



**HAL**  
open science

# Comparaisons des méthodes d'analyse des données binaires ou ordinales corrélées. Application à l'étude longitudinale de l'incapacité des personnes âgées

Isabelle Carrière

► **To cite this version:**

Isabelle Carrière. Comparaisons des méthodes d'analyse des données binaires ou ordinales corrélées. Application à l'étude longitudinale de l'incapacité des personnes âgées. Sciences du Vivant [q-bio]. Université Paris Sud - Paris XI, 2005. Français. NNT: . tel-00107384

**HAL Id: tel-00107384**

**<https://theses.hal.science/tel-00107384>**

Submitted on 18 Oct 2006

**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.

UNIVERSITE PARIS XI

FACULTE DE MEDECINE PARIS-SUD

Année 2005

Thèse n°

## THESE

Pour obtenir le grade de  
DOCTEUR DE L'UNIVERSITE PARIS XI

Spécialité : Santé Publique

Option : Epidémiologie

Présentée et soutenue publiquement

par

Isabelle CARRIERE

Le 26 septembre 2005

Comparaisons des méthodes d'analyse des données binaires ou  
ordinales corrélées. Application à l'étude longitudinale de  
l'incapacité des personnes âgées

Directeur de thèse : M. Jean BOUYER

### JURY

M. Denis HÉMON	Président
M. Jean BOUYER	Directeur de thèse
M. Pierre CHAUVIN	Rapporteur
M. Paul LANDAIS	Rapporteur
Mme Annette LECLERC	Examineur
M. Luc LETENNEUR	Examineur

## Remerciements

Je tiens à remercier tout particulièrement Jean Bouyer d'avoir accepté d'être le directeur de ma thèse. Malgré la distance géographique tu as toujours été très disponible et de bon conseil. J'ai eu la chance de bénéficier de tes connaissances en épidémiologie et en méthodologie de la recherche que tu transmets avec simplicité et pédagogie. Sois assuré de toute ma reconnaissance.

Je remercie Laure Papoz de m'avoir accueillie dans son unité INSERM et de m'avoir offert l'opportunité de débiter cette thèse. Merci d'avoir accompagné ce long chemin de votre bienveillante sollicitude.

Je tiens à remercier Alain Colvez pour m'avoir initiée avec enthousiasme au domaine de la gérontologie et de la prise en charge des personnes âgées. Nos discussions animées sur les concepts de dépendance et de fragilité ont été le point de départ de ce travail de recherche.

Je remercie Karen Ritchie d'avoir accepté que je rejoigne son unité INSERM voici un an et demi et de m'avoir permis de terminer cette thèse dans de très bonnes conditions. Notre nouvelle collaboration me permet de donner un nouvel élan à mes travaux.

Je remercie Cécile Delcourt et Marie Laure Ancelin pour leur compréhension et leurs encouragements à mon égard.

J'exprime toute ma reconnaissance à Denis Hémon pour l'honneur qu'il me fait d'avoir accepté la présidence de ce jury. Votre apport dans le développement de l'épidémiologie et de son enseignement en France est considérable.

Je suis extrêmement reconnaissante au professeur Paul Landais et à Pierre Chauvin d'avoir accepté d'être les rapporteurs de ma thèse. Leurs commentaires et suggestions m'ont permis d'améliorer substantiellement ce document.

Je remercie Annette Leclerc et Luc Letenneur pour leur participation au jury de ma thèse. Vos grandes compétences en épidémiologie et dans le vieillissement nous sont précieuses.

Je remercie chaleureusement tous mes collègues de l'INSERM pour leur soutien amical avec une pensée spéciale pour Brigitte Bonal, Isabelle Jaussent, Annie lacroux et Joanna Norton qui m'ont aidé dans ce travail.

Merci enfin à toute ma famille et à tous mes amis pour leur soutien sans faille et leurs encouragements indéfectibles tout au long de ces années.

## Productions scientifiques issues du travail de la thèse

**Carrière I.**, Bouyer J.

Choosing marginal or random-effects models for longitudinal binary responses. Application to self-reported disability among older persons

BMC Medical Research Methodology 2002; 2:15

**Carrière I.**, Colvez A., Favier F., Jeandel C., Blain H. for the EPIDOS study group

Hierarchical components of physical frailty predict incidence of dependency in a cohort of elderly women

Journal of clinical Epidemiology 2005; accepté

**Carrière I.**, Bouyer J.

Random-effect models for ordinal responses : Application to self-reported disability among older persons.

Rev. Epidémiol. Santé. Publ. 2005; en révision

**Carrière I.**, Bouyer J., Favier F., Colvez A.

Modèles à effets aléatoires pour réponses ordinales. Application à l'évolution de l'incapacité chez les personnes âgées.

Rev. Epidémiol. Santé. Publ. 2004; 52: 1S43. Communication orale.

Delcourt C., **Carrière I.**, Delage M., Descomps B., Cristol JP., Papoz L. and the POLA study group

Associations of cataract with antioxidant enzymes and other risk factors: the French age-related eye diseases (POLA) prospective study.

Ophthalmology. 2003; 110:2318-2326.

Delcourt C., Lacroux A., **Carrière I.** and the POLA Study Group.

The three-year incidence of age-related macular degeneration: the POLA prospective study.

American Journal of Ophthalmology 2005; sous presse

# Table des matières

Introduction, objectifs de la thèse .....	9
Partie I : Dépendance des personnes âgées .....	13
1 Cadre conceptuel .....	16
2 Instruments de mesure.....	20
3 Spécificités statistiques de la dépendance et modélisation .....	22
4 Cohorte EPIDOS .....	26
Partie II : Réponses binaires.....	29
1 Méthodes d'analyse des réponses binaires corrélées .....	30
1.1 Modèle linéaire généralisé .....	33
1.2 Qualité d'ajustement et tests de comparaison des modèles .....	35
1.3 Modèle marginal .....	37
1.4 Modèle mixte linéaire généralisé .....	40
1.5 Effets fixes et effets aléatoires .....	45
1.6 Données manquantes.....	46
1.7 Covariables dépendantes du temps .....	48
1.8 Interprétation des effets fixes .....	50
2 Comparaisons GEE/GLMM sur les données de dépendance binaires.....	52
2.1 Description des données.....	53
2.2 Choix de l'origine du temps.....	55
2.3 Analyse de l'évolution selon l'âge et le temps.....	58
2.4 Analyse des autres facteurs de risque.....	65
2.5 Discussion .....	67
3 Exemple d'utilisation du GLMM pour réponses binaires : construction d'un score prédictif de la dépendance des personnes âgées.....	70
3.1 Objectifs .....	70
3.2 Matériel et méthode.....	71
3.3 Résultats .....	75
3.4 Discussion .....	82
3.4.1 Pertinence de la population ciblée.....	82
3.4.2 Définition de l'indicateur d'incapacité.....	83
3.4.3 Mise en évidence d'une hiérarchie entre les composants de la fragilité.....	84

3.4.4	Applicabilité du score prédictif.....	86
	Partie III : Réponse ordinaire.....	88
1	Méthodes d'analyse des réponses ordinales corrélées.....	89
1.1	Modèles logistiques, ordinaux sans effet aléatoire.....	90
1.1.1	Modèle logistique polytomique.....	90
1.1.2	Modèle à odds proportionnels.....	91
1.1.3	Modèle non contraint, à odds partiellement proportionnels.....	92
1.1.4	Modèle contraint, à odds partiellement proportionnels.....	93
1.1.5	Modèle logistique, ordonné, généralisé.....	93
1.1.6	Modèle Stéréotype.....	93
1.1.7	Modèle de ratios de continuation (Continuation ratio model).....	95
1.1.8	Modèle logistique sur catégories adjacentes (adjacent-category logistic model) .....	96
1.2	Modèles à effets aléatoires.....	96
1.2.1	Modèle logistique polytomique à effets aléatoires.....	97
1.2.2	Modèle logistique à odds proportionnels et effets aléatoires.....	98
1.2.3	Modèle logistique, ordonné, généralisé avec effets aléatoires.....	99
1.2.4	Modèle Stéréotype avec effets aléatoires.....	101
1.2.5	Méthodes d'estimation des modèles logistiques ordinaux à effets aléatoires.	102
1.3	Sélection des modèles et critères d'adéquation aux données ordinales.....	102
1.4	Conclusion.....	103
2	Exemple d'utilisation du GLMM pour réponses ordinales.....	105
2.1	Matériel et méthodes.....	105
2.2	Sélection des modèles.....	107
2.3	Qualité du modèle.....	112
2.4	Conclusion.....	113
	Discussion générale.....	115
	Bibliographie.....	123
	Annexe 1 : Articles originaux.....	131
	Choosing marginal or random effects models for longitudinal binary responses: application to self-reported disability among older persons.....	131
	Hierarchical components of physical frailty predict incidence of dependency in a cohort of elderly women.....	141

Random-effect models for ordinal responses: Application to self-reported disability among older persons. ....	156
Annexe 2 : Notations.....	172
Annexe 3 : liste des abréviations.....	173



## Figures

Figure 1 : Schéma conceptuel de Nagi.....	17
Figure 2 : Schéma conceptuel de la classification OMS de 1980.....	18
Figure 3 : Schéma conceptuel de la classification OMS de 2001.....	19
Figure 4 : Simulation de la probabilité d'une réponse positive dans un modèle logistique avec une ordonnée à l'origine aléatoire en fonction d'une covariable continue .....	32
Figure 5 : Extensions du modèle linéaire aux données binaires groupées.....	41
Figure 6 : Probabilité à 5 ans d'être en incapacité en fonction de l'âge à l'inclusion chez les femmes sans incapacité à l'inclusion et survivant jusqu'à la fin de l'étude. ....	63
Figure 7 : Evolution de la probabilité d'être en incapacité chez les femmes sans incapacité à l'inclusion, vivantes à la fin de l'étude, âgées de 75 et 85 ans à l'inclusion.....	63
Figure 8 : Abaque de correspondance entre le score et la probabilité d'être en incapacité.....	82

# Introduction, objectifs de la thèse

Notre travail se situe au carrefour de la biostatistique et de l'épidémiologie. Il a pour ambition de mettre à la disposition des épidémiologistes, des méthodes statistiques innovantes répondant à une problématique concrète. Partant d'un exemple pratique de recherche épidémiologique, nous comparerons ces méthodes d'analyse, définirons l'outil statistique le plus adapté à notre contexte et appliquerons cette stratégie d'analyse à l'élaboration d'un score prédictif. Les conclusions obtenues à chaque étape seront discutées dans une perspective plus large afin de voir dans quelle mesure elles peuvent être généralisées à d'autres domaines.

En réponse au problème important de santé publique que constitue le vieillissement de la population dans les pays développés, les recherches sur les mécanismes conduisant à la perte d'autonomie des personnes âgées ont connu un grand essor qui s'est rapidement heurté à des problèmes d'ordre méthodologique, révélant un besoin de conceptualisation du phénomène de l'incapacité et de développement d'outils aidant à son analyse. Bien que beaucoup plus riches que les études transversales, les études longitudinales n'ont jusqu'alors été utilisées que dans peu de recherches, probablement en raison des difficultés que pose l'analyse statistique de ces études. Notre objectif est donc de rechercher les méthodes permettant de décrire et d'analyser au mieux le phénomène de l'incapacité chez les personnes âgées.

Cette recherche s'inscrit dans le cadre plus général de l'analyse des données corrélées auxquelles nous sommes souvent confrontés avec le développement de bases de données informatisées aussi variées que les fichiers de l'Assurance Maladie, les registres de maladies, les études de cohortes. Ce terme générique de données corrélées recouvre une multitude de structures de données telles que des données en grappes (familles, hôpitaux, centres d'essais thérapeutiques), des mesures répétées, des données longitudinales ou des données spatiales.

Parmi l'ensemble de ces données corrélées, les études longitudinales nécessitent probablement les modélisations les plus élaborées. Diggle et al [1] recensent trois sources de variabilité. La

première concerne la variabilité inter-individuelle ou l'hétérogénéité des profils d'évolution. La seconde est l'association dans le temps, caractérisée par le fait que des mesures proches dans le temps sont plus semblables que des mesures plus espacées. Cette association dépend alors de la distance entre les mesures. Enfin, il existe toujours l'erreur de mesure qui induit, pour des tests délicats, que même des répliques immédiates sont empreintes d'une certaine variation. Dans les données groupées en grappes, seules la première et la dernière source de variabilité existent, tandis que dans les données spatiales, ce sont les deux dernières sources qui sont présentes.

Ces considérations font que l'adaptation de modèles pour données longitudinales aux données groupées en grappes ou spatiales est, en général, plutôt aisée et que les méthodes étudiées sont facilement transposables aux types de données corrélées autres que les données longitudinales.

Outre la structure des données, le type de la variable à expliquer (ou variable réponse) est une caractéristique essentielle à l'analyse statistique des données. Les méthodes pour réponses continues sont les plus développées et les mieux mises en oeuvre dans les logiciels. Ceci est principalement dû aux propriétés de la distribution normale qui permettent d'importantes simplifications mathématiques. Les logiciels offrent de nombreux outils dans ce domaine. Cependant les réponses catégorielles (binaires, nominales ou ordinales) sont très répandues dans la pratique. Citons par exemple, les scores de qualité de vie, les stades d'une maladie, la présence d'un symptôme.

La nature de ces réponses oriente vers des méthodes statistiques plus complexes, encore rarement utilisées dans le cadre des données corrélées. Notre objectif est de comparer ces méthodes et plus particulièrement le modèle marginal et le modèle à effets aléatoires.

Dans cette thèse, nous reviendrons dans une première partie sur les concepts de dépendance et d'incapacité des personnes âgées et nous verrons les spécificités statistiques que présentent les instruments de mesure associés. Nous présenterons la cohorte Epidos sur laquelle se fondent toutes nos analyses.

La partie II concernera les modèles d'analyse des données binaires corrélées que nous replacerons dans le cadre théorique des modèles linéaires généralisés. Nous insisterons sur l'interprétation des paramètres estimés pour les modèles marginaux et les modèles à effets aléatoires, et nous porterons notre attention sur les problèmes posés par le choix des effets aléatoires, les données manquantes et les covariables dépendantes du temps.

Un chapitre sera consacré à la comparaison des modèles marginaux et des modèles à effets aléatoires dans l'analyse de l'incapacité considérée comme une réponse binaire.

Le modèle à effets aléatoires sera ensuite utilisé pour mettre au point un score prédictif de l'incapacité issu de l'analyse des facteurs de risque potentiels, disponibles dans notre base de données Epidos.

Dans la partie III, nous décrirons les différents modèles logistiques, utilisés dans la littérature pour analyser les réponses ordinales. Nous verrons comment ces modèles peuvent être étendus à l'analyse de données longitudinales. Nous nous attarderons alors sur la question de la sélection des modèles et des critères d'adéquation des modèles aux données.

Enfin, dans un dernier chapitre, nous appliquerons les modèles logistiques, ordonnés, mixtes à la recherche d'effets différenciés des facteurs de risque sur les stades d'incapacité.

# Partie I : Dépendance des personnes âgées

La prévalence élevée de l'incapacité chez les personnes âgées, dans une classe d'âge en constante augmentation, représente un des problèmes de santé publique les plus importants auxquels sont confrontés les systèmes de santé des pays développés. L'incapacité des personnes âgées implique une utilisation croissante des services de santé et de lourds besoins en aide formelle et informelle.

Ainsi en France, la proportion des personnes de 60 ans et plus dans la population totale est passée de 18% en 1970 à 21,8% en 2004 soit 13,5 millions de personnes [2]. Les projections démographiques dessinent une augmentation inéluctable : les personnes âgées de 60 ans et plus seraient au nombre de 21 millions en 2035 soit 32,8% de la population. Cette croissance se ralentirait cependant vers 2035-2040, lorsque les générations moins nombreuses parviendront à leur tour à des âges élevés et que commenceront à disparaître les générations nombreuses nées entre 1945 et 1975. En 2050, 22,4 millions d'habitants pourraient être âgés de plus de 60 ans, soit 35% de la population totale et 11,6 millions pourraient être âgés de 75 ans et plus, soit 18,1% de la population totale [3]. L'évolution du nombre de personnes âgées dépendantes est très liée aux mouvements démographiques et au passage des générations dans les différentes classes d'âge. Les projections concernant le nombre de personnes âgées dépendantes sont plus délicates car outre l'évolution de la mortalité elles doivent tenir compte de l'évolution de la prévalence de la dépendance selon les générations. Cependant, elles prévoient une augmentation de 14 à 18% selon le scénario le plus optimiste qui suppose que les taux de prévalence de la dépendance continuent à baisser au même rythme que celui observé dans les années 1990. En termes d'effectifs, on passerait d'environ 800 000 personnes âgées dépendantes en 2000 à 910 000 en 2020 [4].

Dans ce contexte, l'analyse des facteurs de risque d'entrée en incapacité représente un enjeu majeur pour la santé publique et les politiques de protection sociale. En effet, la connaissance des mécanismes aboutissant à une diminution de l'autonomie dans la vie quotidienne pourrait

permettre de retarder le processus. Des actions de dépistage et de prévention des personnes à risque pourraient être mises en place pour allonger l'espérance de vie sans incapacité. Ces actions ciblées pourraient avoir un grand intérêt dans la mesure où retarder l'entrée en incapacité de quelques années dans cette tranche d'âge où la mortalité est élevée réduirait de manière sensible la prévalence.

Cependant, plusieurs problèmes méthodologiques subsistent tant en ce qui concerne la mesure de l'incapacité que la modélisation de son évolution au cours du temps. Ceci rend difficile l'évaluation précise de l'importance du problème et les comparaisons de prévalence entre pays ou études.



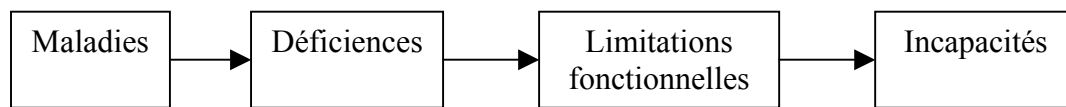
# 1 Cadre conceptuel

Bien qu'indispensables, les concepts de base et les définitions des termes employés ne sont pas unanimement reconnus. Le terme général d'autonomie se définit par la liberté de choisir son mode de vie, c'est-à-dire la capacité d'une personne à choisir les règles de conduite, l'orientation de ses actes et les risques qu'elle est prête à courir. Ce terme ne doit pas être confondu avec l'autonomie fonctionnelle. La dépendance est définie comme l'état pour lequel l'action d'un tiers est nécessaire pour atteindre ses propres buts. En France, la loi de 1997 instaurant la prestation spécifique dépendance (PSD) indiquait que la dépendance est "l'état de la personne qui, nonobstant les soins qu'elle est susceptible de recevoir, a besoin d'être aidée pour l'accomplissement des actes essentiels de la vie et requiert une surveillance régulière". La loi du 20 juillet 2001 relative à l'allocation personnalisée d'autonomie (APA) indique que "cette allocation [...] est destinée aux personnes qui, nonobstant les soins qu'elles sont susceptibles de recevoir, ont besoin d'une aide pour l'accomplissement des actes essentiels de la vie ou dont l'état nécessite une surveillance régulière".

Avec l'expansion du phénomène de chronicité et les besoins en matière de réadaptation et d'indemnisation, il était devenu nécessaire de développer des modèles conceptuels.

La première tentative de classification émane du sociologue américain Nagi [5] qui en 1976 propose de distinguer quatre niveaux : 1) les pathologies (maladies, blessures, malformations congénitales), 2) les déficiences qui sont les conséquences des pathologies au niveau des organes et qui comprennent les anomalies anatomiques, physiologiques, intellectuelles, 3) les limitations fonctionnelles définies comme "des limitations dans les performances au niveau du corps entier ou de la personne" qui englobent le fonctionnement sensori-moteur de l'organisme, les capacités intellectuelles et de raisonnement 4) les incapacités définies comme

"des limitations des performances dans les rôles sociaux et les tâches dans un environnement social ou culturel". Les concepts de Nagi demeurent largement utilisés en Amérique du Nord et peuvent se résumer selon le schéma suivant.



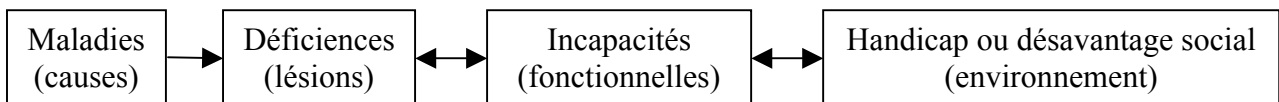
**Figure 1 : Schéma conceptuel de Nagi.**

C'est en 1980 que la Classification internationale des déficiences, des incapacités et des handicaps vit le jour sous l'égide de l'OMS. Plus récemment en 2001, une nouvelle Classification Internationale du Fonctionnement, du handicap et de la santé (CIF) a été adoptée par l'OMS.

La première classification, inspirée des travaux de Wood [6-8], conçoit le handicap comme un phénomène individuel et décrit de manière linéaire les liens de cause à effet : la maladie engendre une déficience organique et fonctionnelle, laquelle mène à une incapacité au niveau des comportements et des activités de la personne, incapacité qui, à son tour, produit un désavantage. Dans ce découpage conceptuel, les déficiences traduisent comme pour Nagi des lésions au niveau de l'organisme : déficience visuelle, déficience de l'appareil locomoteur ou déficience des fonctions cognitives. L'incapacité fait référence à l'impossibilité ou la difficulté à accomplir les activités considérées comme normales pour l'être humain. Elle est évaluée par des épreuves standardisées par exemple examen de la marche, épreuves psychométriques. Elle correspond globalement à ce que Nagi a appelé les limitations fonctionnelles. Le désavantage ou handicap traduit les conséquences dans la vie sociale. Il est apprécié en se fondant sur les performances des personnes pour six "rôles de survie" : la mobilité physique, l'indépendance physique pour les actes élémentaires de la vie quotidienne (manger, s'habiller, se laver, aller aux toilettes), l'indépendance pour des actes plus élaborés appelés activités instrumentales de la vie quotidienne (préparer les repas, faire les courses, gérer un budget, remplir des papiers

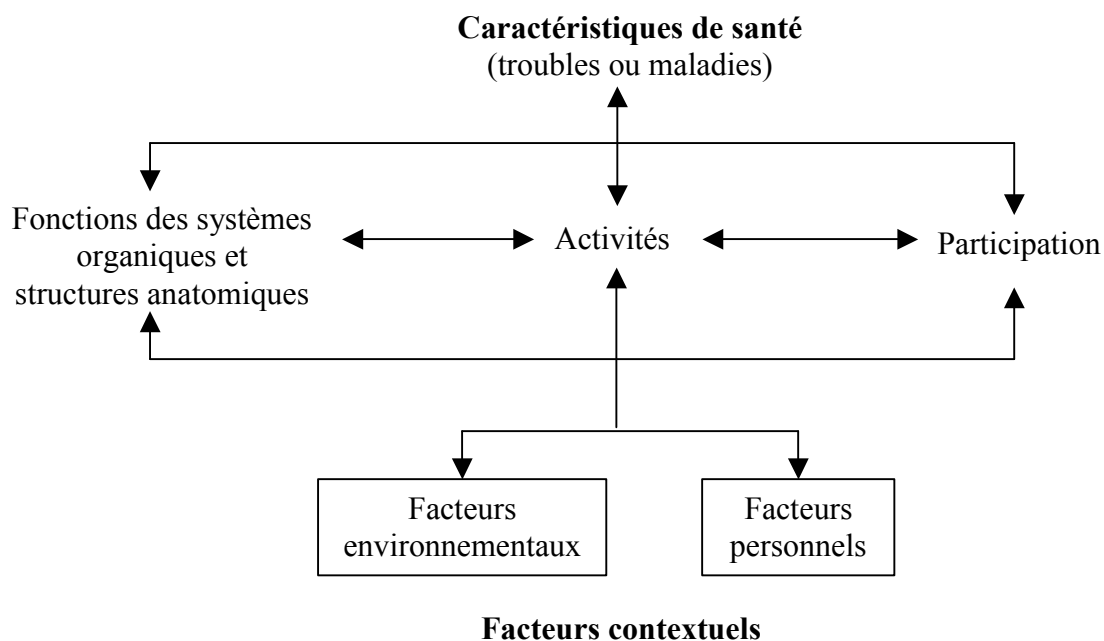
administratifs, utiliser le téléphone, prendre des médicaments), l'intégration sociale, la suffisance économique, l'orientation dans le temps et dans l'espace. Ce désavantage social est proche de la notion d'incapacité définie par Nagi.

