10.1186/s12955-014-0192-2. PubMed PMID : 25551580 ; PubMed Central PMCID : PMC4326524.
- Barbieri A, Anota A, Conroy T, Gourgou-Bourgade S, Juzyna B, Bonnetain F, Lavergne C, Bascoul-Mollevi C. *Applying the Longitudinal Model from Item Response Theory to Assess the Health-Related Quality of Life in the PRODIGE 4/ACCORD 11 Randomized Trial*. Medical Decision Making. 2015 Dec 18. pii : 0272989X15621883. [Epub ahead of print] PubMed PMID : 26683246.
- Barbieri A, Peyhardi J, Lavergne C, Mollevi C. *Mixed regression models for the longitudinal analysis of health-related quality of life*, Soumis à Statistical Methods in Medical Research.
- Barbieri A, Marin JM, Florin K. *A fully Bayesian approach for the Behrens-Fisher problem in the presence of historical studies*. Soumis au Journal de la Société Française de Statistique.
- Barbieri A, Tami M, Bry X, Gourgou-Bourgade S, Azria D, Bascoul-Mollevi C, Lavergne C. *Use EM algorithm in structural equation modeling and mixed models for the longitudinal study of quality of life*. Soumission prévue à Statistics in Medicine.

Communications orales

Internationales

- Barbieri A, Lavergne C, Bascoul-Mollevi C. *Random effect Models for Quality of Life Analysis in Oncology*. 35th Annual Conference of the International Society for clinical Biostatistics (ISCB), August 2014, Vienne (Autriche).

Annexe C. Productions scientifiques

- Barbieri A, Lavergne C, Conroy T, Gourgou-Bourgade S, Juzyna B, Bascoul-Mollevi C. *Mixed-effects regression Models for longitudinal Analysis of Quality of Life in Oncology*. 21th Annual Conference of the International Society for Quality of Life Research (ISOQOL), October 2014, Berlin (Allemagne).
- Jarlier M, Barbieri A, Azria D, Gourgou-Bourgade S, Bascoul-Mollevi C. *Health-related Quality of Life (HRQoL) as a co-primary endpoint in cancer clinical trials : some elements of design and analysis of EORTC QLQ-C30*. Oncology Havana 2014 Conference, Cuba.

Nationales

- Barbieri A, Lavergne C, Conroy T, Gourgou-Bourgade S, Juzyna B, Bascoul-Mollevi C. *Modèle à crédit partiel à effets aléatoires pour l'analyse longitudinale de la Qualité de Vie en oncologie*. 9ième Journées du Cancéropole GSO, 16-18 Octobre 2013, Limoges.
- Barbieri A, Anota A., Conroy T, Gourgou-Bourgade S, Bonnetain F, Juzyna B, Lavergne C and Mollevi C. *Partial Credit Model with random effects for Quality of Life longitudinal analysis*. Workshop « Evaluation et analyse de la qualité de vie : nouveaux développements méthodologiques », avril 2014, Montpellier.
- Barbieri A, Anota A, Conroy T, Gourgou-Bourgade S, Bonnetain F, Juzyna B, Lavergne C and Mollevi C. *Modèles à effets aléatoires pour l'analyse longitudinale de la qualité de vie relative à la santé (Dans l'essai randomisé PRODIGE4/ACCORD11)*. EPICLIN 8 / 21èmes Journées des Statisticiens des CLCC, Mai 2014, Bordeaux.
- Barbieri A, Mollevi C and Lavergne C. *Modèles à effets aléatoires pour l'analyse de la Qualité de Vie en Oncologie*. 46èmes journées de Statistique de la SFDS, Juin 2014, Rennes.
- Tami M, Barbieri A, Bry X, Azria D, Gourgou S, Mollevi C et Lavergne C. *Estimation de modèles à équations structurelles par algorithme EM pour l'analyse longitudinale de la qualité de vie relative à la santé en cancérologie*. EPICLIN 9/22èmes Journées des Statisticiens des CLCC, Mai 2015, Montpellier.
- Barbieri A, Cousson-Gélie F, Gourgou S, Lavergne C et Mollevi C. *Étude longitudinale de la qualité de vie en oncologie par mélanges de modèles mixtes*. EPICLIN 10/23èmes Journées des Statisticiens des CLCC, Mai 2016, Montpellier.
- Barbieri A, Cousson-Gélie F, Gourgou S, Lavergne C et Mollevi C. *Étude longitudinale de la Qualité de Vie en cancérologie par mélanges de modèles mixtes*. 48ème journées de Statistique (SFDS), Juin 2016, Montpellier.

Séminaires

- Séminaire des doctorants de l'I3M (Université de Montpellier, département de Mathématiques), le 16 Avril 2014, Montpellier : *Longitudinal analysis of Health-Related Quality of Life in Oncology*.
- Séminaire, Laboratoire Epsilon, 3 Mars 2016, Montpellier : *Analyses de trajectoires - Étude longitudinale de la Qualité de Vie en cancérologie par mélanges*

de modèles. mixtes.