Dans cette définition, les incapacités sont moins en relation avec le milieu social ou familial que dans la définition de Nagi. Les incapacités dans la classification de l'OMS sont objectivées par des tests en dehors de la situation de vie. Comme indiqué dans le schéma suivant, dans cette classification les flèches sont à double sens car on prend en compte que l'environnement social peut induire des incapacités et que les incapacités peuvent induire des déficiences.



**Figure 2 : Schéma conceptuel de la classification OMS de 1980.**

Ces deux classifications furent critiquées car elles ne prenaient pas suffisamment en compte la participation sociale et le contexte environnemental. La nouvelle classification CIF [9] propose un cadre conceptuel permettant de décrire la santé et ses domaines connexes. Elle tend à faire la synthèse entre le modèle médical qui voit l'incapacité au niveau de la personne comme le produit de maladies ou de problèmes de santé nécessitant une prise en charge médicale et le modèle social qui voit l'incapacité comme un problème créé par un environnement mal adapté et qui nécessite une réponse politique. Le modèle sur lequel se fonde la CIF est appelé "bio-psychosocial", il est décrit dans le diagramme suivant.



**Figure 3 : Schéma conceptuel de la classification OMS de 2001.**

L'état de fonctionnement et de handicap d'une personne est le résultat de l'interaction dynamique entre son état de santé et les facteurs contextuels. Les facteurs contextuels comprennent les facteurs environnementaux externes (structures sociales, caractéristiques architecturales, climat, ...) et les facteurs personnels (sexe, âge, niveau d'éducation, profession, ...). L'état de fonctionnement comprend trois domaines : les fonctions organiques ou anatomiques (déficiences), les activités individuelles (limitations des capacités), la participation dans un contexte social (restrictions de performances). Tous ces domaines interagissent les uns sur les autres.

La CIF est une classification cadre qui propose une analyse selon un schéma, un langage et des codes universels. Elle ne fournit pas les instruments de mesures pour les différentes dimensions décrites. Elle laisse aux professionnels sur le terrain l'élaboration des outils et des critères d'évaluation.

L'utilisation en épidémiologie de ces concepts permet de prendre en compte les différents aspects de la question et offre des grilles d'interprétation du phénomène.

## 2 Instruments de mesure

Les instruments de mesure utilisés se focalisent principalement sur les limitations d'activités et la restriction des performances individuelles. Chaque enquête définit l'incapacité souvent de manière différente. Le développement et l'homogénéisation d'indicateurs couvrant les différents aspects du domaine n'en sont qu'à leurs balbutiements.

Les deux indicateurs utilisés au niveau international sont l'indicateur de Katz [10] et l'indicateur défini par Lawton [11]. Ils se situent tous deux dans le cadre conceptuel de Nagi.

Le premier synthétise les besoins d'aide pour les six activités de la vie quotidienne (AVQ) : s'habiller, faire sa toilette, aller aux toilettes et les utiliser, se déplacer du lit au fauteuil, se nourrir, contrôler ses selles et ses urines. Cet indicateur est donc destiné à mesurer les incapacités lourdes. Le second concerne les activités instrumentales de la vie quotidienne (AIVQ) : faire ses courses, préparer les repas, entretenir sa maison, laver son linge, se servir du téléphone, utiliser des moyens de transport, prendre ses médicaments, gérer son budget. Pour chacune de ses activités, le degré d'aptitude de la personne est relevé, en général au moyen des niveaux suivants : fait seul sans aide et sans difficulté ; fait seul, sans aide avec quelques difficultés ; fait seul, sans aide avec beaucoup de difficultés ; ne fait pas sans aide. L'indicateur des AIVQ révèle donc des incapacités plus précoces mais qui rendent cependant la personne dépendante de son entourage. Il a été aussi proposé un indicateur de mobilité [12] comprenant 4 niveaux : confiné au lit ou au fauteuil, besoin d'aide pour la toilette et l'habillement, besoin d'aide pour sortir, sort sans aide.

En France, l'attribution des aides pour les personnes âgées dépendantes est effectuée sur la base d'une grille d'évaluation originale : Autonomie Gérontologique - Groupes Iso-Ressources (AGGIR). La grille AGGIR évalue l'autonomie selon certains critères et regroupe les malades

en six groupes "iso-ressources". Un groupe iso-ressources comprend des personnes qui nécessitent une même mobilisation de ressources pour faire face à leur dépendance. L'observation porte sur les activités effectuées par la seule personne âgée, en excluant tout ce que font les aidants et les soignants. Les aides matérielles et techniques sont considérées comme faisant partie de la personne : lunettes, prothèses auditives, fauteuil roulant, poche de colostomie, etc. AGGIR comporte dix variables à coder : cohérence, orientation, toilette, habillage, alimentation, élimination, transfert, déplacement à l'intérieur, déplacement à l'extérieur, communication à distance. C'est selon cette grille que les statistiques françaises sur le nombre de personnes âgées dépendantes sont publiées. Malheureusement cette classification n'est utilisée qu'en France et ne permet pas une comparaison avec les autres pays.

Très récemment, de nouveaux instruments intégrant l'évaluation et la définition de champs d'intervention ont été mis à disposition des professionnels. Parmi ceux-ci, le RAI (Résident Assessment Instrument) avec sa nouvelle version pour personnes vivant à domicile comprend l'évaluation de la situation de la personne par le Minimum de Données Standardisées (MDS), une stratégie d'intervention pour chaque domaine repéré grâce à un guide d'analyse et une évaluation des résultats [13]. Au Canada, un Système de Mesure de l'Autonomie Fonctionnelle (SMAF) a été développé [14, 15] . Il consiste en une échelle de 29 items mesurant l'autonomie fonctionnelle : AVQ, mobilité, communication, fonctions mentales, AIVQ. Il constitue la base d'un système de 14 profils d'autonomie qui vont guider la prescription des services.

Ces deux derniers outils sont plus orientés vers une évaluation de la personne dans le but de définir un plan d'aide individuel. Mais une partie des informations ainsi recueillies pourrait être utilisée comme indicateur statistique.

### 3 Spécificités statistiques de la dépendance et modélisation

Indépendamment du critère d'évaluation retenu, le phénomène de la dépendance présente des caractéristiques statistiques qui rendent son analyse complexe. La manière dont les indicateurs sont utilisés affecte largement l'information qu'ils fournissent.

La caractéristique principale est que l'incapacité n'est pas toujours un état définitif (absorbant), mais qu'une proportion non négligeable de personnes ressortent de cet état et retrouvent la totalité de leur autonomie fonctionnelle. Cette caractéristique est communément nommée "réversibilité". L'incapacité doit donc être perçue comme un épisode de durée variable pouvant aller de quelques semaines à plusieurs années [16]. Lors d'une enquête transversale tendant à évaluer la prévalence du phénomène, l'estimation obtenue mélange les incapacités transitoires (à court terme) et les incapacités chroniques (à long terme) sans pouvoir les distinguer. De plus, les incapacités à court terme ont toutes les chances d'être sous estimées car une photo ponctuelle les capture moins efficacement. Cette sous estimation touche aussi, de façon variable, les causes des incapacités. En effet, le statut fonctionnel des personnes est la résultante, en dehors des facteurs environnementaux et comportementaux, de l'impact négatif d'affections médicales, parfois multiples, de sévérité variée. L'importance des affections induisant une incapacité brutale mais avec un taux de réversibilité élevé est alors probablement sous estimée. Par exemple, une lombalgie ou une fracture du col du fémur, pathologies très fréquentes, seront faiblement associées à l'incapacité car, au moment de l'interview, le sujet aura recouvré son autonomie fonctionnelle. Inversement des pathologies, moins fréquentes comme la maladie de Parkinson, provoquant en revanche des incapacités progressives et permanentes seront identifiées comme contribuant fortement à la dépendance.

Lors d'une enquête ponctuelle, la prévalence des incapacités dépendra du moment de l'évaluation au cours de l'année. En effet, certaines pathologies à l'origine d'incapacités réversibles ont un aspect saisonnier : par exemple, les pneumonies, les allergies, et même les infarctus. Le nombre d'épisodes transitoires non capturés dépendra donc de la saison à laquelle l'évaluation est faite.

Pour les études longitudinales, l'ampleur de la sous-estimation des épisodes d'incapacité transitoire est proportionnelle à l'intervalle de temps entre les évaluations successives et inversement proportionnelle à la longueur de l'épisode. De plus, du fait de la durée variable de l'état d'incapacité, le taux de réversibilité observé dépend de l'intervalle entre deux mesures. Si les mesures sont proches dans le temps, on recueillera davantage d'incapacités à court terme et le taux de réversibilité global sera plus élevé.

Ces différents aspects expliquent en grande partie pourquoi le taux de réversibilité varie autant d'une étude à l'autre. La plupart des études se basent sur des intervalles d'évaluation de l'incapacité de 12 à 24 mois et trouvent des taux de réversibilité de l'ordre de 25% à 30% [17, 18] mais lorsque les intervalles entre évaluations deviennent plus courts, les taux de réversibilité sont nettement plus élevés particulièrement chez les sujets entrés récemment en incapacité [19]. La prise en compte de la réversibilité dans la modélisation statistique est donc indispensable.

Une autre caractéristique très importante à prendre en compte lorsque l'on veut évaluer ou analyser la dépendance des personnes âgées est la fréquence élevée des censures. Dans une étude longitudinale, lorsque l'intervalle entre les évaluations est d'un an ou plus, une part de la sous-évaluation de l'incapacité est due aux décès ou aux perdus de vue [20]. En effet, une proportion importante de personnes est en incapacité seulement au cours des quelques mois qui précèdent le décès. Ces incapacités, pourtant très fréquentes dans ces âges élevés, risquent de ne pas être observées. De même, certaines personnes peuvent être perdues de vue de



manière temporaire ou définitive. Ces absences de participation sont souvent dues à des hospitalisations, des prises en charge en institutions, des affections intercurrentes, des états dépressifs, autant de conditions qui peuvent être associées à un état d'incapacité. Une classification des réponses manquantes en fonction de leur relation avec l'événement d'intérêt non observé sera détaillée plus loin (Partie II, paragraphe 1.6) et nous verrons que cette relation a de fortes implications sur le choix du modèle statistique.

Enfin, une question commune à tous les indicateurs concerne la définition du seuil à partir duquel le sujet est considéré en incapacité. Les indicateurs sont pour la plupart des variables ordinales comprenant plusieurs niveaux d'incapacité ou plusieurs stades, et selon les études le point de coupure varie. Ce choix du seuil va, bien sûr, influencer sur le taux de prévalence mais aussi sur le taux de réversibilité dans les études longitudinales. En effet, la réversibilité est plus fréquente pour des incapacités légères que pour des incapacités plus lourdes atteignant un plus grand nombre de champs d'activité.

La majorité des analyses étiologiques est limitée à des études transversales. Or, de façon générale en épidémiologie, les études longitudinales sont plus informatives. Elles permettent, dans la recherche de causalité entre un facteur de risque et un événement, de pouvoir mettre en évidence l'antériorité du facteur. Mais les études longitudinales sont plus délicates à analyser statistiquement. Dans notre domaine, les études transversales souffrent de limitations spécifiques, et en particulier les différences de proportions de personnes en incapacité entre les classes d'âge se confondent avec celles entre générations [21].

La réponse est une variable binaire ou ordinale, recueillie pour chaque sujet plusieurs fois à des temps différents. Certains auteurs ont analysé ces mesures répétées de l'incapacité en comparant deux à deux les temps d'évaluation. Cette méthode implique un grand nombre de

tests et conduit à des résultats partiels pour des périodes fixes. Les analyses statistiques doivent prendre en compte quatre caractéristiques essentielles des données longitudinales :

- le délai peut être une variable explicative,
- les observations répétées pour un sujet sont corrélées,
- les covariables peuvent être dépendantes du temps c'est-à-dire qu'elles peuvent varier au cours du temps pour un même sujet,
- les réponses manquantes de façon intermittente au cours du temps ou les sorties définitives d'étude peuvent introduire un biais.

Comme nous l'avons vu, l'incapacité n'est pas un évènement "absorbant" comme le serait, par exemple, le décès et il est fréquent que la personne retrouve son autonomie fonctionnelle. Aussi, les modèles de survie, analysant le délai jusqu'à la survenue d'un évènement tel que le modèle de Cox, ne rendent que très partiellement compte de la complexité du phénomène. Certains auteurs [21-24] ont utilisé une extension du modèle de Cox, les modèles multi-états qui modélisent l'occurrence de différents types d'évènements ou états entre lesquels on observe des transitions. L'effet des covariables sur ces transitions est ajusté sur l'histoire du processus étudié. En général, ces modèles sont destinés à des observations régulièrement espacées dans le temps sous l'hypothèse que le processus de transition n'est pas hétérogène selon les sujets [25]. Les estimations des intensités de transition sont parfois difficiles à interpréter.

Il nous a semblé intéressant de replacer l'étude de l'évolution de l'incapacité dans le cadre général des extensions des modèles logistiques pour données corrélées.

Avant de décrire les modèles que nous allons comparer dans la partie II, nous présentons la cohorte EPIDOS (Epidémiologie de l'ostéoporose) qui servira d'exemple tout au long de notre exposé.

## 4 Cohorte EPIDOS

Entre janvier 1992 et janvier 1994, 7575 femmes de 75 ans et plus ont été recrutées dans cinq centres en France : Amiens, Lyon, Montpellier, Paris et Toulouse. Les listes électorales ou les fichiers de l'assurance maladie ont servi de bases de recrutement, les femmes éligibles ont été alors contactées par courrier. Les femmes qui n'étaient pas autonomes à la marche ou qui avaient eu une prothèse de hanche bilatérale ont été exclues au moment de l'inclusion [26]. L'examen d'inclusion pratiqué dans un centre clinique comprenait un questionnaire, des examens cliniques et fonctionnels et une mesure de la densité minérale osseuse. Après leur inclusion, les femmes étaient contactées par courrier ou par téléphone tous les quatre mois pour recueillir les informations sur les chutes éventuelles et sur les fractures intervenues au cours de la période. Tous les ans, les femmes remplissaient un questionnaire plus complet qui détaillait les nouvelles hospitalisations, les nouveaux évènements de santé, les changements de poids, le changement d'habitat, les médicaments utilisés, la santé subjective. De plus, ce questionnaire comportait l'auto évaluation de leur incapacité grâce à deux indicateurs : sortir de chez soi sans aide (indicateur Colvez) [12] et les AIVQ. Ces deux indicateurs seront donc nos variables d'intérêt, analysées sous forme soit de réponse binaire soit de réponse ordinale pour les AIVQ. Pour chacune des huit activités de l'indicateur AIVQ (faire le ménage, la lessive, les courses, la cuisine, téléphoner, prendre les transports en commun, remplir entièrement seule ses papiers, prendre entièrement seule ses médicaments), le questionnaire proposait une réponse dichotomique : non, j'en suis incapable/ oui, j'en suis capable. Le score final représente le nombre d'activités qu'une femme se dit incapable de faire, ce score varie donc de zéro à huit.

Le protocole initial prévoyait un suivi de quatre ans pour tous les centres. Cependant le centre de Montpellier a poursuivi l'enquête pendant trois ans de plus. Nous analyserons donc les 7 ans de suivi de la sous-cohorte de ce centre qui comporte 1548 femmes à l'inclusion.

Pour tester nos modèles, nous allons utiliser des covariables d'inclusion, présumées facteurs de risque d'incapacité ainsi que des covariables dépendantes du temps, évaluées chaque année.

A l'inclusion, ces covariables comprennent des performances à des tests physiques standardisés :

- vitesse de marche, temps pour marcher sur une longueur de six mètres à l'allure habituelle,
- nombre de pas effectués sur ces 6 mètres,
- temps pour se lever d'une chaise et s'asseoir cinq fois, les bras croisés,
- temps pour taper 10 fois de suite avec la pointe de son pied en alternance, dans deux cercles placés à 30 centimètres de distance, en position assise,
- temps d'équilibre en position pieds tandem (un pied devant l'autre, le talon du premier touchant la pointe du second),
- temps d'équilibre en position pieds semi-tandem (pieds légèrement décalés),
- capacité à effectuer une marche funambule (faire 4 pas en mettant le talon du pied avant touchant la pointe du pied arrière),
- force de préhension manuelle,
- force du quadriceps.

Nous testerons aussi des variables issues de l'examen clinique initial : l'acuité visuelle (échelle de Snellen), la composition corporelle (poids, masse grasse, masse maigre, indice de masse corporelle ( $\text{poids}(\text{kg})/\text{taille}^2(\text{mètre})$ ), le rapport du tour de taille sur le tour de hanches), la diminution de la taille depuis l'âge de 30 ans, l'activité physique actuelle. Nous utiliserons aussi plusieurs variables issues du questionnaire d'inclusion : le niveau d'éducation,

l'isolement social (vivre seule ou pas), l'état de santé subjectif, l'état de santé comparé aux personnes du même âge, la peur de tomber, les douleurs affectant le bas du dos ou les membres inférieurs.

Enfin, les variables dépendantes du temps, recueillies dans l'auto questionnaire annuel, seront : avoir été hospitalisée au cours de l'année écoulée, être restée confinée au lit temporairement au cours de la dernière année et le nombre de chutes déclarées au cours de l'année.

Plusieurs autres variables d'inclusion ont été testées mais ne sont pas présentées dans ce travail.

Rappelons que l'objet de cette thèse n'est pas l'analyse épidémiologique des données d'EPIDOS qui est faite par ailleurs, mais de s'intéresser aux problèmes de méthodologie statistique qui sont posés.

## Partie II : Réponses binaires

# 1 Méthodes d'analyse des réponses binaires corrélées

Dans un premier temps, considérons une réponse binaire : être ou ne pas être en incapacité. Pour chaque sujet, l'incapacité est évaluée à plusieurs reprises au cours du temps ce qui constitue une suite individuelle de réponses. Ces réponses sont appelées des réponses groupées, chaque sujet constituant un groupe. Plus généralement, on parlera aussi de réponses groupées pour des observations faites sur les membres d'une même famille, le groupe est alors la famille, ou pour des observations faites sur des organes multiples d'un patient (reins, yeux, ...) ou sur des lésions multiples (brûlures, tumeurs, ...), le groupe est alors le patient. Comparativement à celles provenant de deux groupes différents, les réponses à l'intérieur d'un groupe sont davantage semblables, elles sont corrélées. On utilisera alors le terme de réponses corrélées, synonyme pour nous de réponses groupées.

Les groupes sont généralement plus homogènes, et la variabilité entre groupes est plus grande, que ce que l'on observerait dans le cas non corrélé. Dans notre exemple, les états d'incapacité observés à un an d'intervalle sur une femme sont corrélés entre eux. Ne pas tenir compte de ces corrélations intra-groupe conduit, en général, à une sous estimation de la variance qui risque de faire apparaître des significations à tort.

Quand les réponses corrélées sont continues et considérées comme approximativement gaussiennes, une classe de modèles linéaires est maintenant largement utilisée : les modèles linéaires mixtes (LMM pour Linear Mixed Model). Quand les réponses corrélées sont binaires ou nominales, l'extension naturelle correspondante est la classe des modèles mixtes linéaires généralisés (GLMM pour Generalized Linear Mixed Model), mais des problèmes techniques apparaissent. Cela est principalement dû à l'absence de distribution discrète analogue à la distribution multinormale.

Une autre approche a été développée, il s'agit de l'approche marginale avec la formulation d'équations d'estimation généralisées (GEE pour Generalized Estimating Equations ou GEE1). Même si le modèle mixte permet de mieux caractériser l'hétérogénéité entre groupes, dans le cas de réponses corrélées, continues et approximativement normales, les deux approches, modèle mixte ou modèle marginal, aboutissent à des interprétations semblables alors que pour des réponses discrètes ce n'est pas le cas. Il est donc important pour l'épidémiologiste ou le statisticien d'accorder beaucoup d'attention au choix entre ces deux familles de modèles en fonction des objectifs de l'analyse et des possibles sources de corrélation.

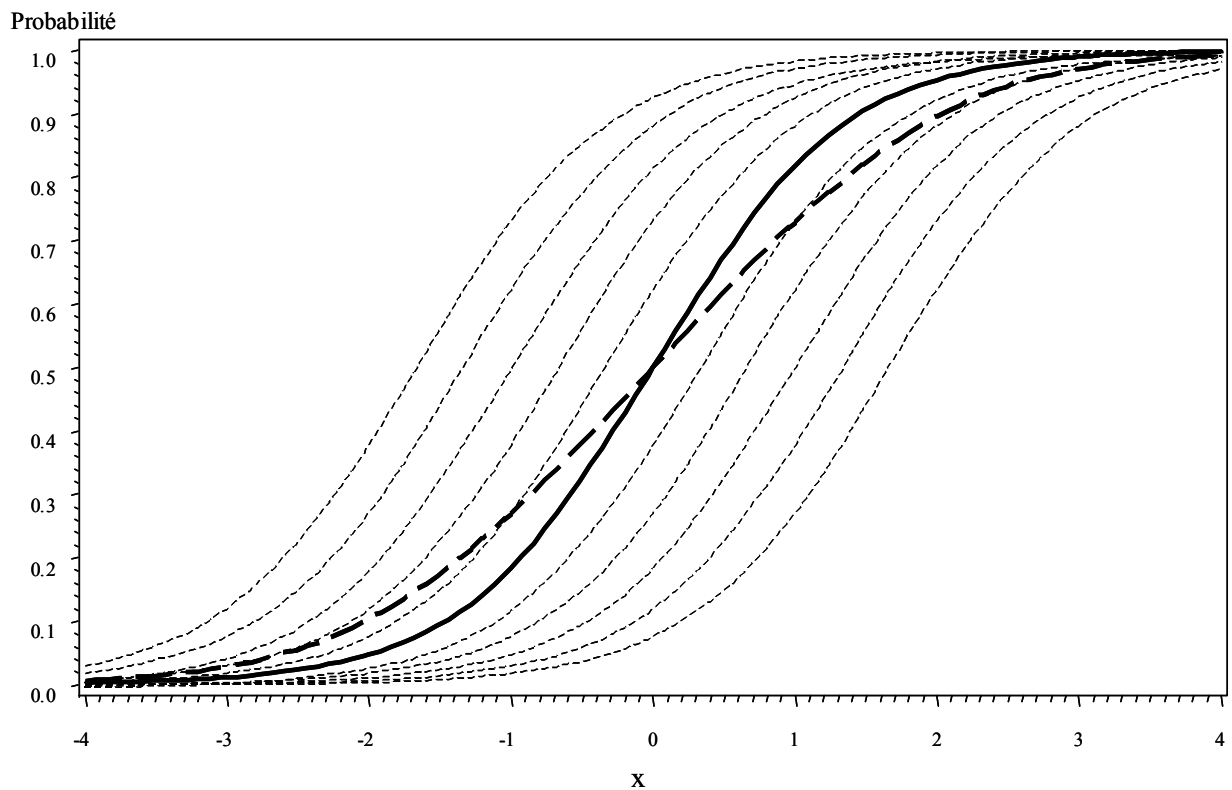
Dans ce qui suit, nous allons décrire de manière plus détaillée ces deux types de modèles en nous intéressant plus particulièrement aux données longitudinales binaires.

Le modèle marginal met en relation les covariables directement avec la probabilité marginale de la réponse. Par probabilité marginale, on entend la réponse moyenne de la sous population qui présente des valeurs communes pour les covariables du modèle. Le modèle mixte (ou modèle à effets aléatoires) prend explicitement en compte l'hétérogénéité entre individus en décrivant comment une réponse individuelle dépend de l'historique du sujet : les réponses précédentes, les covariables initiales, éventuellement les covariables individuelles dépendantes du temps. Le modèle marginal analyse une séquence de moyennes de réponses c'est-à-dire la trajectoire des moyennes, alors que le modèle à effets aléatoires étudie les trajectoires individuelles et, pour les effets fixes, donne leur relation avec la moyenne des trajectoires individuelles. Pour illustrer ce propos, nous présentons dans le graphique ci-dessous (figure 1), une simulation de la probabilité d'une réponse positive dans un modèle logistique en fonction d'une covariable continue,  $X$  [1]. Chaque individu est représenté par une fine ligne en pointillés. La probabilité suit une fonction logistique avec un coefficient en  $X$  fixe mais une ordonnée à l'origine aléatoire, c'est-à-dire qui varie d'un sujet à l'autre. Pour le sujet  $i$ , on a donc :



$$P(x)_i = \frac{1}{1 + \exp-(\alpha_i + \beta x)}$$

Les probabilités individuelles ont la même forme mais sont décalées selon les valeurs de l'ordonnée à l'origine aléatoire. La ligne pleine symbolise la courbe pour un individu avec un effet aléatoire moyen. Elle correspond à ce qu'estime le modèle à effets aléatoires pour la covariable fixe, X. Il s'agit de la courbe pour un individu typique. A contrario, la courbe en pointillés épais, est la courbe des moyennes de probabilités pour chaque valeur de X. Cette dernière courbe correspond au modèle marginal. La pente diffère des courbes individuelles, elle est atténuée. Ce résultat est général quand les probabilités ne sont pas linéaires.



**Figure 4 : Simulation de la probabilité d'une réponse positive dans un modèle logistique avec une ordonnée à l'origine aléatoire en fonction d'une covariable continue**

Il apparaît dans cet exemple simple, que l'approche par un modèle marginal diffère de celle d'un modèle mixte, correspondant à des objectifs différents. En effet, le modèle marginal exprime des relations moyennées pour chaque valeur de X, alors que le modèle à effets

aléatoires fournit des associations conditionnellement au fait d'avoir certaines caractéristiques individuelles modélisées par les effets aléatoires.

De plus, il faut noter que ces deux modèles prennent en compte de manière différente la dépendance entre les observations d'un même sujet. Cette caractéristique est capitale dans les études longitudinales. Le modèle marginal modélise séparément la corrélation intra-sujet des réponses et la régression sur les variables explicatives tandis que dans le modèle à effets aléatoires, cette corrélation est prise en compte au moyen des effets aléatoires individuels et elle est modélisée conjointement avec les variables explicatives.

Nous allons, dans ce qui suit, revenir sur le modèle linéaire généralisé, puis détailler les deux types de modèles, modèle marginal et modèle mixte, pour une réponse binaire en utilisant une régression logistique.

## 1.1 Modèle linéaire généralisé

Dans le cadre des données indépendantes, les modèles linéaires généralisés (GLM pour generalized linear model) permettent une extension des modèles linéaires classiques aux réponses discrètes [27].

Si l'on considère un modèle linéaire classique, notons  $n$  le nombre d'observations et  $\mathbf{Y}$  le vecteur des réalisations  $(Y_i)$ ,  $i=1, n$  de la variable aléatoire  $\mathbf{Y}$  dont les composants sont indépendants et de moyenne  $\boldsymbol{\mu}$ . La partie systématique du modèle est le  $(n \times 1)$  vecteur  $\boldsymbol{\mu}$  qui prend alors la forme:

$$\boldsymbol{\mu} = \sum_{l=1}^p \mathbf{X}_l \beta_l$$

Les  $\mathbf{X}_l$  sont les  $(n \times 1)$  vecteurs des  $p$  variables explicatives,  $\beta_l$  sont les paramètres à estimer à partir des données. Si nous considérons les  $n$  observations indexées par  $i$  alors la partie systématique du modèle peut être écrite :

$$E(Y_i) = \mu_i = \sum_{l=1}^p X_{il} \beta_l ; i=1, \dots, n$$

$X_{il}$  est la valeur de la covariable  $l$  pour l'observation  $i$ .

Si nous utilisons la notation matricielle,  $\boldsymbol{\mu}$  est un vecteur ( $n \times 1$ ),  $\mathbf{X}$  est la matrice ( $n \times p$ ) et  $\boldsymbol{\beta}$  est le vecteur ( $p \times 1$ ), le modèle linéaire classique s'écrit alors :

$$\boldsymbol{\mu} = \mathbf{X} \boldsymbol{\beta}$$

La partie aléatoire est constituée des erreurs indépendantes et de variance constante  $\sigma^2$ .

Le modèle linéaire généralisé permet deux types d'extension : la distribution de la partie aléatoire,  $\mathbf{Y}$ , appartient à la famille exponentielle et la fonction de lien qui relie le prédicteur linéaire,  $\boldsymbol{\eta} = \mathbf{X} \boldsymbol{\beta}$ , à l'espérance  $\boldsymbol{\mu}$  peut être n'importe quelle fonction monotone, différentiable.

Dans le cas du modèle linéaire classique, nous avons vu que cette fonction était l'identité, en effet  $\boldsymbol{\eta}$  et  $\boldsymbol{\mu}$  peuvent prendre n'importe quelle valeur réelle et le lien identité est possible. Par contre, pour une distribution binomiale,  $\boldsymbol{\mu}$  est comprise entre 0 et 1 car il s'agit de la probabilité que  $\mathbf{Y}$  soit égal à 1. La fonction de lien doit permettre de transformer cet intervalle (0,1) vers l'ensemble des réels. Une fonction de lien très courante en épidémiologie est la fonction de lien Logit, pour laquelle nous avons alors :

$$\text{logit}(\boldsymbol{\mu}) = \log \left\{ \frac{\boldsymbol{\mu}}{1 - \boldsymbol{\mu}} \right\} = \mathbf{X} \boldsymbol{\beta}$$

Les paramètres  $\boldsymbol{\beta}$  sont estimés par la méthode du maximum de vraisemblance. On a alors un système de  $p$  équations en  $\boldsymbol{\beta}$ ,  $p$  étant le nombre de covariables dans le modèle. Leur résolution requiert des méthodes itératives dans lesquelles interviennent le hessien pour l'algorithme de Newton-Raphson ou la matrice d'information pour les scores de Fisher. A chaque itération, l'algorithme calcule une nouvelle approximation de  $\boldsymbol{\beta}$ . L'approximation finale correspond à celle pour laquelle la convergence est atteinte.

Dans le cas où la fonction de lien du modèle linéaire généralisé utilisée est la fonction canonique associée à la distribution de la partie aléatoire (logit pour une distribution binomiale), plusieurs simplifications interviennent. On montre que le hessien est égal à la matrice d'information et donc les méthodes de résolution du score de Fisher et de Newton-Raphson coïncident [27].

Cette estimation peut être mise en œuvre très facilement dans la plupart des logiciels de statistique.

## 1.2 Qualité d'ajustement et tests de comparaison des modèles

Une fois l'estimation des paramètres d'un modèle multivarié faite, vient la question de l'évaluation de la qualité de ce modèle. Cette évaluation peut se faire soit de manière absolue en confrontant le modèle calculé aux données observées, soit de manière relative en comparant le modèle en cours à un autre modèle candidat.

Dans le premier cas, il s'agit d'évaluer la qualité d'ajustement du modèle sur la base des différences entre les observations et les estimations. Il existe deux importantes mesures d'adéquation du modèle aux données : la statistique de vraisemblance ou déviance et la statistique de Pearson [28].

Notons  $L(\boldsymbol{\mu}; \mathbf{y})$  le logarithme de la fonction de vraisemblance exprimée selon la moyenne  $\boldsymbol{\mu}=(\mu_1, \mu_2, \dots, \mu_n)'$  pour un vecteur de données  $\mathbf{y}=(y_1, y_2, \dots, y_n)'$ . De même notons  $L(\hat{\boldsymbol{\mu}}; \mathbf{y})$  le maximum du log de la vraisemblance obtenu pour un modèle donné avec  $p$  paramètres et  $L(\mathbf{y}; \mathbf{y})$  le maximum du log de la vraisemblance pour le modèle saturé. Alors la déviance du GLM est définie par :

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = -2[L(\hat{\boldsymbol{\mu}}; \mathbf{y}) - L(\mathbf{y}; \mathbf{y})]$$

Ce qui est équivalent au carré du rapport des vraisemblances. On montre qu'asymptotiquement,  $D$  suit une loi du  $\chi^2$  à  $(n-p)$  degrés de liberté. Le nombre de degrés de liberté est la différence entre le nombre d'observations et le nombre de paramètres du modèle. Ceci permet de construire un test de rejet ou d'acceptation du modèle.

Un test du  $\chi^2$  est également utilisé pour comparer les valeurs observées,  $y_i$ , à leur prévision par le modèle. La statistique du test de Pearson est définie par :

$$X^2 = \sum_i (y_i - \hat{\mu}_i)^2 / \text{Var}(\hat{\mu}_i)$$

$X^2$  suit aussi asymptotiquement une distribution du  $\chi^2$ .

Lorsque le but est de comparer deux modèles entre eux, pour aider au choix entre plusieurs modèles emboîtés candidats, deux critères sont habituellement proposés : le rapport de vraisemblance et le test de Wald.

Le rapport de vraisemblance ou la différence des déviances est une évaluation de l'apport des variables explicatives supplémentaires dans l'ajustement du modèle. La différence des déviances entre deux modèles emboîtés avec respectivement  $p_1$  et  $p_2$  variables explicatives suit approximativement une loi du  $\chi^2$  à  $(p_2-p_1)$  degrés de liberté pour les lois à un paramètre (binomiale, Poisson) et une loi de Fisher pour les lois à deux paramètres (gaussienne). Ceci permet de tester la significativité de la diminution de la déviance engendrée par l'ajout de variables explicatives ou la prise en compte d'interactions.

Le test de Wald est basé sur la forme quadratique faisant intervenir la matrice de covariance des paramètres. Si la matrice  $\mathbf{L}$ , dite contraste, définit l'ensemble des hypothèses à tester sur les paramètres :  $\mathbf{L}'\boldsymbol{\beta} = 0$

Alors la statistique de Wald s'écrit :

$$S = (\mathbf{L}' \hat{\boldsymbol{\beta}})' (\mathbf{L}' \hat{\boldsymbol{\Sigma}} \mathbf{L})^{-1} (\mathbf{L}' \hat{\boldsymbol{\beta}})$$

Avec  $\hat{\boldsymbol{\beta}}$ , l'estimateur du maximum de vraisemblance et  $\hat{\boldsymbol{\Sigma}}$  la matrice de covariance estimée.

La distribution asymptotique de  $S$  est une loi du  $\chi^2$  avec comme nombre de degrés de liberté le rang de la matrice  $\mathbf{L}$ .

Ce test peut ne pas être correct lorsque le nombre d'observations est faible.

On peut aussi utiliser le critère d'information de Akaike (AIC pour Akaike Information Criterion) ou le critère de Schwarz (SC pour Schwarz Criterion ou BIC pour Bayesian Information Criterion). Ces critères soulignent le principe de parcimonie des paramètres en pénalisant la vraisemblance quand le nombre de paramètres augmente.

$$AIC = -2 \log(L(\hat{\boldsymbol{\mu}}; y)) + 2p$$

$$SC = -2 \log(L(\hat{\boldsymbol{\mu}}; y)) + p \log(n)$$

Plus la valeur de ces deux critères est faible, plus le modèle est intéressant.

### 1.3 Modèle marginal

Dans le cas des réponses longitudinales, corrélées, pour un sujet,  $i$ , la variable que nous voulons modéliser, est constituée de plusieurs observations de l'incapacité à des temps différents. Notons  $Y_{ij}$ , la  $j^{\text{ème}}$  réponse du sujet  $i$ , observée au temps  $t$ .

Le modèle marginal, appelé aussi modèle moyenné sur la population, suppose que la relation entre le vecteur des réponses,  $\mathbf{Y}$ , et la matrice des covariables,  $\mathbf{X}$ , est la même pour tous les sujets. L'équation du modèle marginal s'écrit alors de la façon suivante :

$$\text{Logit}(E(Y_{ij} | \mathbf{X}_{ij})) = \text{Logit}(P(Y_{ij} = 1 | \mathbf{X}_{ij})) = \mathbf{X}'_{ij} \boldsymbol{\beta} = \mu_{ij}$$

avec  $\mathbf{X}_{ij}$  le vecteur des  $p$  covariables observées au moment de la  $j^{\text{ème}}$  réponse pour le sujet  $i$ .

Le vecteur des paramètres,  $\boldsymbol{\beta}$ , est estimé grâce aux équations d'estimation généralisées (GEE) proposé par Liang et Zeger [29, 30]. Ce système d'équations est une extension au cas corrélé des équations d'estimation pour observations indépendantes.

$$\sum_{i=1}^n \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = 0$$

avec :  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})'$ ,

$\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{in_i})'$ ,

$\mathbf{V}_i = \text{var}(\mathbf{Y}_i) = \phi \mathbf{A}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$  est la matrice de covariance qui décrit la dépendance entre les observations,

$\mathbf{A}_i$  est une matrice ( $n_i \times n_i$ ) diagonale avec  $\phi \mu_{ij} (1 - \mu_{ij})$  comme  $j^{\text{ème}}$  élément diagonal,

$\mathbf{R}(\boldsymbol{\alpha})$  est une matrice de corrélation de travail.

Cette matrice est entièrement spécifiée par le vecteur des paramètres  $\boldsymbol{\alpha}$ . Cette corrélation de travail est supposée identique pour tous les sujets et reflète une dépendance moyenne entre les observations répétées des sujets. Si  $\mathbf{R}(\boldsymbol{\alpha}) = \mathbf{I}$ , la matrice identité, le GEE se réduit au cas des observations indépendantes. Sinon, plusieurs choix sont possibles pour la structure de cette matrice selon les hypothèses que l'on peut faire sur la nature des liens entre les observations.

$\mathbf{R}(\boldsymbol{\alpha})$  est appelée  $m$ -dépendante si  $\text{corr}(Y_{ij}, Y_{ij+r}) = \alpha_r$  pour  $r = 1, 2, \dots, m$  et

$\text{corr}(Y_{ij}, Y_{ij+r}) = 0$  pour  $r > m$ ,  $\mathbf{R}(\boldsymbol{\alpha})$  est dite échangeable si  $\text{corr}(Y_{ij}, Y_{ik}) = \alpha$  pour  $j \neq k$ ,  $\mathbf{R}(\boldsymbol{\alpha})$

est dite non structurée si  $\text{corr}(Y_{ij}, Y_{ik}) = \alpha_{ik}$  pour  $j \neq k$  et  $\mathbf{R}(\boldsymbol{\alpha})$  est autorégressive si

$\text{corr}(Y_{ij}, Y_{ij+r}) = \alpha^r$  pour  $r = 0, 1, \dots, n_i - j$ . Il faut remarquer que le délai séparant deux réponses

n'est pas pris en compte dans ces corrélations (car le délai réel entre les observations  $j$  et  $j+r$  peut varier d'un sujet à l'autre). Liang et Zeger ont proposé un algorithme modifié du score de

Fisher pour résoudre le système d'équations d'estimation. Cet algorithme consiste en les étapes suivantes :

- Calcul du vecteur des paramètres  $\beta$ , comme valeurs initiales des itérations, avec un modèle linéaire généralisé ordinaire sans dépendance entre les observations,
- Calcul de la matrice de corrélation de travail  $R$  à l'aide des résidus des valeurs estimées de  $\beta$  en tenant compte de la structure de cette matrice,
- Calcul de la matrice globale de covariance :  $V_i = \phi A_i^{1/2} \hat{R}(\alpha) A_i^{1/2}$ ,
- Calcul de  $\beta$  à partir des estimations de l'itération précédente,

$$\beta^{(r+1)} = \beta^{(r)} + \left[ \sum_{i=1}^n \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right]^{-1} \left[ \sum_{i=1}^n \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} (Y_i - \mu_i) \right]$$

- Répétition des étapes 2 à 4 jusqu'à la convergence.

Deux estimateurs de la variance des  $\hat{\beta}$  ont été proposés : l'estimateur basé sur le modèle et l'estimateur robuste. La variance robuste est peu sensible à la spécification de la matrice de travail. Elle corrige, en quelque sorte, la variance des estimateurs quand la matrice de variance-covariance est mal spécifiée.

Le résultat important, démontré par Liang et Zeger [29] et qui explique la popularité du GEE, est que les estimateurs des paramètres et les estimateurs robustes de leur variance sont consistants c'est-à-dire que pour chaque estimateur, sa variance ainsi que son biais tendent asymptotiquement vers zéro, même si la structure de corrélation (c'est-à-dire  $R(\alpha)$ ) n'est pas bien spécifiée.

Cependant il s'agit d'un résultat asymptotique et le choix d'une structure de corrélation la plus appropriée possible, améliore l'efficacité des estimateurs [25]. Il est donc recommandé de spécifier la structure de corrélation le mieux possible en se fondant sur l'observation de l'évolution longitudinale des réponses.



## 1.4 Modèle mixte linéaire généralisé

Le modèle mixte appelé aussi modèle à effets aléatoires ou bien modèle spécifique du sujet, suppose que les relations entre les covariables et la survenue de l'événement,  $Y$ , diffèrent d'un sujet à l'autre. Les paramètres,  $\beta$ , associés aux covariables sont donc dépendants du sujet. Le modèle s'écrit de la manière suivante :

$$\text{logit}\left(P\left(Y_{ij}=1|\mathbf{X}_{ij}\right)\right)=\mathbf{X}_{ij}'\beta_i$$

Cette écriture peut se décomposer de la façon suivante :

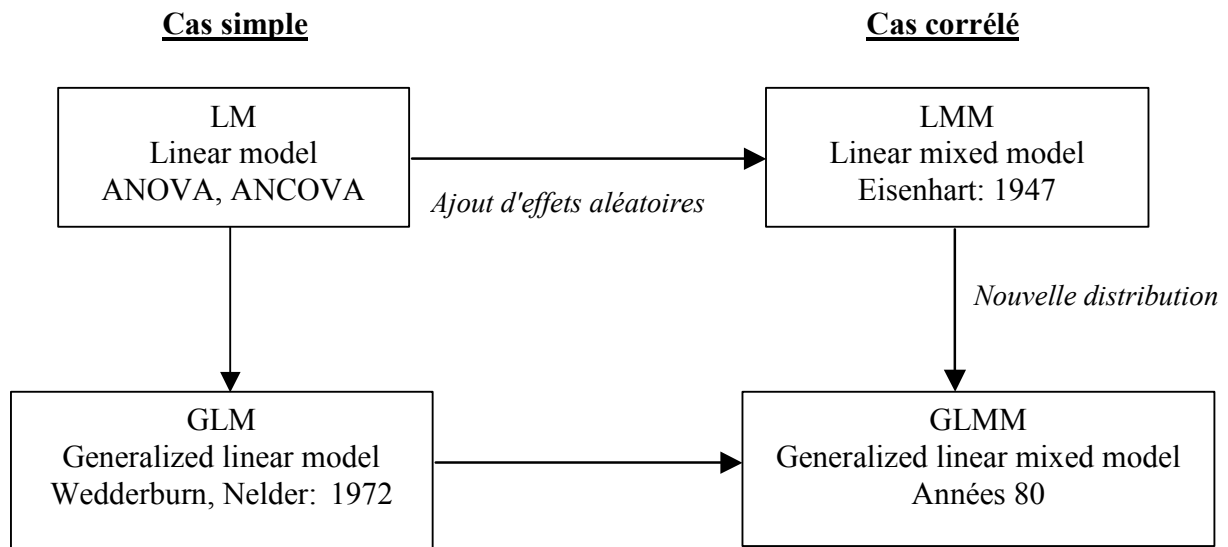
$$\text{logit}\left(E\left(Y_{ij} \mid \mathbf{X}_{ij}, \mathbf{U}_i\right)\right)=\mathbf{X}_{ij}'\left(\beta^* + \mathbf{U}_i\right)=\mathbf{X}_{ij}'\beta^* + \mathbf{X}_{ij}'\mathbf{U}_i$$

$$\text{logit}\left(E\left(Y_{ij} \mid \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{U}_i\right)\right)=\mathbf{X}_{ij}'\beta^* + \mathbf{Z}_{ij}'\mathbf{U}_i$$

Les effets aléatoires,  $\mathbf{U}_i$ , sont supposés indépendants d'un sujet à un autre, mais distribués selon une loi commune qui est en général une loi normale de moyenne nulle et de variance  $\mathbf{D}$ .

On parle alors d'un modèle "logistique-normal".  $\mathbf{Z}_{ij}$  est un sous vecteur de  $\mathbf{X}_{ij}$  ; les effets aléatoires concernent donc seulement une partie des covariables et l'ordonnée à l'origine. La variance  $\mathbf{D}$ , doit être estimée et représente l'amplitude de la variabilité inter-individus, non expliquée.

En résumé, comme indiqué dans le schéma ci-dessous, par rapport au modèle linéaire simple, le GLMM est le produit d'une double extension. La première est une extension en termes de loi avec la prise en compte d'une famille de distributions plus large que la loi normale, ce qui a donné naissance au GLM. La deuxième est une extension en termes d'introduction d'effets aléatoires, ce qui a abouti à la classe des modèles linéaires mixtes (LMM) d'une part, et au GLMM, d'autre part.



**Figure 5 : Extensions du modèle linéaire aux données binaires groupées**

Un point délicat du modèle GLMM est l'estimation des paramètres car la distribution normale des effets aléatoires n'est pas la conjuguée de la distribution binomiale des réponses (la conjuguée serait une distribution bêta).

Plusieurs procédures d'estimation ont été proposées pour le GLMM. Nous nous intéressons d'abord aux méthodes qui maximisent directement une approximation de la vraisemblance. Dans notre cas de modèle logistique à effets aléatoires gaussiens, si nous considérons seulement une ordonnée à l'origine aléatoire ( $Z_i = 1$ ), correspondant à un paramètre individuel, constant au cours du temps, révélateur d'une propension à être en incapacité, les paramètres à estimer sont d'une part le vecteur  $\beta$  et d'autre part la variance du scalaire  $U_i$ ,  $\sigma_u^2$ .