Communications affichées

- **(Internationales)** Anota A, Barbieri A, Savina M, Pam A, Gourgou-Bourgade S, Bonnetain F, Bascoul-Mollevi C. *Comparison of Three Longitudinal Analysis Models for the Health-Related Quality of Life*. 21th Annual Conference of the International Society for Quality of Life Research (ISOQOL), 15-18 octobre 2014, Berlin (Allemagne).
- **(Internationales)** Bascoul-Mollevi C, Barbieri A, Anota A, Savina M, Azria D, Bonnetain F, Gourgou-Bourgade S. *EORTC QLQ-C30 Transversal and Longitudinal Analyses with STATA Procedures*. 21th Annual Conference of the International Society for Quality of Life Research (ISOQOL), 15-18 octobre 2014, Berlin (Allemagne).
- **(Nationale)** Barbieri A, Lavergne C, Conroy T, Gourgou-Bourgade S, Juzyna B, Bascoul-Mollevi C. Partial credit model with random effects for longitudinal analysis of Quality of life in oncology. « GSO international workshop - Mathematic, biostatistics and epidemiology of cancer », 10-11 Octobre 2013, Bordeaux.
- **(Nationale)** Barbieri A, Lavergne C, Conroy T, Gourgou-Bourgade S, Juzyna B, Bascoul-Mollevi C. Modèle à crédit partiel à effets aléatoires pour l'analyse longitudinale de la Qualité de Vie en oncologie. 9ième Journées du Cancéropole GSO, 16-18 Octobre 2013, Limoges.
- **(Nationale)** Barbieri A, Tami M, Bry X, Azria D, Gourgou S, Mollevi C and Lavergne C. *Analyse longitudinale de la qualité de vie en cancérologie par modèles à équations structurelles et à effets aléatoires*. 11ième Journées Cancéropôle Grand Sud-Ouest, Novembre 2015, Bordeaux, France.

Méthodes longitudinales pour l'analyse de la Qualité de Vie Relative à la Santé en Cancérologie

Résumé : L'étude de la qualité de vie relative à la santé est un objectif prioritaire des essais cliniques en cancérologie pour évaluer l'efficacité d'une prise en charge ; elle est mesurée par le biais d'auto-questionnaire. Dans ce travail, nous proposons différentes modélisations statistiques pour l'analyse longitudinale de ce critère, ainsi que leur application sur des données issues de plusieurs essais cliniques. Une première partie présente les modèles issus de la théorie de réponse à l'item (IRT) pour réaliser une analyse longitudinale directement sur les données brutes (multi-réponses ordinales) et ce par dimension. Une fois replacés dans le contexte des modèles linéaires généralisés mixtes, une sélection conceptuelle de modèles IRT a conclu que le Graded response model semble le mieux adapté. Dans une seconde partie, nous proposons un modèle à équation structurelle permettant de prendre en compte conjointement l'aspect multidimensionnel et longitudinal de la qualité de vie. À l'aide de facteurs reflétés par des ensembles de variables observées, il permet de lier à chaque temps de mesure toutes les observations issues du questionnaire, tout en considérant également des variables explicatives. L'analyse longitudinale est réalisée sur le statut global de santé et les facteurs réduisant ainsi le nombre de tests. Enfin, une approche par mélanges de modèles mixtes est proposée pour obtenir des classes latentes à partir de trajectoires de qualité de vie. Cette approche a permis de caractériser des sous-populations homogènes et d'associer différente évolution de la qualité de vie suivant des profils particuliers de patients.

Mots clés : Qualité de vie relative à la santé, Analyses longitudinales, Modèles mixtes, Modèles IRT, Modèle à équation structurelle, Mélange de modèles mixtes.

Longitudinal methods for the health-related quality of life analysis in oncology

Abstract : The health-related quality of life is a major objective in oncology clinical trials to improve patients' care and better evaluate the impact of the treatments on their everyday life. Auto-questionnaires are usually used to measure this endpoint. In this work, different statistical models for the longitudinal analysis of health-related quality of life in oncology are proposed and applied to clinical trial data. First, we present different models derived from the item response theory (IRT) to achieve a longitudinal analysis directly on raw data (multi-response outcomes) for each dimension. Within the generalized linear mixed model background, a conceptual selection of the IRT models concluded that the graded response model seems to be the most suitable. Then, we propose a structural equation model which allows taking into account the multidimensional nature of data at each time and the longitudinal aspect induced by the repeated measurements. At each measurement time, the model allows to link all the observed variables issued from the questionnaire given explanatory variables. Two factors are estimated, each summarizing a set of observed variables. The longitudinal analysis is performed on the global health status and on the factors, thus reducing the number of tests. Finally, an approach based on a mixture of mixed models is used to obtain latent classes from quality of life trajectories. The approach has resulted in the identification of homogeneous subpopulations and their latent trajectory according to specific patient profiles.

Keywords : Health-related quality of life, Longitudinal analysis, Mixed models, IRT models, Structural equation model, Mixture of mixed models.