La vraisemblance exprimée en fonction des paramètres inconnus s'écrit alors :

$$L(\beta^*, \sigma_u; y) = \prod_{i=1}^n \int f(y_i | u_i) f(u_i; \sigma_u) du_i .$$

Plus généralement, la vraisemblance est l'intégrale sur les effets aléatoires non observés de la distribution conjointe des données et des effets aléatoires. Dans le cas des modèles non linéaires, cette intégrale n'a pas de forme simple et des méthodes numériques d'intégration sont nécessaires telles que les méthodes de quadrature gaussienne ou de simulation Monte

Carlo par chaînes de Markov (MCMC pour Markov Chain Monte Carlo). Ces méthodes d'intégration, très utilisées en sciences physiques, demandent parfois de longs temps de calculs. De nombreux algorithmes ont été proposés et sont toujours en cours d'amélioration.

Dans les logiciels statistiques tels que SAS avec la procédure NLMIXED ou STATA avec le programme GLLAMM, c'est la méthode de la quadrature de Gauss–l'Hermite [31] qui est utilisée. Il s'agit d'une extension sophistiquée de la méthode des trapèzes utilisée pour calculer une intégrale simple. La fonction à calculer est remplacée par une série de polynômes. Les méthodes de Gauss ajoutent l'utilisation d'abscisses non régulièrement espacées et de pondérations sur les intervalles. La formule d'intégration de Gauss à N points devient exacte jusqu'à des polynômes de degré  $2N-1$ , ce qui augmente la précision de l'évaluation sans qu'il soit nécessaire d'augmenter le nombre de points à calculer. Une fois le calcul de la vraisemblance fait, des méthodes standard de maximisation, telles que celle de Newton-Raphson, permettent d'obtenir les estimations des paramètres [28].

Cette méthode présente certaines limites. Les calculs peuvent parfois être très longs du fait de la nécessité d'utiliser un grand nombre de points dans le calcul de l'intégrale avant que l'approximation ne soit raisonnable. La durée des calculs dépend en particulier fortement du nombre d'effets aléatoires introduits dans le modèle. Cette méthode limite donc le choix de la structure de covariance dans un modèle, c'est-à-dire le choix des covariables auxquelles on adjoint un effet aléatoire. Il est difficile voire impossible d'introduire des structures complexes avec plusieurs effets aléatoires corrélés entre eux.

Une méthode alternative est alors celle des simulations MCMC (pour Markov Chain Monte Carlo) [32]. Les algorithmes MCMC sont définis dans le cadre bayésien. Dans cette approche, les paramètres sont traités comme des variables aléatoires et interprétés en termes de degrés de certitude ou d'incertitude selon leurs valeurs. Partant d'une loi a priori pour les paramètres,

une série d'échantillons est simulée, définis sur les valeurs possibles des paramètres, et ayant pour loi stationnaire la loi recherchée, à savoir la loi a posteriori des paramètres.

À chaque étape, chaque paramètre est simulé selon sa loi conditionnelle courante sachant les observations. Le principe fondateur des méthodes d'estimation de Monte Carlo est d'approcher les estimateurs d'après une série d'échantillons. Néanmoins, en présence de lois de probabilité non standard et multidimensionnelles difficilement simulables, un moyen de produire des échantillons simulés suivant une loi d'intérêt donnée est d'utiliser une méthode de simulation par chaînes de Markov. Le principe d'une telle méthode de simulation est de générer une suite d'échantillons de manière récursive :  $\theta_r$  suit une loi  $q(\theta_r | \theta_{r-1})$  avec une loi  $q(\cdot | \cdot)$  aisément simulable, de telle manière que les échantillons  $\theta_r$  soient asymptotiquement distribués suivant la distribution d'intérêt. Cette distribution de transition doit vérifier certaines conditions pour qu'un tel schéma de simulation, produisant par définition des échantillons corrélés, puisse converger et générer des variables aléatoires suivant une loi invariante. Cette méthode d'estimation est implémentée dans le logiciel BUGS (Bayesian inference Using Gibbs Sampling). Comme l'indique son titre, ce logiciel utilise la méthode d'échantillonnage de Gibbs [33] pour construire une chaîne de Markov dont les valeurs convergent vers la distribution d'intérêt. L'idée de base est que seules des distributions conditionnelles univariées sont considérées, c'est-à-dire quand toutes les variables aléatoires sont fixées, excepté une. Cette situation est plus simple à simuler que des distributions jointes complexes. Pour simplifier, si on considère le cas bivarié où on cherche à estimer  $\theta = (\phi, \psi)$  l'algorithme est le suivant :

- 1- initialisation à  $\phi^{(0)}$
- 2- simulation de  $\psi^{(r)}$  à partir de  $p(\psi | \phi^{(r-1)}, y)$
- 3- simulation de  $\phi^{(r)}$  à partir de  $p(\phi | \psi^{(r)}, y)$

L'avantage des méthodes MCMC est leur souplesse et leur efficacité sur de petits échantillons, mais leur mise en œuvre demande un certain savoir-faire en particulier dans le choix des distributions a priori des paramètres et dans la détermination du moment où la série de simulations atteint son régime stationnaire [34].

Pour contourner la difficulté que constitue l'intégration de la vraisemblance, une autre stratégie consiste à utiliser des méthodes approchées. Breslow et Clayton par exemple, ont introduit une pseudo-vraisemblance ou une pseudo-vraisemblance restreinte [35, 36] sur un modèle approché pour les réponses moyennes. La procédure comporte trois étapes :

1- Calcul d'une formule approchée de la variable dépendante grâce à un développement de Taylor d'ordre un :  $\nu = \text{logit}(\hat{\mu}) + (y - \hat{\mu})\text{logit}'(\hat{\mu})$  en considérant  $\hat{\beta}^*$  et  $\hat{U}_i$  connus.

2-  $\nu$  a une distribution gaussienne et la matrice de variance  $\mathbf{D}$ , est estimée grâce à un modèle linéaire mixte pondéré, en maximisant la pseudo-vraisemblance ou la pseudo-vraisemblance restreinte.

3- Nouvelle estimation de  $\hat{\beta}^*$  et  $\hat{U}_i$ .

Ces trois étapes sont répétées jusqu'à l'obtention de la convergence.

Cette approche est implémentée dans SAS avec la macro GLIMMIX, mais elle comporte des inconvénients. En particulier, Rodriguez et Goldman ont démontré qu'elle donnait des estimateurs très biaisés dans le cas de modèles mixtes pour réponses binaires [37, 38]. Leurs simulations révèlent qu'à la fois les effets fixes et les composants de la variance présentent un biais d'atténuation important dans certains cas.

## 1.5 Effets fixes et effets aléatoires

Une importante question lorsque l'on utilise les modèles mixtes est le choix des effets aléatoires. De manière générale, on modélise les observations individuelles à l'intérieur de chaque groupe. Les effets fixes dépendent du plan d'expérience et donc du protocole ; si un nouvel échantillon était constitué, ces effets fixes seraient les mêmes. Les effets aléatoires, par contre, dépendent de l'échantillon, on fait seulement l'hypothèse qu'ils sont distribués selon une même loi. Dans les études longitudinales, l'introduction d'effets aléatoires est principalement destinée à prendre en compte l'hétérogénéité entre sujets [39]. Pour cela on introduit des ordonnées à l'origine aléatoires qui prennent en compte les différences entre les sujets, constantes dans le temps, sur les probabilités de réponses positives, différences qui ne sont pas expliquées par les covariables disponibles. Dans notre exemple sur l'incapacité, ces effets aléatoires individuels peuvent être interprétés comme une "fragilité" inhérente au sujet qui induit une plus grande propension à être en incapacité. On peut aussi introduire un effet-temps aléatoire. Cela revient à autoriser qu'en plus de cette fragilité individuelle constante dans le temps, il existe une fragilité qui augmente ou diminue avec le temps.

La définition des effets aléatoires détermine dans un modèle mixte la structure des corrélations entre observations. Dans le modèle marginal décrit plus haut, les estimations des paramètres sont asymptotiquement valides quelles que soient les hypothèses faites sur la structure de covariance. Dans un modèle à effets aléatoires, les paramètres fixes et aléatoires sont simultanément estimés. Le choix d'une structure de covariance influence les estimations des effets fixes et réciproquement l'introduction de nouvelles covariables fixes peut influencer l'estimation de la covariance. Pour un modèle mixte appliqué à des réponses gaussiennes, Littell et al [40] recommandent de considérer d'abord le modèle le plus général avec toutes les covariables pertinentes, de spécifier au mieux la structure de covariance (c'est-à-dire les effets aléatoires), d'estimer les paramètres et enfin d'essayer de réduire la partie fixe.

On peut considérer que des méthodes telles que les méthodes de quadrature gaussienne ou de simulation MCMC fournissent des estimateurs du maximum de vraisemblance [41] c'est-à-dire que les approximations convergent vers les estimateurs du maximum de vraisemblance quand le nombre de points de quadrature augmente pour l'intégration numérique ou quand la taille d'échantillon de Monte Carlo augmente pour la méthode MCMC. (Ceci n'est par contre, pas vrai pour les méthodes approchées comme celles basées sur la pseudo-vraisemblance.) Les estimateurs des effets fixes,  $\beta^*$ , ont les propriétés de normalité asymptotique des estimateurs du maximum de vraisemblance et l'on peut utiliser les tests asymptotiques de rapport des vraisemblances pour comparer des modèles se différenciant uniquement par l'ajout ou la suppression d'effets fixes. Par contre, pour les éléments de la covariance il est difficile de calculer la distribution asymptotique et donc de tester l'égalité à zéro de ces éléments.

En pratique, comme nous l'avons vu, les algorithmes d'estimation ne permettent en général pas d'introduire une structure de covariance très complexe.

## 1.6 Données manquantes

En épidémiologie, il peut arriver que les données des variables réponses ou des variables explicatives soient incomplètes. Nous nous intéresserons uniquement aux données incomplètes pour la variable réponse. Ces données peuvent être manquantes par intermittence ou de manière définitive (perdus de vue par exemple). Little et Rubin [42] ont proposé une classification pour définir le mécanisme du processus de données manquantes.

Ce processus de données manquantes peut être :

- complètement aléatoire (MCAR pour Missing Completely at Random) si la probabilité de réponse ne dépend pas des caractéristiques individuelles de l'individu (observées ou non),

- aléatoire (MAR pour Missing At Random) si la probabilité de réponse dépend seulement des données observées (réponses ou covariables observées antérieurement),
- informatif (ou non aléatoire ou MNAR pour Missing Not At Random ou non-ignorable) si le processus de réponse dépend des données non observées.

Dans le cas particulier des perdus de vue dans les études longitudinales, Diggle et Kenward [43] ont suivi cette taxonomie pour définir le processus de sortie d'étude comme suit :

- complètement aléatoire (CRD pour Completely Random Drop-out) ; le processus de sortie d'étude et de mesure sont complètement indépendants,
- aléatoire (RD pour Random Drop-out), le processus de sortie d'étude dépend des mesures déjà observées c'est-à-dire qui précèdent la sortie,
- informatif (ID pour Informative Drop-out), le processus de sortie dépend des mesures qui auraient été observées si le sujet n'était pas perdu de vue.

Les différences entre les trois catégories de données manquantes peuvent être illustrées dans notre exemple. Un processus MCAR voudrait dire que les personnes âgées "oublient" de répondre à un ou plusieurs questionnaires annuels, totalement au hasard. Un processus MAR implique que pour une personne donnée, la probabilité de ne pas répondre une année est directement liée aux niveaux d'incapacité recueillis lors des précédentes interviews. Enfin, le processus serait MNAR si en plus de la dépendance aux états précédents, une personne qui par exemple, se retrouverait brutalement à un niveau d'incapacité élevé (accident vasculaire cérébral) sans que les données précédentes puissent le prévoir, aurait une probabilité plus faible de répondre (impossibilité d'utiliser le téléphone ou d'écrire). Le fait que le questionnaire annuel soit léger et rapide, qu'il soit recueilli par téléphone ou par courrier éventuellement auprès d'un proche et qu'il ait été fait des relances chez les non-répondants, minimise le risque d'avoir des données manquantes MNAR, mais ne l'exclut pas.



De manière générale, une méthode qui peut prendre en compte des réponses espacées de manière inégale dans le temps et un nombre de réponses différent selon les sujets, peut prendre en compte des réponses MCAR. Dans ce cas, une méthode du type GEE peut permettre d'analyser de manière simple des données MCAR.

Par contre, des méthodes basées sur les moments, telles que les méthodes GEE, sont biaisées pour des données manquantes MAR [44] tandis que les méthodes basées sur la vraisemblance, comme le modèle mixte, ne le sont pas.

Il faut noter que tous les modèles marginaux ne nécessitent pas la condition très restrictive de données manquantes complètement au hasard. Certains modèles marginaux basés sur la vraisemblance sont valides pour des données MAR. Robins et al [45] ont aussi proposé une extension de la méthode GEE adaptée aux données MAR. Ils ont proposé une classe d'équations d'estimation pondérées, qui donne une estimation sans biais des paramètres après avoir spécifié correctement le mécanisme de données manquantes. Leur approche consiste en une version pondérée de la méthode GEE dans laquelle chaque élément du vecteur de résidus  $(Y_i - \mu_i)$  est pondéré par l'inverse de la probabilité d'avoir observé cette réponse.

Quand les réponses manquantes sont informatives, à la fois les méthodes basées sur la vraisemblance et celles basées sur les moments sont biaisées.

## 1.7 Covariables dépendantes du temps

Nous venons de voir que la nature d'un évènement survenant au cours du temps, comme l'absence ponctuelle de réponse, interfère avec la méthode de prise en compte de la dépendance longitudinale. Il en est de même pour les covariables dépendantes du temps. Dans notre étude, un certain nombre de variables étaient recueillies chaque année, par questionnaire. Nous pouvons par exemple citer : le fait d'avoir été hospitalisé ou alité au cours de l'année précédente, le nombre de chutes dans l'année, les nouvelles pathologies, une perte de poids survenue dans l'année précédente, la santé subjective ou les difficultés pour effectuer

certaines mouvements au moment de l'interview. Par opposition aux variables d'inclusion, ces variables sont mises à jour lors de chaque nouvelle évaluation. La prise en compte dans un modèle de telles covariables et leur interprétation n'est pas toujours aisée comme le montre la littérature sur l'analyse des durées de survie [46]. Cependant, s'agissant de la comparaison entre modèle marginal et modèle mixte une attention particulière doit être portée à cet aspect. Considérons la distribution conjointe de la variable réponse,  $\mathbf{Y}$ , et des variables dépendantes du temps,  $\mathbf{Z}$ , conditionnellement aux covariables d'inclusion,  $\mathbf{X}$ , pour un sujet :

$$f(\mathbf{Y}, \mathbf{Z} | \mathbf{X}) = f(\mathbf{Y}_0, \mathbf{Z}_0 | \mathbf{X}) \prod_t f(\mathbf{Y}_t | \mathbf{Z}_t, \mathcal{F}_t^y, \mathcal{F}_t^z, \mathbf{X}) f(\mathbf{Z}_t | \mathcal{F}_t^y, \mathcal{F}_t^z, \mathbf{X})$$

$\mathcal{F}_t^y$ , étant l'historique des réponses répétées individuelles jusqu'à juste avant le temps  $t$  et  $\mathcal{F}_t^z$  étant l'historique correspondant mais pour les covariables dépendantes du temps,  $f(\mathbf{Y}_0, \mathbf{Z}_0 | \mathbf{X})$ , étant la distribution des données de baseline.

Les covariables dépendantes du temps sont dites exogènes (ou externes) si elles vérifient :

$$f(\mathbf{Z}_t | \mathcal{F}_t^y, \mathcal{F}_t^z, \mathbf{X}) = f(\mathbf{Z}_t | \mathcal{F}_t^z, \mathbf{X})$$

C'est-à-dire si elles ne dépendent pas de l'historique des réponses répétées qui précède leur mesure.

Par opposition, on dit d'une covariable dépendante du temps qu'elle est endogène (ou interne) si sa valeur à un temps  $t$ , dépend des réponses de la variable d'intérêt précédemment observées [39].

Dans notre étude, nous pouvons par exemple considérer que les variables "santé subjective" ou "difficultés à faire certains mouvements" sont des variables endogènes car elles sont probablement liées pour un sujet, aux états d'incapacité qui précèdent. Par contre, la survenue d'une pathologie aiguë serait, elle, une variable exogène.

D'un point de vue statistique, si les deux distributions conditionnelles de  $\mathbf{Y}$  et  $\mathbf{Z}$  n'ont pas de paramètres en commun, on peut utiliser un modèle mixte avec des covariables endogènes,

mais construire directement un modèle marginal contenant des covariables endogènes est incorrect [47].

## 1.8 Interprétation des effets fixes

Comme nous l'avons vu au paragraphe 1.3, dans le modèle marginal les différences entre sujets sont ignorées et ce sont les moyennes qui sont modélisées. Pour les variables explicatives et la variable réponse, le modèle marginal décrit une série non connectée de moyennes dans chaque marge. Dans le modèle mixte, l'hétérogénéité des sujets est prise en compte au moyen des effets aléatoires et le modèle décrit comment chaque réponse individuelle varie en lien avec l'histoire du sujet (les réponses passées ou les covariables dépendantes du temps).

Si nous revenons à notre étude et si nous reprenons la notation précédente avec  $\beta$  le vecteur des paramètres pour le modèle marginal et  $\beta^*$  le vecteur des paramètres pour le modèle à effets aléatoires, si nous considérons par exemple la covariable recueillie à l'inclusion : vivre seule, l'odds ratio,  $OR^* = \exp(\beta^*)$ , représente l'odds de l'évènement (être en incapacité) pour une personne vivant seule comparativement à cette même personne supposée ne vivant pas seule. On peut considérer qu'il s'agit d'un odds ratio ajusté sur les caractéristiques individuelles du sujet, c'est-à-dire sur les covariables individuelles non incluses dans le modèle, ces covariables étant observées ou bien non prévues dans le protocole.

Dans le modèle marginal,  $OR = \exp(\beta)$  représente l'odds moyen du sous-groupe ayant la caractéristique de vivre seul comparativement au sous-groupe ne vivant pas seul. Concernant l'évolution dans le temps le modèle marginal se comporte comme s'il s'agissait de sujets différents à chaque temps ; il ne reflète pas le fait que les investigateurs ont recueilli des données répétées sur chaque sujet.

Nous avons vu sur la figure 4, que les paramètres estimés par le modèle marginal sont atténués par rapport à ceux du modèle à effets aléatoires. Ce phénomène a été étudié par Neuhaus et al [48] qui ont démontré, dans le cas d'un modèle avec seulement une ordonnée à l'origine aléatoire, que si la variance des effets aléatoires est non nulle ( $\text{Var}(U_i) > 0$ ) alors les propriétés suivantes sont vérifiées :

- $|\beta_k| \leq |\beta_k^*|, k=1, \dots, p;$
- $|\beta_k| = |\beta_k^*|$ , si et seulement si  $\beta_k^* = 0;$
- la différence entre  $\beta_k$  et  $\beta_k^*$  augmente avec  $\text{Var}(U_i)$ .

En particulier, si  $U_i$  suit une loi gaussienne de moyenne nulle et de variance  $\sigma_u^2$ , Zeger et al. [30] ont montré que :

$$\text{logit}(E(Y_{ij})) \approx \frac{X'_{ij}\beta^*}{\sqrt{(c^2\sigma_u^2 + 1)}}$$

avec  $c = \frac{16\sqrt{3}}{15\pi}$

et donc :  $\beta \approx \frac{\beta^*}{\sqrt{(c^2\sigma_u^2 + 1)}}$

avec  $c^2 \approx 0,346$

Ces relations prouvent que les estimations des paramètres moyennés sur la population, sont approximativement inférieures ou égales aux paramètres analogues, spécifiques du sujet. De plus, l'écart entre les deux termes de cette inégalité augmente avec la variance de l'effet aléatoire c'est-à-dire avec l'hétérogénéité entre sujets.

## 2 Comparaisons GEE/GLMM sur les données de dépendance binaires<sup>1</sup>

Pour illustrer les points développés dans le chapitre précédent décrivant les méthodes d'analyse de réponses binaires répétées, nous allons maintenant comparer les résultats obtenus en utilisant le modèle marginal GEE et le modèle à effets aléatoires sur les données de la cohorte Epidos. Cet exemple nous permettra de présenter une stratégie d'analyse de ces réponses longitudinales et d'expliquer comment choisir entre ces deux modèles.

Plus précisément, nous allons d'abord décrire l'évolution de l'incapacité au cours des 6 années de suivi pour lesquelles les données étaient disponibles au moment de l'analyse. En effet, l'analyse statistique a été faite alors que les données de la septième année n'étaient pas encore entièrement saisies et validées. Nous nous concentrerons plus particulièrement sur un problème inhérent à l'analyse des données longitudinales qui est le choix de l'origine du temps et la manière de prendre en compte l'influence du temps sur les réponses. Nous présenterons ensuite un exemple d'incorporation de facteurs de risque dans les modèles, à la fois sous la forme de variables fixes dans le temps et de variables dépendantes du temps.

Dans cette partie, nous utiliserons comme variable réponse l'indicateur de mobilité à quatre niveaux : confiné au lit ou au fauteuil, besoin d'aide pour la toilette et l'habillage, besoin d'aide pour sortir, sort sans aide. Compte tenu du fait que les femmes de la cohorte Epidos étaient relativement autonomes à l'inclusion et dans les premiers temps de suivi, nous avons transformé cet indicateur en une réponse binaire. La personne sera considérée en incapacité si elle est confinée chez elle ou si elle a besoin d'aide pour sortir. Pour chacune des femmes,

---

<sup>1</sup> Ce chapitre a fait l'objet d'un article publié dans *BMC Medical Research Methodology*.

nous avons une série de 7 réponses correspondant à l'inclusion et aux 6 années de suivi, ces réponses annuelles sont soit égales à 1 si le sujet est en incapacité, soit égales à 0 si le sujet est sans incapacité ou soit en données manquantes. Ces données manquantes peuvent être dues à des non-réponses occasionnelles ou au fait que le sujet est perdu de vue ou qu'il est décédé. Les réponses manquantes intermittentes étaient souvent dues à une maladie, une hospitalisation ou un événement familial comme le décès de l'époux. L'analyse présentée dans cette section porte sur les 1548 femmes de la cohorte.

## 2.1 Description des données

La moyenne (écart type) d'âge à l'inclusion était de 80 ans (3,7) et seulement 10% des femmes incluses avaient 90 ans ou plus. La proportion des femmes rapportant ne pas pouvoir sortir de chez elles sans aide augmente modérément dans le temps sauf au cours de la première année où le pourcentage de femmes en incapacité passe de 29% à 39% (Tableau 1). Les séquences des 7 états d'incapacité successifs (trajectoires) diffèrent fortement d'un sujet à l'autre. Seulement 268 femmes (17%) étaient capables de sortir sans aide du début à la fin de l'étude, 77 (5%) étaient en incapacité tout au long des 6 années. De nombreuses autres trajectoires sont bien sûr possibles, mais la fréquence de chacune d'elle est très faible, de l'ordre de 1% à 2%. Le tableau 2 présente les dix trajectoires les plus fréquentes pour illustration.

Au cours des 6 années de suivi, la réversibilité, c'est-à-dire le fait de ressortir d'un état d'incapacité, est assez importante. Au total 498 femmes (32%) présentaient une réversibilité au moins une fois. Cette fréquence de réversibilité est un des arguments importants pour utiliser des modèles sur mesures répétées et non pas des modèles analysant le délai jusqu'à l'entrée en incapacité comme le modèle de Cox.

**Tableau 1 - Description des données par année d'évaluation**

	Inclusion	Année 1	Année 2	Année 3	Année 4	Année 5	Année 6
Nombre de femmes ayant une mesure d'incapacité	1548	1503	1450	1409	1294	1020	970
Pourcentage de femmes en incapacité	29%	39%	41%	46%	47%	48%	52%
Délai depuis l'inclusion (ans)							
Médiane		0,99	1,98	2,97	3,95	4,96	6,12
Etendue		0,91; 1,67	1,95; 2,44	2,93; 3,37	3,87; 4,53	4,87; 6,35	5,88; 6,94
Nombre de décès au cours de l'année précédente		23	34	40	63	58	76

**Tableau 2 : Description des dix trajectoires les plus fréquentes**

Inclusion	Année 1	Année 2	Année 3	Année 4	Année 5	Année 6	N	%
0	0	0	0	0	0	0	268	17,3
1	1	1	1	1	1	1	77	5,0
0	1	1	1	1	1	1	66	4,3
0	0	0	0	0	0	DM	38	2,5
0	0	0	0	0	0	1	34	2,2
1	1	1	1	1	DM	DM	31	2,0
1	1	1	1	DM	DM	DM	30	1,9
0	0	0	0	0	DM	DM	27	1,7
1	1	1	1	1	1	DM	26	1,7
0	1	1	1	1	DM	DM	25	1,6

0: peut sortir sans aide, 1: a besoin d'aide pour sortir ou reste confinée chez elle, DM: donnée manquante

L'étude était initialement programmée pour durer 4 ans. Après cette période, un amendement au protocole a permis de proposer aux femmes de continuer le suivi. Certaines femmes ont alors refusé de continuer et sont sorties de l'étude. C'est ce qui explique en partie que le nombre de questionnaires retournés diminue plus rapidement entre l'année 4 et l'année 5 (tableau 1). Il faut aussi noter que certaines femmes donnent des réponses intermittentes. Par exemple, parmi les 970 femmes ayant encore fourni une réponse à l'année 6, 150 femmes

(15%) avaient au moins une réponse manquante dans les années précédentes. Le processus de non-réponse dépend en partie d'évènements antérieurs observés ou non, tels que la survenue d'une maladie ou d'un événement familial. Il n'est donc probablement pas complètement aléatoire (MCAR). Les conditions d'utilisation du modèle marginal GEE ne sont alors pas complètement respectées.

Notons aussi que le délai entre les évaluations annuelles et l'inclusion variait largement d'une femme à une autre et que l'étendue des délais pour une année donnée avait tendance à augmenter à la fin du suivi. Nous avons par conséquent considéré le délai depuis l'inclusion comme une variable continue. Notons enfin que, grâce à une enquête auprès des mairies, le statut vital était connu pour toutes les femmes même pour celles qui étaient sorties d'étude. Le nombre total de décès survenus au cours des 6 ans était de 294 (19%).

## 2.2 Choix de l'origine du temps

La réponse,  $Y_{i0}$ , donnée par le sujet  $i$ , à l'inclusion, peut être prise en compte de deux façons différentes. Elle peut être considérée soit comme faisant partie de la série de réponses longitudinales et alors, le vecteur réponse a 7 éléments correspondant aux temps 0,1, ..., 6 ans, soit comme une covariable de baseline et alors le vecteur réponse n'a que 6 éléments. Ces deux possibilités conduisent à des structures différentes pour les données brutes comme les tableaux 3 et 4 l'indiquent.





Dans les modèles étudiés (marginal ou à effets aléatoires), l'association entre deux réponses successives est supposée garder la même structure tout au long du suivi. Si l'association entre l'incapacité recueillie à la baseline et les autres réponses est différente des associations entre les réponses du suivi alors, il vaut mieux considérer la réponse d'inclusion comme une covariable.

Dans notre étude, comme dans beaucoup d'études de cohorte, l'examen effectué à l'inclusion était très complet. Cette évaluation comprenait en particulier des examens cliniques et fonctionnels (une série de tests standard mesurant la performance physique) et une mesure de la densité minérale osseuse effectuée à l'hôpital. Il est probable que seules les femmes qui se sentaient en suffisamment bonne forme ont accepté de participer à ces examens, introduisant ainsi un biais de sélection à l'inclusion. Par contre, au cours du suivi, les femmes devaient simplement renvoyer par la poste un questionnaire léger, et même les personnes ayant des limitations physiques pouvaient accomplir cette tâche. Ainsi, au cours du suivi, le profil de notre échantillon évolue, les femmes devenant de plus en plus physiquement dépendantes. Au début de l'étude, les femmes en incapacité devaient probablement l'être de façon légère et pouvaient donc ressortir de cet état. Ensuite, davantage de femmes étaient en incapacité sévère et avaient plus tendance à rester à ce stade jusqu'à la fin de l'étude. Nous pouvons donc supposer que l'association entre les réponses consécutives est plus forte à la fin de l'étude qu'au début. La réponse de baseline devrait donc être considérée comme un cas à part, moins liée aux réponses qui suivent.

Pour tester cette hypothèse, nous examinons la matrice des odds ratios d'être en incapacité à l'année  $j$  ( $Y_{ij}$ ) par rapport à être en incapacité les années précédentes ( $Y_{i(j-r)}$ ). Nous constatons que l'association entre deux temps successifs situés après la baseline est plus forte que celle avec la baseline (tableau 5). Par exemple, si nous observons la diagonale la plus basse qui donne le lien entre deux évaluations séparées d'un an, l'odds ratio entre l'année 1 et la baseline

est de 5,3 tandis que cet odds ratio est supérieur à 16 entre l'année 2 et 1 et pour les paires suivantes. Ces résultats nous confortent dans l'idée que la réponse à la baseline doit être considérée comme une covariable. Les données brutes ont donc été structurées comme indiqué dans le tableau 4.

**Tableau 5 - Odds ratios d'être en incapacité une année par rapport à être en incapacité les années précédentes**

	Inclusion	Année 1	Année 2	Année 3	Année 4	Année 5	Année 6
Inclusion		5,3	5,1	5,7	4,1	4,3	4,8
Année 1			16,3	12,2	11,1	10,0	9,5
Année 2				18,2	14,3	13,1	11,4
Année 3					16,7	17,3	13,1
Année 4						20,2	15,8
Année 5							20,7

Nous remarquons aussi dans ce tableau que les corrélations ont tendance à décroître lorsque le temps qui sépare deux mesures augmente (lignes du tableau 5). Par exemple, l'odds ratio est de 16,3 entre l'année 1 et 2 et de 10,0 entre l'année 1 et 5. Cette constatation nous permet de mieux sélectionner la structure de corrélation dans les modèles. En particulier, la structure de corrélation "échangeable" dans le modèle marginal ne semble pas adaptée. De même, il semble plus pertinent d'introduire une pente aléatoire pour le temps dans le modèle à effets aléatoires.

### 2.3 Analyse de l'évolution selon l'âge et le temps

Notre objectif est maintenant d'évaluer l'impact du temps sur la proportion de femmes en incapacité. Nous avons choisi d'utiliser l'âge à l'inclusion et le délai depuis la baseline pour caractériser cette évolution. Une autre solution possible aurait été de considérer l'âge en le réactualisant à chaque évaluation c'est-à-dire comme une variable dépendante du temps. Mais

notre option nous permet d'introduire une interaction entre l'âge au départ et le délai depuis la baseline pour tester si chez les femmes les plus âgées, l'effet du temps est plus marqué.

Concernant le choix entre modèle mixte et modèle marginal, plusieurs éléments nous permettent déjà d'orienter notre décision. En particulier, le processus de données manquantes impliquant que l'échantillon se modifie à chaque évaluation et le possible biais de sélection à l'inclusion nous oriente vers un modèle à effets aléatoires. De plus, dans les études de cohorte la recherche de relations individuelles, pour un individu type, telle que le permet le modèle mixte, est plus adaptée que l'étude de relations moyennées sur la population [39]. Cependant, il est intéressant de comparer les deux modèles pour montrer combien les estimations des paramètres peuvent différer et pour mieux expliquer ces différences. Pour cela nous avons utilisé deux méthodes pour caractériser les changements d'état d'incapacité au cours du temps : le modèle marginal avec une estimation par GEE et le modèle à effets aléatoires avec intégration numérique de la vraisemblance et des effets aléatoires gaussiens. Nous avons utilisé la procédure GENMOD de SAS avec l'option REPEATED et une matrice de corrélation de travail non structurée pour le modèle marginal et la procédure NLMIXED de SAS avec la méthode d'intégration de la quadrature de Gauss-Hermite pour le modèle mixte. Le programme SAS correspondant est donné ci-après.

code SAS:

```
proc GENMOD data=example descending ;
class num ;
model disab = time age agetime died disab0 / dist = bin
                                                link = logit ;
repeated subject=num / type=un corrw ;
run ;

proc nlmixed data=example qmax=100 ;
parms beta0=-3.6 beta1=0.2 beta2=0.2 beta3=0.02 beta4=1.4 beta5=2.9
s2u=8 s2v=0.2 ;
eta = beta0 + beta1*time + beta2*age + beta3*agetime + beta4*died +
beta5*disab0 + u + v*time ;
expeta = exp(eta);
p = expeta/(1+expeta);
```

```
model disab ~ binary(p);  
random u v ~ normal([0,0],[s2u,0,s2v]) subject=num;  
run;
```

Dans le modèle aléatoire, la structure de la covariance a été déterminée en introduisant successivement deux effets aléatoires : une ordonnée à l'origine aléatoire et une pente aléatoire pour le temps.

Dans les deux modèles, les effets fixes suivants ont été testés : le délai depuis la baseline (variable continue), le délai au carré, l'âge à l'inclusion (en années au-delà de 74 ans), l'interaction (âge x délai) et l'incapacité de baseline. Seul le délai au carré n'était pas significatif et a été finalement retiré du modèle. Nous avons aussi introduit une variable indicatrice pour les femmes décédées au cours des 6 ans. Ceci est une façon simple de prendre en compte les sorties d'étude dues au décès, une difficulté souvent rencontrée dans les études longitudinales chez les personnes âgées. Le statut vital de toutes les femmes de la cohorte étant connu pour les 6 années de l'étude, il était donc possible d'utiliser une telle indicatrice.

La procédure NLMIXED a parfois des difficultés à converger et les estimations ne sont produites que lorsque les paramètres initiaux (indiqués avec la commande PARMS) ne sont pas trop éloignés de la solution finale. Pour contourner cette difficulté, nous avons calculé des modèles successifs en commençant par un modèle simple avec seulement une ordonnée à l'origine aléatoire et en utilisant, à chaque étape, les estimations des paramètres du modèle précédent comme valeurs initiales pour l'estimation suivante.

Les résultats de la comparaison entre les deux types de modèles sont présentés dans le tableau 6.

**Tableau 6 - Analyse de l'influence du temps et de l'âge selon le modèle marginal et le modèle à effets aléatoires.**

	Modèle marginal			Modèle à effets aléatoires		
	GEE			Intégration de la vraisemblance		
	Paramètres	Erreur standard	p value	Paramètres	Erreur standard	p value
Ordonnée à l'origine	-1,83	0,12	< 0,0001	-3,61	0,25	< 0,0001
Délai (ans)	0,12	0,03	< 0,0001	0,17	0,06	0,002
Age à l'inclusion*	0,09	0,02	< 0,0001	0,16	0,03	< 0,0001
Age* x délai	0,01	0,004	0,003	0,04	0,009	< 0,0001
Incapacité de baseline	1,54	0,10	< 0,0001	3,18	0,22	< 0,0001
Décédée au cours de l'étude	0,68	0,12	< 0,0001	1,36	0,26	< 0,0001
Variance de l'ordonnée à l'origine aléatoire	-	-	-	7,25	0,68	< 0,0001
Variance de la pente aléatoire sur le délai	-	-	-	0,18	0,04	< 0,0001

\* âge moins 74 ans

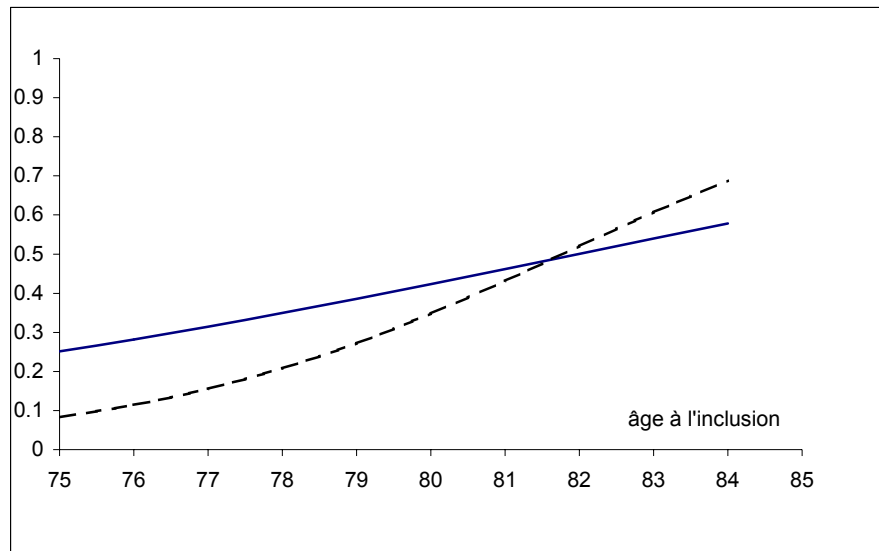
Comme nous l'avons vu au paragraphe 1.4, la méthode d'estimation utilisée par NLMIXED est longue et complexe aussi pour valider ces estimations, nous avons utilisé le logiciel WinBUGS qui produit des estimations selon les méthodes MCMC. La comparaison des estimations produites par les deux logiciels est fournie dans le tableau 7. Dans WinBugs, les estimations des paramètres ont été obtenues sur 7000 échantillons et sont données dans la colonne "moyenne". Ces estimations sont très proches de celles fournies par SAS.

**Tableau 7 - Comparaison entre les estimations du modèle à effets aléatoires calculées selon la méthode MCMC et la méthode d'intégration numérique.**

	WinBUGS						SAS NLMIXED	
	moyenne	Écart type	MC erreur	2,5%	médiane	97,5%	Paramètre estimé	Ecart standard
Ordonnée à l'origine	-3,649	0,2747	0,02456	-4,213	-3,644	-3,13	-3,61	0,25
Temps	0,1798	0,0645	0,005807	0,0576	0,178	0,3109	0,17	0,06
Age	0,1693	0,0372	0,003397	0,0978	0,1689	0,2475	0,16	0,03
Age x temps	0,0370	0,0103	9,136E-4	0,0172	0,0372	0,0566	0,04	0,009
Incap <sub>0</sub>	3,19	0,2089	0,008613	0,8466	1,347	1,84	3,18	0,22
Décédée	1,349	0,2503	0,009351	2,774	3,189	3,61	1,36	0,26
Précision ordonnée à l'origine	0,1365	0,0129	6,544E-4	0,1133	0,1358	0,1634	7,25*	0,68
Précision pente	5,83	1,356	0,1081	3,892	5,563	9,212	0,18*	0,04

\* variance=1/précision

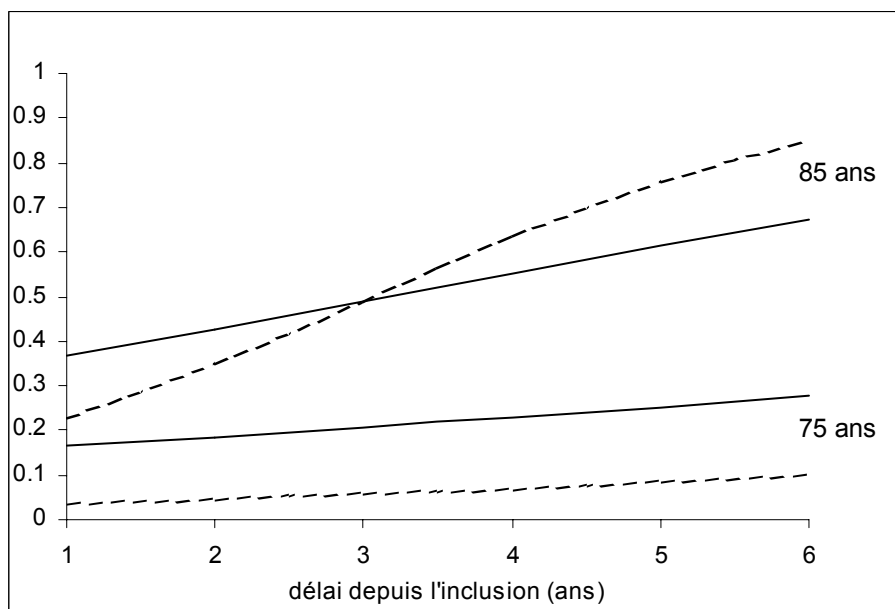
S'agissant de la comparaison entre le modèle marginal et le modèle à effets aléatoires, d'un point de vue général, les conclusions sont identiques pour les deux modèles mais les estimateurs sont différents. Par exemple, le facteur le plus important, l'incapacité de baseline, est très significatif dans les deux modèles mais les paramètres estimés sont égaux à 1,54 et 3,18 (tableau 6). Le risque d'incapacité croissant selon l'âge et le délai depuis l'inclusion est illustré par la figure 6 qui représente la probabilité moyenne d'être en incapacité, 5 ans après l'inclusion, pour le groupe de femmes capables de sortir sans aide à l'inclusion et n'étant pas décédées au cours du suivi. Dans cette cohorte de femmes, recrutées en 1992-1993, la probabilité d'avoir une mobilité réduite, 5 ans plus tard, était élevée, particulièrement chez les femmes âgées de 80 ans et plus.



Légende : modèle GEE marginal ———— , modèle aléatoire-----

**Figure 6 : Probabilité à 5 ans d'être en incapacité en fonction de l'âge à l'inclusion chez les femmes sans incapacité à l'inclusion et survivant jusqu'à la fin de l'étude.**

La figure 7 montre l'évolution avec le temps de deux groupes de femmes âgées de 75 et 85 ans à l'inclusion. Il y avait une interaction significative entre l'âge à l'inclusion et le temps et par conséquent le risque d'incapacité s'accélère davantage chez les femmes les plus âgées.



Légende : modèle GEE marginal ———— , modèle aléatoire-----

**Figure 7 : Evolution de la probabilité d'être en incapacité chez les femmes sans incapacité à l'inclusion, vivantes à la fin de l'étude, âgées de 75 et 85 ans à l'inclusion.**



Comme nous l'avons déjà vu plus haut (paragraphe 1.8), les différences observées entre les estimateurs produits par le modèle mixte et ceux produits par le modèle à effets aléatoires, étaient attendues. Le modèle marginal exprime des relations moyennées sans tenir compte du fait qu'il s'agit des mêmes sujets, revus à chaque temps, alors que le modèle à effets aléatoires fournit des associations conditionnellement au fait d'avoir certaines caractéristiques individuelles modélisées par les effets aléatoires. Nous retrouvons les propriétés sur les estimateurs démontrées par Nehaus (paragraphe 1.8) : les estimations des paramètres données par le modèle marginal sont systématiquement plus faibles que celles du modèle aléatoire. Ces propriétés sont visibles sur les figures 6 et 7 sur lesquelles les courbes du modèle marginal sont plus aplaties que les autres. Comme nous l'avons vu, les différences entre les paramètres estimés selon les deux méthodes sont largement liées à l'hétérogénéité inter-individuelle. Cette hétérogénéité peut être évaluée en regardant dans le modèle à effets aléatoires, la dispersion de l'ordonnée à l'origine et de la pente aléatoire. Dans notre analyse (tableau 6), la variance de l'ordonnée à l'origine aléatoire est égale à 7,25, ce qui donne un intervalle de confiance à 95% de 10,6 et donc une variabilité supplémentaire de cette ordonnée à l'origine. Par comparaison, la variabilité provenant des covariables fixes (âge variant de 75 à 85 ans, incapacité à la baseline et décès variant de 0 à 1) a une largeur est de 6,14 ( $0,16 \times 10 + 3,18 + 1,36$ ). La variabilité expliquée par les trois covariables est donc plus faible que la variabilité inter-individuelle, non expliquée. De façon identique, la pente aléatoire a une largeur à 95% de 1,66, tandis que la variabilité expliquée par l'âge a une largeur à 95% de 0,40 ( $0,04 \times 10$ ).

La probabilité d'être en incapacité est donc davantage due à la fragilité individuelle, non décrite, qu'à l'âge ou l'incapacité de baseline. Cette hétérogénéité inter-individuelle explique les différences entre les estimateurs des paramètres du modèle marginal et du modèle mixte.

## 2.4 Analyse des autres facteurs de risque

Nous allons maintenant comparer les estimations données par le modèle marginal et le modèle à effets aléatoires pour identifier d'autres facteurs de risque de l'incapacité. Ceci va aussi nous permettre de tester le comportement des algorithmes d'estimation lorsque plusieurs covariables sont introduites conjointement.

La comparaison des deux modèles multivariés, ajustés sur les covariables décrites précédemment (âge, délai, interaction entre l'âge et le délai, incapacité de baseline, vivant à la fin du suivi) est donnée dans le tableau 8.

**Tableau 8 - Comparaison du modèle marginal et du modèle à effets aléatoires dans l'analyse de covariables fixes dans le temps et de covariables dépendantes du temps.**

	Modèle marginal*			Modèle à effets aléatoires*		
	GEE			Intégration de la vraisemblance		
	Paramètre	Erreur standard	p value	Paramètre	Erreur standard	p value
<b>Inclusion</b>						
Vivant seul	-0,31	0,09	0,0004	-0,66	0,18	0,0002
IMC (kg/m <sup>2</sup> )						
< 25 (ref)	0			0		
25 – 29	0,30	0,10	0,002	0,69	0,20	0,0005
≥ 29	0,75	0,12	< 0,0001	1,46	0,23	< 0,0001
Acuité visuelle						
≤ 2/10	0,64	0,20	0,001	1,21	0,35	0,0006
> 2/10 (ref)	0			0		
Santé perçue						
Mauvaise ou très mauvaise	0,76	0,15	< 0,0001	1,43	0,28	< 0,0001
<b>Suivi</b>						
Hospitalisée (année précédente )	0,26	0,05	< 0,0001	0,49	0,11	< 0,0001
Temporairement confinée au lit (année précédente)	0,39	0,07	< 0,0001	0,73	0,15	< 0,0001
Nombre de chutes ≥ 2 (année précédente)	0,23	0,08	0,004	0,49	0,15	0,0013

\*ajusté sur le délai depuis l'inclusion, l'âge à l'inclusion, l'interaction âge x délai, l'incapacité de baseline et le décès

Comme précédemment, nous avons utilisé la méthode MCMC, pour valider nos estimations du modèle à effets aléatoires. Ces estimations sont données dans le tableau 9 et sont très proches de celles calculées avec SAS.

**Tableau 9 - Comparaison entre les estimations du modèle à effets aléatoires calculées selon la méthode MCMC et la méthode d'intégration numérique.**

	WinBUGS						SAS NL MIXED	
	moyenne	Écart type	MC erreur	2,5%	médiane	97,5%	Paramètre estimé	Ecart standard
Ordonnée à l'origine	-4,222	0,2891	0,02484	-4,812	-4,206	-3,709	-4,13	0,28
Temps	0,1851	0,0586	0,004964	0,0696	0,1837	0,2984	0,18	0,06
Age	0,1694	0,0330	0,00285	0,1093	0,1688	0,2392	0,16	0,03
Age x temps	0,03425	0,0093	7,833E-4	0,0166	0,0341	0,0518	0,03	0,009
Décédé	1,114	0,2485	0,008195	0,6342	1,113	1,608	1,11	0,24
Incap <sub>0</sub>	2,901	0,2081	0,009459	2,485	2,903	3,317	2,85	0,21
Seule <sub>0</sub>	-0,6739	0,1831	0,009749	-1,022	-0,6759	-0,310	-0,66	0,175
BMI <sub>3</sub> <sub>0</sub>	0,6988	0,2128	0,01252	0,2827	0,6992	1,124	0,69	0,20
BMI <sub>4</sub> <sub>0</sub>	1,482	0,2483	0,01446	0,9915	1,48	1,961	1,46	0,23
Acuité <sub>0</sub>	1,234	0,3613	0,01241	0,5051	1,235	1,944	1,21	0,35
Hospitalisée	0,4914	0,1114	0,00188	0,2784	0,4916	0,7117	0,50	0,11
Alitée	0,7331	0,1528	0,002762	0,4394	0,7324	1,033	0,73	0,15
Santé perçue <sub>0</sub>	1,455	0,2867	0,01093	0,8999	1,456	2,017	1,43	0,28
chutes	0,496	0,1571	0,002904	0,1823	0,4968	0,8055	0,49	0,15
Précision ordonnée à l'origine	0,1596	0,0164	8,968E-4	0,1306	0,1582	0,194	6,03*	0,60
Précision pente	6,281	1,659	0,1499	4,025	5,905	10,34	0,16*	0,03

\* variance=1/précision

Quand les covariables introduites dans le modèle sont nombreuses, la procédure NL MIXED de SAS ne réussit parfois pas à intégrer la vraisemblance ou bien peut donner des estimations

non stationnaires. Pour détecter ces estimations non optimales, il faut systématiquement contrôler le gradient (le vecteur des dérivées premières) du logarithme de la vraisemblance pour chaque paramètre. Ces gradients, systématiquement fournis par SAS, doivent être proches de zéro, sinon la solution proposée ne peut pas être acceptée.

Nous n'avons pas rencontré ces problèmes de convergence lors de la modélisation des covariables présentées dans le tableau 8. Nous avons testé deux catégories de covariables : des variables uniquement collectées à la baseline (vivre seule, indice de masse corporelle (IMC), acuité visuelle, santé perçue) et des variables se modifiant au cours du temps (avoir été hospitalisée au moins une fois au cours de l'année précédente, avoir été confinée au lit de manière temporaire au cours de l'année précédente et le nombre de chutes au cours de l'année précédente). Dans le modèle multivarié, toutes ces variables sont significatives. Comme précédemment, nous retrouvons le fait que si les degrés de significativité sont semblables, les estimations des paramètres sont atténuées dans le modèle marginal.

## 2.5 Discussion

Dans cette partie, nous avons présenté les différentes étapes permettant de choisir le modèle capable de décrire au mieux l'évolution de l'incapacité considérée comme une réponse binaire répétée.

Pour choisir le modèle le plus adapté concernant l'analyse de données répétées, plusieurs points importants doivent être pris en considération.

En premier lieu, le choix entre modèle marginal et modèle à effets aléatoires est d'abord dicté par l'objectif de l'étude. Il faut garder à l'esprit que les deux méthodes ne modélisent pas la même chose et que l'interprétation des paramètres diffère. Dans le modèle marginal, l'exponentielle d'un paramètre est un odds ratio moyenné sur la population et concerne la sous population ayant en commun une caractéristique par rapport à la sous population ne l'ayant pas. Dans le modèle à effets aléatoires, on estime l'odds ratio d'une personne qui a une

caractéristique par rapport à cette même personne supposée ne pas avoir cette caractéristique. Le modèle à effets aléatoires prend en compte la dépendance individuelle sous-jacente. De plus, il faut noter que les hypothèses concernant les distributions sont différentes et que le fait que la distribution conditionnelle soit binomiale n'implique pas que la distribution marginale soit aussi binomiale [47, 49]. En accord avec les résultats démontrés dans la littérature, notre analyse montre que les différences entre les estimations fournies par les deux types de modèle, peuvent être importantes.

En épidémiologie, la conduite suivante peut être adoptée. Si le but est de prédire l'évolution au cours du temps d'une prévalence moyenne d'incapacité selon le sexe et l'âge, le modèle marginal est bien adapté. Si, par contre, le but est d'analyser les facteurs de risque individuels dans une perspective plus étiologique, le modèle à effets aléatoires est recommandé car il permet d'ajuster sur les caractéristiques non observées du sujet et de mieux comprendre le mécanisme sous jacent du phénomène étudié.

En second lieu, il est nécessaire d'examiner le processus des données manquantes. Le modèle marginal avec estimation par GEE suppose que l'échantillon est représentatif à chaque temps de la population étudié et que le processus de réponses manquantes est MCAR. Il est peu probable que ce soit le cas dans notre exemple de cohorte sur les personnes âgées. Le modèle à effets aléatoires, pour lequel les données manquantes doivent être MAR, semble plus approprié. Cependant, nous ne pouvons pas exclure que les états d'incapacité non observés puissent influencer une partie du processus de données manquantes. L'amplitude et le biais potentiel introduit par ces éventuelles réponses manquantes non au hasard devraient être testés. Il serait intéressant de mesurer l'impact des données manquantes sur l'inférence statistique par une analyse de sensibilité [50] mais ceci est en dehors de notre projet.

Un troisième point à considérer, est de savoir si l'incapacité mesurée à l'inclusion doit être considérée comme faisant partie de la réponse ou comme étant une covariable. En

épidémiologie, les cohortes présentent souvent des conditions différentes de recueil de la réponse à l'inclusion et au cours du suivi. La première réponse est donc souvent considérée comme une covariable mais ce choix doit aussi être étayé par l'analyse du tableau des corrélations entre les réponses successives.

Enfin, la structure de la covariance entre les réponses doit aussi faire l'objet d'attention. Dans le modèle marginal, les inférences sur les estimateurs sont asymptotiquement valables quelle que soit la structure imposée, mais il vaut mieux choisir une structure correspondant au mieux aux données. Par contre, dans le modèle aléatoire, les paramètres fixes et aléatoires sont simultanément estimés et la structure de la covariance choisie influence les résultats des estimations.

En conclusion, nous venons de voir, dans notre exemple de cohorte sur les personnes âgées, que le modèle à effets aléatoires est plus approprié. Le modèle marginal a tendance à perdre de l'information et ne mesure pas l'association entre les évolutions intra-sujets des variables et des réponses alors que ces associations sont typiquement la raison pour laquelle on a préféré une étude longitudinale.

### 3 Exemple d'utilisation du GLMM pour réponses binaires : construction d'un score prédictif de la dépendance des personnes âgées<sup>2</sup>.

Dans le chapitre précédent nous avons vu que, dans notre exemple sur la cohorte Epidos et l'analyse de l'incapacité, le modèle logistique GLMM est le plus adapté. Nous allons maintenant approfondir l'analyse de l'évolution de l'incapacité et de ses facteurs de risque en utilisant ce modèle.

#### 3.1 Objectifs

Cette nouvelle analyse s'appuie sur le constat fait par les gérontologues que le déclin avec l'âge des performances physiques n'est pas uniforme [49]. Des interventions de prévention pourraient alors se révéler efficaces dans l'objectif d'un vieillissement réussi. Quelques programmes d'intervention ont montré que les personnes âgées ayant encore un bon niveau fonctionnel sont celles qui tirent le plus grand bénéfice de ces programmes [51]. Ceci suggère que les interventions ciblées au début du processus de déclin fonctionnel, sont potentiellement les plus efficaces.

Un concept de "sujets âgés fragiles" est parfois introduit pour définir la population cible.

Dans cette optique, nous nous proposons donc d'identifier les facteurs de risque du déclin fonctionnel et de caractériser de façon précoce les états fonctionnels qui seront ensuite associés à une incapacité. Nous pensons que cette période précoce est un moment crucial dans la préservation de l'autonomie et d'une bonne qualité de vie des personnes âgées. Pour cela, nous nous polarisons sur l'analyse des personnes sans incapacité à l'inclusion et nous prenons

---

<sup>2</sup> Ce chapitre a fait l'objet d'un article accepté dans *Journal of Clinical Epidemiology*

comme variable d'intérêt, le critère d'incapacité : perte d'au moins une activité instrumentale de la vie quotidienne (AIVQ).

Le premier objectif de notre étude est donc d'identifier à l'inclusion les meilleurs prédicteurs de l'apparition d'une AIVQ au cours du suivi de la cohorte Epidos. Le second objectif est de construire un score prédictif qui pourrait permettre soit, dans une approche individuelle, de sélectionner des femmes en bonne santé mais à risque d'incapacité et qui pourraient bénéficier d'un programme d'intervention, soit, dans une approche plus globale, d'évaluer en population le risque d'entrer en incapacité.

### 3.2 Matériel et méthode

Dans cette étude, nous travaillerons sur les 545 femmes montpelliéraines, indemnes d'incapacité à l'inclusion et lors de la première évaluation à un an. En effet pour garantir que nous ne sélectionnons pas des femmes pour lesquelles le processus d'entrée en incapacité a déjà commencé, nous considérons uniquement celles ne présentant pas d'incapacité deux fois de suite (à l'inclusion et à un an). Au moment de cette nouvelle analyse, les données des 7 années de suivi étaient disponibles pour le centre Epidos de Montpellier. Dans un deuxième temps, nous utiliserons les données du centre Epidos de Toulouse pour tester la valeur prédictive de notre score. Cet échantillon indépendant de celui de Montpellier, comprend 807 femmes sans incapacité à l'inclusion suivies pendant 4 ans.

Conformément à notre objectif, nous ne testerons que des covariables recueillies à l'inclusion.

Ces covariables sont les suivantes :

- La densité minérale osseuse pour le corps entier,
- La masse maigre et la masse grasse,
- L'indice de masse corporelle ( $IMC = \text{poids} / \text{taille}^2$ ), le ratio tour de taille sur tour de hanche,
- La perte de taille depuis l'âge de 30 ans,



- Le niveau d'éducation sous la forme d'une variable dichotomique : avoir obtenu ou non le certificat d'étude,
- L'activité physique évaluée à l'inclusion par un questionnaire portant sur l'exercice physique (marche, gymnastique, vélo, nage, jardinage) et le gros ménage. Le temps passé dans chaque activité a été converti en équivalents métaboliques (MET) par semaine et ensuite sommé sur l'ensemble des activités,
- La peur de tomber et la variable indiquant si au moins une chute s'était produite au cours des 6 derniers mois,
- La santé perçue évaluée par deux variables : se considérer en très bonne, bonne, mauvaise ou très mauvaise santé et se considérer en meilleure, aussi bonne ou plus mauvaise santé que les personnes du même âge,
- Les résultats aux tests de performance physique : force du quadriceps, force de préhension, périmètre du mollet, temps mis pour se lever et s'asseoir 5 fois d'une chaise sans utiliser les bras, temps pour taper 10 fois de suite avec la pointe de son pied en alternance en position assise, vitesse de marche et longueur des pas,
- Les résultats aux tests d'équilibre : Temps d'équilibre en position tandem, semi-tandem, capacité à effectuer une marche funambule,
- Acuité visuelle mesurée dans les conditions de la vie courante, avec les lunettes.

Notre variable d'intérêt est donc une réponse binaire, recueillie chaque année : avoir besoin d'assistance pour au moins une AIVQ. Nous analyserons aussi les réponses "avoir besoin d'assistance pour au moins deux AIVQ" et "au moins trois AIVQ" pour voir si les mêmes facteurs de risques sont retrouvés.

L'analyse statistique comportera 6 étapes :

- 1- La description des femmes sélectionnées pour les calculs (n = 545) et la comparaison avec celles exclues de l'analyse parce qu'elles n'avaient pas d'évaluation d'incapacité

au cours du suivi ou avaient des données manquantes sur les covariables d'inclusion.

Le test t de Student ou le test de Wilcoxon seront utilisés pour les variables quantitatives et le test du Chi2 pour les variables qualitatives.

- 2- Une analyse exploratoire recherchant les facteurs liés à l'évolution de l'incapacité. Chaque facteur potentiel sera testé dans un modèle univarié ajusté sur le délai depuis l'inclusion et sur l'âge à l'inclusion.
- 3- Les covariables trouvées significatives à l'étape précédente seront regroupées en catégories selon leur origine clinique. Cinq catégories sont retenues : la mobilité, l'équilibre statique et dynamique, la force (quadriceps et préhension), la composition corporelle, la santé perçue. Certains facteurs seront laissés hors de ces catégories : l'activité physique, l'acuité visuelle, le niveau d'éducation et l'indice de masse corporelle (IMC). Pour chaque catégorie, les facteurs seront inclus dans un modèle multivarié afin de sélectionner le ou les facteurs les plus pertinents. Cette étape permet d'éviter d'entrer dans le modèle final des variables très corrélées entre elles.
- 4- Les facteurs sélectionnés dans l'étape précédente seront inclus simultanément dans un modèle multivarié. Une sélection finale sera effectuée par une méthode pas à pas descendante.
- 5- Afin de créer un score final facile à utiliser en milieu clinique, les coefficients de régression fixes du modèle final seront multipliés et arrondis. L'algorithme de Cole [52] sera utilisé pour calculer ces coefficients arrondis de manière optimale. Cet algorithme consiste en la recherche du plus petit commun multiplicateur, k, qui permet de transformer chaque coefficient estimé en un entier, sans que la précision en soit trop affectée.
- 6- La valeur prédictive de ce nouveau score sera évaluée sur le même jeu de données et en utilisant les données fournies par le centre Epidos de Toulouse. Les performances

de ce score prédictif seront mesurées par la c-statistique qui est, dans le cas d'un évènement binaire, identique à l'aire sous la courbe de ROC [53]. Nous calculerons cette statistique sur l'ensemble des réponses répétées des 7 ans de suivi pour le centre de Montpellier et des 4 ans de suivi pour le centre de Toulouse.

Les covariables quantitatives ont été transformées en 3 catégories selon les quartiles : le quartile le plus bas, les deux quartiles du milieu réunis et le quartile le plus haut. Ainsi les odds ratios ne dépendent plus de l'unité de mesure et peuvent être comparés entre tous les facteurs. Ceci nous permettra d'établir une hiérarchie entre les facteurs pronostiques allant de ceux ayant l'odds ratio le plus élevé à ceux ayant l'odds ratio le plus bas.

Dans les étapes 2, 3 et 4, le modèle utilisé sera un modèle logistique, à effets aléatoires, défini comme suit :

$$\text{Logit}\left(P(Y_{ij} = 1 | t_{ij}, X_{il}, u_i)\right) = \alpha_0 + \alpha_1 t_{ij} + \alpha_2 \text{age} + \alpha_3 (\text{age} * t_{ij}) + \sum_{l=1}^p \beta_l X_{il} + u_i$$

Avec  $Y_{ij}$ , la réponse binaire (AIVQ > 0), correspondant à la  $j^{\text{ème}}$  réponse, mesurée au temps  $t_{ij}$ , sur le sujet  $i$  ( $i=1$  à 545) et  $X_{il}$ , les covariables de baseline ( $l=1$  à  $p$ ).

Les paramètres  $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_l$  correspondent aux effets fixes communs à tous les sujets tandis que les  $u_i$  sont les ordonnées à l'origine aléatoires, indépendantes et identiquement distribuées selon une loi normale. Les effets aléatoires décrivent les déviations individuelles par rapport au score moyen prédit par la partie fixe du modèle. Dans notre précédente étude exposée au paragraphe 2, nous avons opté pour une structure de covariance avec une ordonnée à l'origine et une pente aléatoire pour le temps. Bien que la structure du paragraphe 2 semble plus adaptée, dans cette nouvelle analyse nous utilisons une structure avec uniquement une ordonnée à l'origine aléatoire. Ceci nous permet de garder des temps de calcul acceptables malgré le nombre important de modèles successivement estimés.

Le score prédictif final porte sur les effets fixes, il est obtenu en multipliant les paramètres estimés,  $\alpha_1, \alpha_2, \alpha_3, \beta_l$ , par le coefficient d'échelle,  $k$ , et en arrondissant ces paramètres à

l'entier le plus proche. L'effet aléatoire n'est pas utilisé dans la prédiction car il est particulier à chaque sujet et ne peut donc être généralisé à de nouveaux sujets. Le score prédictif est donc un score commun à tous les sujets ayant les caractéristiques modélisées par les effets fixes mais il est ajusté sur les caractéristiques inconnues propres à chaque sujet.

Pour cette analyse, nous avons utilisé la procédure NLMIXED de SAS. Comme certains problèmes de convergence peuvent survenir avec cette procédure, la sélection du modèle final a aussi été effectuée avec le logiciel WinBUGS utilisant les méthodes MCMC.

### 3.3 Résultats

Parmi les 691 femmes sans incapacité à l'inclusion et à la première évaluation, les données étaient manquantes pour au moins une covariable de baseline chez 146 femmes. L'analyse finale a donc été effectuée sur 545 femmes. La comparaison entre ces 545 femmes sélectionnées et les 146 femmes exclues de l'analyse est présentée dans le tableau 10. Les femmes incluses dans l'analyse sont plus jeunes (âge médian égal à 79 ans au lieu de 80 ans,  $p < 0,005$ ). Notons que la plupart des femmes vivaient encore chez elles à l'inclusion (92,7%).

**Tableau 10 - Description des femmes sélectionnées pour l'analyse versus celles exclues pour cause de données manquantes**

	Sélectionnées pour l'analyse (n = 545)	Avec données manquantes (n = 146)	Valeur de p *
Age en années [médiane (IIQ)]	79 (76-81)	80 (77-82)	< 0,005
IMC en kg/m <sup>2</sup> [moyenne (ET)]	25,2 (3,8)	25,9 (4,0)	0,05
Santé perçue (%)			
En meilleure santé	43,3	34,1	
En aussi bonne santé	55,0	65,9	
En plus mauvaise santé	1,7	0,0	0,04
Vivant à domicile (%)	92,7	95,2	0,22
Niveau d'éducation : école primaire (%)	46,8	43,0	0,41
Hospitalisée l'année précédente (%)	15,6	18,6	0,38
Antécédent de cancer (%)	19,8	21,9	0,57
Antécédent de maladie cardiovasculaire (%)	26,4	23,3	0,44
Antécédent de dépression (%)	0,9	2,1	0,38

\* test de Wilcoxon, ou test *t* de Student, ou test du Chi 2, ou test exact de Fisher.

La proportion des femmes en incapacité (AIVQ > 0) croît de 22,1% à la fin de la première année jusqu'à 52,1% à la fin de la septième année (Tableau 11). Au cours de ces 7 années, la réversibilité, c'est-à-dire un score d'AIVQ revenant à zéro, a été observée au moins une fois chez 155 femmes (28,4%). Certaines femmes ont donné des réponses intermittentes. Par exemple, parmi les 432 femmes donnant encore une réponse à la septième année, 83 femmes (19,2%) avaient au moins une réponse manquante au cours des années précédentes. Le délai réellement observé entre les réponses annuelles et l'inclusion varie d'une femme à l'autre et comme dans l'étude du paragraphe 2, ceci nous conduit à considérer le délai comme une variable quantitative dans les modèles présentés ci-dessous. Le statut vital était renseigné pour toutes les participantes au long des 7 ans de suivi et au total, à la fin de l'étude, 63 femmes étaient décédées (11,6%).

**Tableau 11 - Description des données par année de suivi**

	Année 2	Année 3	Année 4	Année 5	Année 6	Année 7
Effectif de femmes avec évaluation de l'incapacité	542	525	498	438	413	432
Nombre de décès au cours de l'année précédente	1	8	15	13	15	11
Effectif de femmes sans évaluation de l'incapacité	2	11	23	70	80	50
Femmes en incapacité (%)						
AIVQ > 0	22,1	27,4	33,7	36,3	42,4	52,1
AIVQ > 1	10,0	10,7	18,1	17,8	27,9	33,8
AIVQ > 2	5,4	5,7	10,2	12,1	19,4	24,1
Délai depuis l'inclusion (ans)						
Médiane	1,98	2,97	3,94	4,96	6,06	6,99
Etendue	1,95-2,37	2,95-3,26	3,90-4,26	4,89-5,89	5,92-6,87	6,91-7,59

Les tableaux 12a et 12b donnent les odds ratios et leur intervalle de confiance à 95% de tous les facteurs de risque significativement associés à l'incapacité (AIVQ > 0) lorsque ces facteurs sont testés dans des modèles séparés comprenant en outre l'âge à l'inclusion, le délai et l'interaction (âge x délai). Ces facteurs prédictifs sont présentés selon les groupes définis pour la pré-sélection avant le modèle final. Au total 21 facteurs se sont avérés significativement associés à l'incapacité.

**Tableau 12a - Modèles univariés ajustés sur l'âge, le délai depuis l'inclusion et l'interaction âge x délai.**

	Odds-ratio	Intervalle de confiance à 95%
<b>MOBILITE</b>		
Vitesse de marche(m/sec)		
< 0,78	6,31 <sup>****</sup>	3,26 ; 12,19
0,78 – 1,0	2,32 <sup>**</sup>	1,32 ; 4,08
≥ 1,0	1	
Nombre de pas pour marcher 6 mètres		
< 11	1	
11 – 13	1,01	0,50 ; 2,05
≥ 13	5,31 <sup>****</sup>	2,43 ; 11,59
Temps pour se lever 5 fois d'une chaise (sec)		
< 9,6	1	
9,6 – 13	2,35 <sup>**</sup>	1,30 ; 4,24
≥ 13	7,62 <sup>****</sup>	3,86 ; 15,04
<b>EQUILIBRE</b>		
Temps pour effectuer 10 pointers du pied (sec)		
< 4,25	1	
4,25 – 6,6	2,05 <sup>*</sup>	1,14 ; 3,67
≥ 6,6	5,35 <sup>****</sup>	2,75 ; 10,43
Temps d'équilibre en position tandem (sec)		
0	6,07 <sup>****</sup>	2,86 ; 12,88
0,1 – 5	1,74	1,00 ; 3,01
≥ 5	1	
Difficultés pour effectuer une marche pieds tandem		
Capable de faire 4 pas consécutifs	1	
Incapable sans trébucher	1,27	0,74 ; 2,17
Incapable	8,36 <sup>***</sup>	2,70 ; 25,91
<b>FORCE</b>		
Force de préhension (Kpa)		
< 47	2,81 <sup>**</sup>	1,43 ; 5,52
47 - 61	1,36	0,77 ; 2,40
≥ 61	1	
Force du quadriceps (Newton/longueur de la jambe)		
< 3,52	2,69 <sup>**</sup>	1,38 ; 5,24
3,52 – 4,95	1,50	0,85 ; 2,66
≥ 4,95	1	
<b>Acuité visuelle</b>		
≤ 4	1,77	0,97 ; 3,26
> 4	1	

\* p < 0,05, \*\* p < 0.,01, \*\*\* p < 0.,001, \*\*\*\* p < 0.,0001

**Tableau 12b - Modèles univariés ajustés sur l'âge, le délai depuis l'inclusion et l'interaction âge x délai**

	Odds-ratio	Intervalle de confiance à 95%
<b>COMPOSITION CORPORELLE</b>		
Poids (kg)		
< 53	1	
53 – 65	0,94	0,52 ; 1,70
≥ 65	2,02*	1,02 ; 3,98
Masse maigre/masse totale		
< 0,54	1	
0,54 – 0,63	0,45**	0,25 ; 0,80
≥ 0,63	0,35**	0,18 ; 0,69
Masse grasse/ masse totale		
< 0,31	1	
0,31 – 0,40	1,09	0,61 ; 1,93
≥ 0,40	2,87**	1,52 ; 5,42
Indice de masse corporelle (kg/m <sup>2</sup> )		
< 22,8	1	
22,8 – 27,6	0,87	0,49 ; 1,54
≥ 27,6	2,97**	1,55 ; 5,70
Tour de taille/tour de hanches		
< 0,81	1	
0,81 – 0,90	1,40	0,79 ; 2,48
≥ 0,90	1,82	0,93 ; 3,58
Gain de poids depuis l'âge de 30 ans (cm)		
> 0	1	
0 – 12	1,02	0,56 ; 1,85
≤ 12	2,02*	1,02 ; 4,02
Gain d'IMC depuis l'âge de 30 ans (cm)		
> 2	1	
0 – 7	1,10	0,62 ; 1,95
≤ 7	2,15*	1,09 ; 4,24
Activité physique (MET/semaine)		
< 6,9	3,53***	1,78 ; 7,01
6,9– 36,6	1,58	0,90 ; 2,79
≥ 36,6	1	
Niveau d'étude		
Ecole primaire	2,33***	1,46 ; 3,72
> Ecole primaire	1	
<b>SANTÉ PERÇUE</b>		
Santé subjective		
Mauvaise ou très mauvaise	7,26**	2,01 ; 26,25
Bonne ou très bonne	1	
Santé comparée / personnes du même âge		
En aussi bonne ou moins bonne santé	2,42***	1,51 ; 3,90
En meilleure santé	1	
Peur de tomber		
Oui	2,57****	1,61 ; 4,12
Non	1	

\* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001, \*\*\*\* p < 0,0001



Le modèle final (après l'étape de sélection pas à pas descendante) est présenté dans le tableau 13 qui indique les odds ratios pour le délai depuis l'inclusion, l'âge à l'inclusion, l'interaction (âge x délai) et pour les neuf autres facteurs de risque encore significatifs. Il faut noter que, dans un souci de simplicité, pour les trois covariables vitesse de marche, activité physique et indice de masse corporelle, la catégorie du milieu a été regroupée avec la catégorie de référence car les paramètres estimés étaient très proches de zéro.

**Tableau 13 - Modèle multivarié et coefficients du score après simplification**

	Modèle final			Coefficient simplifié arrondi (k=16,44)*
	Paramètre	Odds-ratio	Intervalle de confiance à 95%	
Délai depuis l'inclusion (ans)	0,38	1,46	1,33 ; 1,60	6
Age-74 ans x Délai depuis l'inclusion	0,03	1,03	1,02 ; 1,05	1
<b>MOBILITE</b>				
Vitesse de marche (m/sec)				
< 0.78	0,57	1,76	1,04 ; 2,98	9
Temps pour se lever 5 fois d'une chaise (sec)				
9.6 – 13	0,48	1,61	0,92 ; 2,82	8
≥ 13	1,23	3,41	1,74 ; 6,67	20
<b>SANTE PERÇUE</b>				
Santé subjective				
Mauvaise ou très mauvaise	1,25	3,47	1,07 ; 11,33	20
Peur de tomber	0,46	1,58	1,01 ; 2,46	7
<b>EQUILIBRE</b>				
Temps d'équilibre en position tandem (sec)				
0	1,11	3,04	1,49 ; 6,21	18
0,1 – 5	0,33	1,39	0,83 ; 2,34	5
<b>NUTRITION</b>				
Indice de masse corporelle (kg/m <sup>2</sup> )				
≥ 27,6	0,89	2,43	1,48 ; 3,98	15
<b>FORCE</b>				
Force de préhension (Kpa)				
< 47	0,58	1,78	1,07 ; 2,95	9
Activité physique (MET/sem)				
< 6,9	0,51	1,67	0,99 ; 2,82	8
Niveau d'éducation				
Ecole primaire	0,45	1,57	1,02 ; 2,44	7

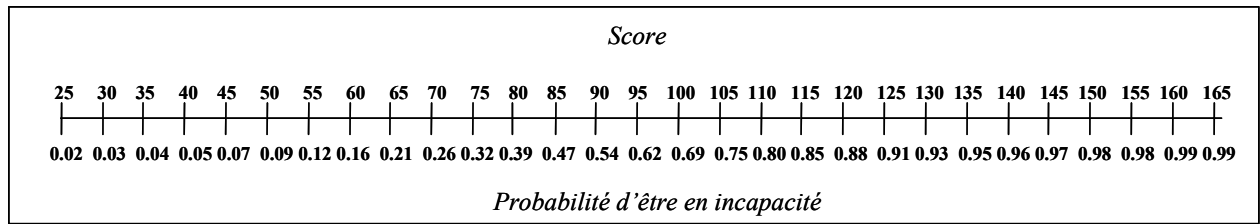
\* k : coefficient multiplicateur donné par l'algorithme de Cole

Les facteurs prédictifs de l'examen d'inclusion, les plus importants, appartiennent au groupe "mobilité" : le temps pour se lever 5 fois d'une chaise, la vitesse de marche et le nombre de pas pour parcourir 6 mètres avec des odds ratios (IC 95%) du quartile des plus basses performances comparativement au quartile des meilleures performances respectivement égaux à 7,62 (3,86 ; 15,04), 6,31 (3,26 ; 12,19) et 5,31 (2,43 ; 11,59) (tableau 12). Dans le modèle final, les deux premiers facteurs subsistent (tableau 13).

Vient ensuite le groupe relatif aux tests d'équilibre (effectuer 10 pointers avec le pied, équilibre en position tandem et marche funambule), mais dans le modèle multivarié la variable "équilibre en position tandem" est la seule qui persiste. Enfin, le groupe "composition corporelle" est lui aussi prédictif avec des mesures telles que l'IMC et la masse grasse relative très significatives dans le modèle univarié. Dans le modèle final, l'IMC semble être la mesure la plus prédictive de l'incapacité. La force de préhension et la force du quadriceps sont toutes deux significatives dans le modèle univarié, mais seule la force de préhension est conservée dans le modèle final.

Concernant les covariables recueillies par questionnaire, le groupe "santé perçue" est particulièrement remarquable avec deux variables restant dans le modèle final : santé subjective et peur de tomber (tableau 13). L'activité physique déclarée était aussi très significative ainsi que le niveau d'éducation.

L'algorithme de Cole calculant un score simplifié à partir du modèle final, produit un multiplicateur optimal unique,  $k$ , égal à 16,44 (Tableau 13). Ainsi pour chaque femme, son score prédictif de dépendance fonctionnelle peut être facilement calculé en sommant les coefficients, donné dans la colonne de droite du tableau 13, pour chaque facteur de risque observé. L'abaque ci-dessous permet de passer facilement du score à la probabilité d'être en incapacité.



**Figure 8 : Abaque de correspondance entre le score et la probabilité d'être en incapacité**

Par exemple, une femme âgée de 78 ans, sans incapacité, avec une force de préhension < 47 Kpa, un IMC  $\geq 27.6$ , une mauvaise santé subjective et qui a peur de tomber, aura 3 ans après, un score égal à  $(6 \times 3 \text{ ans depuis la baseline}) + (4 \text{ ans au-delà de } 74 \text{ ans} \times 3 \text{ ans depuis la baseline}) + 9 \text{ (force de préhension)} + 15 \text{ (IMC)} + 20 \text{ (santé subjective)} + 7 \text{ (peur de tomber)} = 81$  ce qui correspond à une probabilité d'être en incapacité de 0,39.

L'évaluation de la valeur prédictive de ce score a été effectuée sur le même échantillon de 545 femmes observées pendant 7 ans et sur l'échantillon indépendant des 807 femmes du centre Epidos de Toulouse. La statistique, c, était égale respectivement à 0,72 et 0,71 ce qui indique un bon niveau de prédiction global.

## 3.4 Discussion

### 3.4.1 Pertinence de la population ciblée

Les résultats de cette étude sont limités aux femmes âgées de 75 ans et plus. Mais ce segment de la population représente probablement la cible la plus intéressante pour une intervention tendant à différer l'entrée en incapacité. En effet, non seulement les femmes sont largement majoritaires dans la population des personnes âgées, mais de plus il semble que leur profil d'évolution soit différent de celui des hommes. Concernant les incapacités légères, mesurées par l'indicateur AIVQ ou par un score de capacité des membres inférieurs, il a été suggéré que les taux d'incidence sont presque identiques dans les deux sexes, mais que les taux de prévalence sont supérieurs chez les femmes indiquant qu'elles restent plus longtemps en

incapacité. Ceci a été confirmé par d'autres études montrant qu'à la fois le taux de réversibilité et le taux de décès, sont plus faibles chez les femmes [54]. De plus, dans la comparaison avec les hommes du même âge, pour des incapacités plus lourdes mesurées par l'indicateur AVQ, Strauss et al ont trouvé un taux d'incidence plus élevé chez les femmes de 90 ans et une prévalence de la morbidité plus forte chez les femmes de 85 ans et plus [55]. Tous ces résultats cumulés suggèreraient donc que les femmes avec des incapacités légères survivent aux hommes de même âge entre 75 et 85 ans et sont ensuite à plus haut risque de développer des incapacités sévères aux âges très avancés. C'est donc dans cette tranche, des femmes les moins âgées commençant à avoir des incapacités légères, que les interventions de prévention auraient le plus d'intérêt et devraient être conduites en priorité.

Dans la classe d'âge de notre étude, l'espérance de vie est encore élevée (13 ans et 6,5 ans pour les Françaises respectivement âgées de 75 et 85 ans). Cependant, comme le processus de déclin fonctionnel avait déjà commencé pour certains de nos sujets nous avons dû sélectionner un échantillon de femmes sans incapacité pour deux années consécutives afin d'être sûrs de n'étudier strictement que l'incidence de l'incapacité. Notre échantillon est donc constitué de femmes particulièrement alertes pour leur âge et il est donc plus proche d'une classe d'âge plus jeune. Des études supplémentaires sont nécessaires pour définir la classe d'âge pour laquelle le dépistage des sujets à risque serait le plus profitable. Cette classe d'âge optimale est probablement autour de 65 à 75 ans, mais elle tend à se déplacer vers le haut avec la continuelle augmentation de l'espérance de vie.

### 3.4.2 Définition de l'indicateur d'incapacité

Comme notre propos était d'examiner les stades précoces de dépendance fonctionnelle, nous avons préféré l'échelle des AIVQ. Wolinsky et al.[56] ont montré que la détérioration du score AIVQ est un prédicteur important de futures aggravations. Ce score semble plus approprié à notre objectif que celui des AVQ de Katz qui ne capture que les incapacités de la fin du

processus et qui n'est pas apte à déceler les incapacités légères en population vivant à domicile où la prévalence des AVQ est faible [57]. La perte d'une AIVQ a une origine à la fois cognitive et motrice [58]. En particulier les activités complexes telles que gérer son budget ou se servir du téléphone ou prendre ses médicaments sont corrélées aux détériorations cognitives [59]. Notre score prédit donc l'incapacité globale comprenant aussi les détériorations de la cognition.

### 3.4.3 Mise en évidence d'une hiérarchie entre les composants de la fragilité

Notre analyse statistique nous a d'abord permis de modéliser l'impact du temps sur la survenue de l'incapacité. Chaque année supplémentaire, après les observations de baseline, conduit à une augmentation importante de la probabilité d'être en incapacité. De plus, nous avons trouvé une interaction entre l'âge au moment des mesures et le délai, ce qui signifie que le risque s'accélère chez les femmes les plus âgées.

Notre modèle final permet aussi de mesurer le poids relatif de chaque facteur de risque dans la prédiction de l'incapacité au cours des sept années de suivi. Cette combinaison de prédicteurs s'apparente à un score de fragilité physique. La fragilité physique a été définie comme un état physiologique de vulnérabilité croissante à des stress induit par une diminution des réserves physiologiques de multiples systèmes [60, 61]. Dans la population des personnes vivant à domicile, cet état de fragilité est souvent asymptomatique car les individus utilisent de fortes stratégies de compensation. Cet équilibre instable peut être facilement rompu par des changements minimes de l'environnement plus exigeants pour les individus.

Les experts cliniques et les gériatres ont proposé plusieurs critères pour définir la fragilité. Campbell et al [62] définissent la fragilité comme une perte des capacités à faire face aux stress mineurs de l'environnement. Ils suggèrent que ces capacités appartiennent à quatre systèmes : la fonction musculo-squelettique, la capacité aérobie (capacité de faire de l'exercice sans s'essouffler), la fonction cognitive et l'état nutritionnel. Fried et al [63] ont

introduit le concept de "syndrome de fragilité" s'appuyant sur cinq domaines : la nutrition (perte de poids), la force musculaire (test de préhension), l'endurance (fatigabilité), la mobilité (vitesse de marche), et l'activité physique. Au cours de ces dernières années, trois autres domaines ont été cités dans la littérature : l'équilibre, l'activité motrice (vitesse des mouvements, coordination) et la cognition [64, 65].

Notre étude prospective nous a permis de passer en revue un grand nombre de facteurs potentiels, de confirmer le rôle important joué par cinq de ces domaines et d'établir entre eux la hiérarchie suivante : mobilité (vitesse de marche et lever d'une chaise), équilibre (test en position tandem), nutrition (IMC), force musculaire (préhension) et activité physique. Nous avons montré aussi l'impact important de la santé perçue (santé subjective et peur de tomber) sur la probabilité d'entrée en dépendance. Plusieurs auteurs ont démontré une association de la santé perçue avec la mortalité [66], les états chroniques [67], la perte de capacités fonctionnelles [68], l'inactivité physique [69]. Mais, sauf pour la mortalité, cette association a été trouvée dans des études transversales et interprétée comme étant une conséquence de la détérioration physique. En revanche, nos résultats suggèrent que le sentiment d'avoir une santé altérée, augmente le risque de présenter plus tard, des limitations fonctionnelles. La santé auto-évaluée pourrait être un marqueur important de la vulnérabilité des femmes sans incapacité et apparemment indemnes d'affections cliniques.

En accord avec d'autres études [70], les femmes avec un niveau d'éducation faible avaient un risque plus élevé. Un niveau d'éducation faible est souvent synonyme de faibles revenus, de conditions de vie difficiles ou de limitations dans l'accès au soin et aux programmes de prévention autant de paramètres susceptibles de précipiter l'entrée en incapacité.

La dimension nutritionnelle du syndrome de fragilité est habituellement définie comme un processus d'amaigrissement se manifestant par une perte de poids ou de masse maigre [63]. Nous trouvons, par contre, dans notre analyse univariée une association positive entre le

risque d'incapacité et le poids corporel, l'IMC, la masse grasse et le gain de poids depuis l'âge de 30 ans. Dans une étude prospective, Launer et al [71] ont suggéré qu'un IMC élevé est facteur de risque d'une diminution de la mobilité chez les femmes âgées au même titre que la perte de poids. Dans notre cohorte, les femmes ayant des IMC faibles sont rares ce qui ne nous a pas permis d'identifier le mécanisme à travers lequel l'IMC joue sur l'incapacité et de tester si la courbe de risque en "J" trouvée pour la relation entre IMC et mortalité [72] est aussi valable pour l'incapacité.

Dans la cohorte Epidos, les détériorations cognitives n'ont pas été évaluées à l'inclusion et nous n'avons donc pas pu tester cette dimension. Cependant dans notre échantillon très sélectionné de femmes âgées, aucun diagnostic de démence modérée ou sévère n'a été recueilli à l'inclusion et nous pouvons supposer que s'il existe des détériorations cognitives elles restent probablement légères.

Dans notre analyse, certains facteurs repérés significatifs dans des études transversales [73] n'ont pas été retrouvés : antécédent de dépression, de maladie cardiovasculaire, de cancer et la densité minérale osseuse. Ceci suggère que, dans la sélection des femmes à risque d'incapacité, les tests de performance fonctionnelle sont plus pertinents pour évaluer les capacités en réserve qu'un examen des pathologies chroniques. De même, les variables d'inclusion sur la nutrition n'étaient pas significatives dans notre étude mais le questionnaire alimentaire investiguait surtout l'apport en calcium et n'était pas suffisamment exhaustif.

#### 3.4.4 Applicabilité du score prédictif

Nous avons mis au point un score prédictif à partir de la combinaison des facteurs de risque avec des coefficients entiers. La bonne correspondance entre les réponses d'incapacité observées et celles prédites, mesurée sur un échantillon indépendant, nous permet d'avancer que cette nouvelle échelle représente une évaluation relativement précise du risque.

Cependant, une validation de notre score sur une nouvelle cohorte plus large, permettrait de le valoriser.

Une condition importante pour qu'un score soit utilisé en pratique est sa simplicité. Or, il se trouve que les prédicteurs sélectionnés dans le modèle final sont très simples à obtenir. Il s'agit d'une combinaison de données recueillies par auto-questionnaire et de quelques mesures faciles à réaliser (indice de masse corporelle, tests de force, d'équilibre et de mobilité). De plus, le changement d'échelle et le fait d'avoir arrondi les coefficients contribuent à cette simplicité.

Notre étude a essayé de quantifier la fragilité physique considérée comme un précurseur physiologique de l'incapacité. Notre score a donc une valeur pratique très importante pour les cliniciens et pour les professionnels de santé publique. Son utilisation peut se faire soit dans un but individuel ou dans un but collectif selon que l'on cherche à 1) sélectionner des sujets apparemment en bonne santé mais à risque de devenir fonctionnellement dépendants et qui pourraient tirer bénéfice de programmes d'entraînement physique dans le but de retarder la dépendance, ou à 2) évaluer, avec un critère précoce, l'impact d'un programme de santé publique destiné aux personnes âgées.



## Partie III : Réponse ordinale

# 1 Méthodes d'analyse des réponses ordinales corrélées

Jusqu'à présent nous avons considéré l'incapacité sous forme d'une variable binaire. Or, l'incapacité des personnes âgées se caractérise par différents niveaux correspondant à des états de plus en plus graves. Tous les indicateurs destinés à mesurer l'incapacité sont des échelles distinguant ces différentes phases. L'analyse de ces indicateurs sous forme de variables binaires oblige à choisir un seuil pour découper la réponse initiale. Cette dichotomisation, forcément arbitraire, induit une perte importante d'information. De plus, il est probable que certains facteurs de risque agissent de manière différente sur le passage vers des états d'incapacité légère et sur le passage vers des incapacités plus lourdes. Au cours du temps, la trajectoire empruntée varie d'une personne à l'autre, par exemple, certaines entrent directement dans un état d'incapacité sévère alors que d'autres peuvent rester plusieurs années dans des états légers avec parfois des périodes de retour à un état normal. Les mécanismes aboutissant à ces différents profils sont probablement différents. Il nous a donc paru intéressant d'essayer de modéliser l'impact des facteurs de risque sur les différents stades d'incapacité. Pour cela, nous prenons en compte le caractère ordonné de notre variable réponse en utilisant les extensions du modèle logistique aux données ordinales. Plusieurs modèles logistiques ont été proposés dans le cas simple des réponses ordinales indépendantes et nous nous attacherons d'abord à les détailler [74, 75]. Nous verrons ensuite comment certains d'entre eux peuvent être adaptés aux cas des réponses ordinales corrélées. Pour cela, nous utiliserons le modèle à effets aléatoires qui, nous l'avons déjà montré, est plus adapté à nos données de cohorte. En effet les arguments en faveur du modèle à effets aléatoires, développés dans la comparaison avec le modèle marginal pour les réponses binaires, répétées restent les mêmes dans le cas de réponses ordinales [76].

De manière plus large, les échelles ordinales sont fréquemment utilisées en épidémiologie et recherche clinique pour évaluer des dimensions telles que la qualité de vie, la fatigue, la dépression, les problèmes cognitifs. Parallèlement, avec le développement des études de cohorte, ces échelles sont recueillies de manière répétée dans le temps. Dans ce contexte, la mise à disposition de méthodes d'analyse de ces données ordinales répétées présente un important intérêt et notre étude peut être facilement transposée dans d'autres domaines de recherche.

## 1.1 Modèles logistiques, ordinaux sans effet aléatoire

Dans cette partie, nous avons réuni les différents modèles décrits dans la littérature pour analyser des réponses ordinales indépendantes. Tous ces modèles peuvent être estimés par la méthode du maximum de vraisemblance.

### 1.1.1 Modèle logistique polytomique

Le modèle logistique polytomique (ou multinomial) [77] n'est pas adapté aux réponses ordinales mais il constitue la base de plusieurs modèles présentés ci-dessous. Nous le décrivons donc dans ce paragraphe et introduisons les notations utilisées.

Ce modèle est une extension directe du modèle logistique pour réponses binaires au cas des réponses catégorielles. Dans ce modèle, l'ordre des modalités de la réponse n'est pas pris en compte, on considère simplement une catégorie de référence et on modélise la probabilité de survenue de chacune des autres catégories par rapport à cette catégorie de référence.

$$\log \left( \frac{P(Y_i = c | \mathbf{X}_i)}{P(Y_i = 0 | \mathbf{X}_i)} \right) = \alpha_c + \mathbf{X}_i' \boldsymbol{\beta}_c$$

$i$  : indice du sujet

$c$  : indice de la modalité de la réponse

$\mathbf{X}_i$  : vecteur des  $p$  covariables

$\boldsymbol{\beta}_c$  : vecteur des  $p$  paramètres pour la modalité  $c$

Il faut noter que dans ce modèle, pour chaque covariable,  $\mathbf{X}_i$ , le paramètre  $\beta_{lc}$  dépend de la catégorie  $c$  de la réponse. L'estimation de ce modèle est disponible dans plusieurs logiciels dont STATA avec la commande `mlogit` et SAS avec la procédure `logistic` assortie de l'option `link=glogit`.

### 1.1.2 Modèle à odds proportionnels

Dans ce modèle appelé aussi "modèle logistique cumulatif" [78], l'ordre des modalités de la réponse est pris en compte mais l'effet d'une covariable est supposé le même pour le passage d'un niveau de réponse au niveau suivant quels que soient ces niveaux.

$$\log\left(\frac{P(Y_i \geq c | \mathbf{X}_i)}{P(Y_i < c | \mathbf{X}_i)}\right) = \alpha_c + \mathbf{X}_i' \boldsymbol{\beta}$$

avec la contrainte :  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_{K-1}$

On suppose donc que les paramètres des covariables,  $\beta_l$ , sont indépendants du niveau des réponses,  $c$ . Seules les ordonnées à l'origine,  $\alpha_c$ , absorbent le caractère ordinal de la réponse ce qui les contraint à être ordonnées. Cette hypothèse implique que le log-odds ratio pour une covariable donnée est identique pour tous les niveaux de la réponse. Ceci a conduit McCullagh à nommer ce modèle "proportional odds model" [79].

L'aspect cumulatif de ce modèle induit plusieurs propriétés pratiques : 1) le modèle à odds proportionnels est invariant si les codes de la réponse,  $\mathbf{Y}$ , sont inversés et seul le signe des paramètres de régression est changé, 2) le modèle à odds proportionnels est invariant si les catégories adjacentes de la réponse ordinale sont regroupées [74]. Le vecteur des paramètres,  $\boldsymbol{\beta}$ , reste inchangé, seules les ordonnées à l'origine,  $\alpha_c$ , sont modifiées.

Il faut remarquer que le modèle est souvent présenté dans une écriture descendante mais cela ne fait que modifier le signe des paramètres :

$$\log\left(\frac{P(Y_i \leq c | \mathbf{X}_i)}{P(Y_i > c | \mathbf{X}_i)}\right) = -\log\left(\frac{P(Y_i \geq c | \mathbf{X}_i)}{P(Y_i < c | \mathbf{X}_i)}\right) = -\alpha_c - \mathbf{X}_i' \boldsymbol{\beta}.$$

Ce modèle est souvent présenté comme destiné à modéliser une réponse ordinaire avec des catégories liées de façon monotone à une variable latente, continue, sous-jacente,  $\mathbf{Z}$ . Les  $\alpha_c$  sont alors les points de coupure sur  $\mathbf{Z}$  et  $Y_i$  est égal à  $c$  si  $Z_i$  est compris entre  $\alpha_c$  et  $\alpha_{c+1}$ . Cependant, comme l'indique McCullagh [27], l'existence de cette variable latente est, en pratique, souvent invérifiable.

Ce modèle est disponible dans de nombreux logiciels (commande `ologit` de STATA et procédure `logistic` avec l'option `link=logit` dans SAS).

Les hypothèses de proportionnalité très restrictives de ce modèle ont conduit certains auteurs à proposer des modèles partiellement proportionnels avec des contraintes plus ou moins fortes.

### 1.1.3 Modèle non contraint, à odds partiellement proportionnels

Ce modèle [80] permet à un sous-groupe de  $q$  covariables d'avoir des odds non proportionnels selon les niveaux de la réponse. Il permet éventuellement de tester la proportionnalité pour ce sous-groupe.

$$\log\left(\frac{P(Y_i \geq c | \mathbf{X}_i)}{P(Y_i < c | \mathbf{X}_i)}\right) = \alpha_c + \sum_{l=1}^q (\beta_l + \gamma_{cl}) X_{il} + \sum_{l=q+1}^p \beta_l X_{il}$$

avec :  $q$  covariables avec des odds ratios non proportionnels ( $q < p$ )

et  $p-q$  covariables avec des odds ratios proportionnels.

Au total ce modèle a  $(K+q \times K+p)$  paramètres.

Pour la covariable  $l$ , la partie  $\gamma_{cl}X_{il}$ , représente l'incrément du risque associé uniquement au niveau  $c$ , en plus du risque commun à tous les niveaux de la réponse. Le test  $\gamma_{cl}=0$ , permet de tester l'hypothèse de proportionnalité sur cette covariable. Quand tous les  $\gamma_{cl}$  sont égaux à zéro, on retrouve le modèle précédent à odds proportionnels.

#### 1.1.4 Modèle contraint, à odds partiellement proportionnels

Un autre modèle plus contraignant a été proposé [80] pour le cas où les données ne satisferaient pas l'hypothèse des odds proportionnels mais où une augmentation monotone des logarithmes des odds ratios existe pour les différents niveaux de réponse.

$$\log\left(\frac{P(Y_i \geq c | \mathbf{X}_i)}{P(Y_i < c | \mathbf{X}_i)}\right) = \alpha_c + \sum_{l=1}^p \beta_l X_{il} + \tau_c \sum_{l=1}^q \gamma_l X_{il}$$

Comme précédemment, on considère que pour un sous-groupe de q covariables les odds ne sont pas proportionnels selon les niveaux de la réponse mais qu'une relation linéaire existe avec les logarithmes des odds ratios.

#### 1.1.5 Modèle logistique, ordonné, généralisé

Ce modèle est la généralisation la plus large du modèle proportionnel. L'effet de toutes les variables explicatives varie selon le point de coupure sur les catégories de la variable dépendante.

$$\log\left(\frac{P(Y_i \geq c | \mathbf{X}_i)}{P(Y_i < c | \mathbf{X}_i)}\right) = \alpha_c + \sum_l \beta_{cl} X_{il}$$

avec  $X \beta_K \leq X \beta_{K-1} \leq \dots \leq X \beta_2 \leq X \beta_1$   
et  $\alpha_K \leq \alpha_{K-1} \leq \dots \leq \alpha_2 \leq \alpha_1$

Ce modèle est disponible dans STATA avec la commande "gologit".

Comme le modèle à odds proportionnels est emboîté dans ce modèle généralisé, on peut tester l'hypothèse de proportionnalité en comparant ces deux modèles. La différence sur les déviances ( $-2 \log$  vraisemblance) entre les deux modèles fournit un test de l'hypothèse d'égalité des paramètres. Elle est distribuée selon une loi du  $\chi^2$  avec comme degré de liberté la différence du nombre de paramètres entre les deux modèles emboîtés.

#### 1.1.6 Modèle Stéréotype

Dans ce modèle construit à partir du modèle logistique polytomique, l'ordre des modalités de la réponse est pris en compte et il y a une catégorie de référence [81]. L'effet d'une covariable

sur la réponse  $c$  par rapport à la référence est décomposé en un effet global de la covariable plus un effet particulier à cette catégorie  $c$ , commun à toutes les covariables.

$$\log\left(\frac{P(Y_i = c | \mathbf{X}_i)}{P(Y_i = 0 | \mathbf{X}_i)}\right) = \alpha_c + \phi_c \sum_l \beta_l X_{il}$$

avec  $\phi_0 = 0$  et  $\phi_1 \leq \phi_2 \leq \dots \leq \phi_K = 1$

Ce modèle est disponible dans STATA avec la commande "solog".

Il s'agit d'un modèle stéréotype à une dimension où une seule combinaison de covariables est nécessaire pour faire la distinction entre les différents niveaux de la réponse  $\mathbf{Y}$ . Les paramètres  $\phi_c$  indiquent la distance entre les prédicteurs linéaires des niveaux de réponse.

Si au contraire il faut, par exemple, une combinaison de prédicteurs pour distinguer les premiers niveaux et une deuxième combinaison pour distinguer les niveaux suivants de la réponse on utilisera un modèle stéréotype à deux dimensions qui s'écrit :

$$\log\left(\frac{P(Y_i = c | \mathbf{X}_i)}{P(Y_i = 0 | \mathbf{X}_i)}\right) = \alpha_c + \phi_{c1} \sum_l \beta_{l1} X_{il} + \phi_{c2} \sum_l \beta_{l2} X_{il}$$

Un modèle stéréotype de dimension  $(K-1)$  est équivalent au modèle logistique polytomique. Le modèle logistique polytomique est le plus complet et fournit le meilleur ajustement possible mais au prix d'un grand nombre de paramètres tandis que le modèle stéréotype (emboîté dans le précédent) permet de réduire ce nombre de paramètres.

L'étude de ce modèle a donné lieu à l'introduction des concepts de "dimensionnalité" et de "distinguableté". Le premier concept fait référence au nombre de combinaisons de covariables nécessaire pour différencier les réponses. Le second indique si les covariables du modèle permettent de faire la distinction entre tous les niveaux de  $\mathbf{Y}$ , c'est-à-dire si certains  $\phi_c$  sont égaux [81, 82].

Ce modèle est adapté au cas où les catégories de la variable réponse ne correspondent pas à un ordre selon une dimension unique (variable latente) mais sont le résultat d'un agrégat de scores provenant de plusieurs dimensions. Les exemples de modèles stéréotypes

multidimensionnels sont peu fréquents dans la littérature. Lunt [82] présente l'analyse d'une étude sur des patients atteints de polyarthrite inflammatoire et classés en 6 stades selon la gravité des lésions articulaires. Il distingue deux groupes de variables agissant de manière différenciée sur les stades de gravité.

### 1.1.7 Modèle de ratios de continuation (Continuation ratio model)

Ce modèle s'intéresse à la probabilité d'être dans la catégorie  $c$  conditionnellement à être dans une catégorie inférieure à  $c$ .

$$\log\left(\frac{P(Y_i = c | \mathbf{X}_i)}{P(Y_i < c | \mathbf{X}_i)}\right) = \alpha_c + \sum_l \beta_l X_{il}$$

Si la fonction de lien "logit" est remplacée par la fonction de lien "complémentaire log-log", ce modèle est alors celui de Cox pour des données de survie avec des temps discrets [83].

$$\log\left[-\log\left(\frac{P(Y_i = c | \mathbf{X}_i)}{P(Y_i < c | \mathbf{X}_i)}\right)\right] = \alpha_c + \sum_l \beta_l X_{il}$$

Ce modèle peut être estimé dans STATA avec la commande "ocratio" et dans SAS avec la procédure logistic et l'option link=cloglog. Il est recommandé lorsque le processus sous-jacent est irréversible dans le sens où, lorsqu'un sujet a atteint un niveau de réponse, il ne peut retourner à un niveau plus bas [84-86].

Ce modèle ne possède pas les propriétés d'invariance : il n'est pas invariant ni lorsqu'on change l'ordre des catégories de la variable réponse ni lorsqu'on regroupe les catégories.

Comme pour le modèle à odds proportionnels, certains auteurs [84, 85] ont introduit plus de flexibilité en définissant pour chaque covariable des paramètres spécifiques à la catégorie de réponse. Les modèles peuvent être non contraints si pour toutes les covariables, on définit des paramètres spécifiques aux catégories ou bien partiellement contraints si seulement un sous-groupe de covariables a des paramètres spécifiques aux catégories.



### 1.1.8 Modèle logistique sur catégories adjacentes (adjacent-category logistic model)

Ce modèle s'intéresse au quotient de la probabilité d'avoir une réponse à un niveau  $c$  par rapport à la probabilité du niveau adjacent.

$$\log\left(\frac{P(Y_i = c | \mathbf{X}_i)}{P(Y_i = c-1 | \mathbf{X}_i)}\right) = \alpha_c + \sum_l \beta_{cl} X_{il}$$

Les paramètres des covariables dépendent du niveau de la réponse.

Comme le modèle précédent, il est recommandé pour les processus irréversibles.

En résumé, dans le tableau suivant nous donnons le nombre de paramètres à estimer pour chaque modèle lorsque nous considérons  $p$  prédicteurs et  $k$  niveaux de réponse.

**Tableau 14 - Nombre de paramètres pour certains modèles ordinaux.**

Modèle	$\alpha$	$\beta$	$\phi$
Odds proportionnels	$k$	$+p$	
Ordonné généralisé	$k$	$+kp$	
Continuation ratio	$k$	$+p$	
Catégories adjacentes	$k$	$+kp$	
Stéréotype 1-dimension	$k$	$+p$	$+(k-1)$
Stéréotype $d$ -dimensions	$k$	$+dp$	$+d(k-d)$
Polytomique	$k$	$+kp$	

## 1.2 Modèles à effets aléatoires

Dans cette partie, nous décrivons comment les modèles précédents peuvent être étendus au cas des réponses répétées grâce aux modèles à effets aléatoires. Nous nous limiterons aux modèles les plus adaptés à nos données. Nous montrerons comment le choix d'un modèle impose aussi des contraintes sur la structure des effets aléatoires.

Comme nous l'avons vu, certains modèles (modèle de ratios de continuation et modèle sur catégories adjacentes) sont plus spécialement destinés aux phénomènes irréversibles ce qui n'est pas le cas de l'incapacité chez les personnes âgées, quel que soit l'indicateur considéré. Nous laisserons donc ces deux modèles de côté dans cette partie.

### 1.2.1 Modèle logistique polytomique à effets aléatoires

Le modèle peut s'écrire de la façon suivante :

$$\log \left( \frac{P(Y_{ij} = c | \mathbf{X}_{ij}, \mathbf{U}_{ic})}{P(Y_{ij} = 0 | \mathbf{X}_{ij}, \mathbf{U}_{ic})} \right) = \alpha_c + \mathbf{X}_{ij}' \boldsymbol{\beta}_c + \mathbf{Z}_{ij}' \mathbf{U}_{ic}$$

Les  $\mathbf{U}_{ic}$  sont les effets aléatoires spécifiques du sujet  $i$  et de la catégorie de la variable dépendante  $c$ . Dans le cas de réponses nominales où chaque catégorie correspond à une dimension différente, il n'y a pas de raison d'attendre que les effets aléatoires soient similaires pour des valeurs différentes de  $c$  [87]. Pour que la définition du modèle soit complète, il faut donner la distribution de  $\mathbf{U}'_i = (\mathbf{U}'_{i1}, \dots, \mathbf{U}'_{i,K-1})$ . On considère en général une distribution multinormale, indépendante, avec une moyenne nulle et une matrice de covariance  $\boldsymbol{\Sigma}$ .

En pratique, comme dans le modèle logistique pour réponse binaire, le temps de calcul augmente rapidement avec le nombre d'effets aléatoires inclus dans le modèle, les procédures d'estimation numériques peuvent alors échouer. Nous allons donc nous restreindre au cas d'un modèle avec seulement des ordonnées à l'origine aléatoires.

$$\log \left( \frac{P(Y_{ij} = c | \mathbf{X}_{ij}, u_{ic})}{P(Y_{ij} = 0 | \mathbf{X}_{ij}, u_{ic})} \right) = \alpha_c + \mathbf{X}_{ij}' \boldsymbol{\beta}_c + u_{ic}$$

$u_{ic}$  est un paramètre individuel, caractéristique de la catégorie  $c$ , constant au cours du temps.

Pour un sujet donné, quel que soit l'intervalle de temps séparant deux réponses, la force de la dépendance est identique.

Le programme SAS s'écrit alors :

```
title1 ' Polytomique ' ;
title2 ' DELAI ' ;
```

```

proc nlmixed data=table cmax=5000 ;

eta1 = alpha1 + beta11*delai + u1 ;
eta2 = alpha2 + beta12*delai + u2 ;
eta3 = alpha3 + beta13*delai + u3 ;
eta4 = alpha4 + beta14*delai + u4 ;

if (iadl3=0) then
z=exp(eta1) / (exp(eta1)+exp(eta2)+exp(eta3)+exp(eta4)+1) ;
else if (iadl3=1) then
z=exp(eta2) / (exp(eta1)+exp(eta2)+exp(eta3)+exp(eta4)+1) ;
else if (iadl3=2) then
z=exp(eta3) / (exp(eta1)+exp(eta2)+ exp(eta3)+exp(eta4)+1) ;
else if (iadl3=3) then
z=exp(eta4) / (exp(eta1)+exp(eta2)+ exp(eta3)+exp(eta4) +1) ;
else if (iadl3=4) then
z=1 / (exp(eta1)+exp(eta2)+ exp(eta3)+exp(eta4)+1) ;

ll= log(z) ;
model iadl3 ~ general(ll) ;
random u1 u2 u3 u4 ~ normal([0,0,0,0], [s2u1,0,0,0, S2u2,0,0, s2u3,0, S2u4])
subject=sujet;

run ;
title ;

```

## 1.2.2 Modèle logistique à odds proportionnels et effets aléatoires

Dans le modèle à odds proportionnels, la nature ordinaire de la réponse est prise en compte en considérant les probabilités cumulées  $P(Y_{ij} \geq c)$ , ordonnées de la façon suivante :

$$1 = P(Y_{ij} \geq 0) \geq P(Y_{ij} \geq 1) \geq \dots P(Y_{ij} \geq c) \geq \dots P(Y_{ij} \geq K).$$

Si nous considérons uniquement un modèle à ordonnée à l'origine aléatoire, le modèle peut s'écrire :

$$\log \left( \frac{P(Y_{ij} \geq c | \mathbf{X}_{ij}, u_i)}{P(Y_{ij} < c | \mathbf{X}_{ij}, u_i)} \right) = \alpha_c + \mathbf{X}_{ij}' \boldsymbol{\beta} + u_i$$

avec :  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_{k-1}$

Dans ce modèle, les effets aléatoires ( $u_i$ ) et les effets fixes ( $\boldsymbol{\beta}$ ) sont indépendants du niveau de la réponse ( $c$ ). L'utilisation d'effets aléatoires non liés à la catégorie de la réponse s'appuie sur la notion qu'il existe un phénomène unique et continu, sous-jacent à la réponse ordinaire. Il serait possible de considérer un modèle avec des effets aléatoires,  $u_{ic}$ , variant avec la catégorie de la réponse. Mais l'estimation est alors plus complexe et des contraintes supplémentaires

doivent être introduites pour conserver l'ordre des ordonnées à l'origine. En effet, pour une catégorie donnée, l'ordonnée à l'origine est alors  $\alpha_c + u_{ic}$  et l'ordre suivant doit être respecté :

$$\alpha_1 + u_{i1} \geq \alpha_2 + u_{i2} \geq \dots \geq \alpha_K + u_{iK}.$$

Le programme SAS correspondant est le suivant :

```

title1 ' proportional odds model ' ;
title2 ' DELAI ' ;

proc nlmixed data=table qmax=5000 ;      * descendant : Pr(y>=j)/Pr(y<j) ;
bounds i1>0, i2>0, i3>0 ;
eta = beta0 + beta1*delai + beta2*age + u ;

if (iadl3=4) then z=1/(1+exp(-eta)) ;
else if (iadl3=3) then z= 1/(1+exp(-i1-eta)) - 1/(1+exp(-eta)) ;
else if (iadl3=2) then z= 1/(1+exp(-i1-i2-eta)) - 1/(1+exp(-i1-eta)) ;
else if (iadl3=1) then z= 1/(1+exp(-i1-i2-i3-eta)) - 1/(1+exp(-i1-i2-eta)) ;
else if (iadl3=0) then z= 1 - 1/(1+exp(-i1-i2-i3-eta)) ;

ll=log(z) ;
model iadl3 ~ general(ll);
random u ~ normal(0,s2u) subject=sujet;

estimate 'int3' beta0+i1 ;
estimate 'int2' beta0+i1+i2 ;
estimate 'int1' beta0+i1+i2+i3 ;
run;

```

### 1.2.3 Modèle logistique, ordonné, généralisé avec effets aléatoires

Comme dans le cas de réponses non corrélées, l'hypothèse d'odds proportionnels peut s'avérer trop stricte et doit de toute façon être testée. Pour cela, on peut considérer le modèle ordinal généralisé suivant :

$$\log \left( \frac{P(Y_{ij} \geq c | \mathbf{X}_{ij}, u_i)}{P(Y_{ij} < c | \mathbf{X}_{ij}, u_i)} \right) = \alpha_c + \mathbf{X}'_{ij} \boldsymbol{\beta}_c + u_i$$

Dans le modèle ordinal à odds proportionnels, le caractère ordonné de la réponse s'exprime uniquement à travers l'ordre des ordonnées à l'origine,  $\alpha_c$ . Dans ce nouveau modèle, les covariables participent à l'ordre des réponses. Le classement des probabilités cumulées induit la même chose sur les logits :  $\text{logit}(P(Y_{ij} \geq I)) \geq \dots \geq \text{logit}(P(Y_{ij} \geq c)) \geq \dots \geq \text{logit}(P(Y_{ij} \geq K-I))$ . Pour une seule covariable et des effets aléatoires indépendants du niveau de la réponse, cette relation devient :  $\alpha_1 + \beta_1 x \geq \alpha_2 + \beta_2 x \geq \dots \geq \alpha_{K-1} + \beta_{K-1} x$ . Ces K-1 droites de régression ne sont

pas forcément parallèles et peuvent se croiser pour certaines valeurs de  $x$ , ce qui aboutit à la violation de la hiérarchie pré-définie. Pour des covariables dichotomiques, les croisements de ces droites de régression surviennent en dehors des valeurs admissibles (c'est-à-dire  $< 0$  ou  $> 1$ ) [88].

Cependant pour les covariables continues ces croisements peuvent survenir à l'intérieur du spectre des valeurs observables. Cela peut poser un problème si le modèle doit être extrapolé et l'on doit systématiquement vérifier que l'ordre entre les prédicteurs linéaires est respecté. Une solution lorsque le croisement intervient au mauvais endroit pour une covariable continue sur laquelle l'hypothèse de proportionnalité des odds n'est pas valide, est de la transformer en  $m$  classes et de la remplacer dans le modèle par  $m-1$  variables dichotomiques indicatrices. C'est pour cette raison que les effets aléatoires supposés provenir d'une distribution normale continue ne sont en général pas considérés dépendants des niveaux de la réponse [88].

Si l'hypothèse de proportionnalité des odds est rejetée, seulement pour une partie des covariables, on obtient alors un modèle partiellement proportionnel de la forme :

$$\log\left(\frac{P(Y_{ij} \geq c | \mathbf{X}_{ij}, u_i)}{P(Y_{ij} < c | \mathbf{X}_{ij}, u_i)}\right) = \alpha_c + \mathbf{X}_{ij}^1 \boldsymbol{\beta}_c^1 + \mathbf{X}_{ij}^2 \boldsymbol{\beta}^2 + u_i$$

$\mathbf{X}_{ij}^1$ : la sous matrice de  $\mathbf{X}_{ij}$  des  $q$  covariables avec des odds non proportionnels

$\mathbf{X}_{ij}^2$ : la sous matrice de  $\mathbf{X}_{ij}$  des  $p-q$  covariables avec des odds proportionnels

```

title1 ' Generalized ordered logit model ' ;
title2 ' DELAI, AGE ' ;

proc nlmixed data=table qmax=5000 ;      * descendant : Pr(y>=j)/Pr(y<j) ;
    bounds i1>0, i2>0, i3>0 ;
    eta1 = beta0 + beta11*delai + beta21*age+ u ;
    eta2 = beta0 + i1 + beta12*delai + beta22*age + u ;
    eta3 = beta0 + i1 + i2 + beta13*delai + beta23*age + u ;
    eta4 = beta0 + i1 + i2 + i3 + beta14*delai + beta24*age + u ;

    if (iadl3=4) then z=1/(1+exp(-eta1)) ;
    else if (iadl3=3) then z= 1/(1+exp(-eta2)) - 1/(1+exp(-eta1)) ;
    else if (iadl3=2) then z= 1/(1+exp(-eta3)) - 1/(1+exp(-eta2)) ;
    else if (iadl3=1) then z= 1/(1+exp(-eta4)) - 1/(1+exp(-eta3)) ;
    else if (iadl3=0) then z= 1 - 1/(1+exp(-eta4)) ;
    ll=log(z) ;
model iadl3 ~ general(ll) ;

```

```

random u ~ normal(0,s2u) subject=sujet;

estimate 'int3' beta0+i1 ;
estimate 'int2' beta0+i1+i2 ;
estimate 'int1' beta0+i1+i2+i3 ;
run;

```

## 1.2.4 Modèle Stéréotype avec effets aléatoires

Dans le cas où la réponse ordinaire ne pourrait être considérée comme une version discrète d'une variable continue sous-jacente, le modèle stéréotype à effets aléatoires peut être utilisé :

$$\log\left(\frac{P(Y_{ij} = c | \mathbf{X}_{ij}, u_i)}{P(Y_{ij} = 0 | \mathbf{X}_{ij}, u_i)}\right) = \alpha_c + \phi_c \mathbf{X}_{ij}' \boldsymbol{\beta} + u_i$$

avec  $0 < \phi_1 < \phi_2 < \dots < \phi_{k-1} = 1$

```

title1 ' Stereotype' ;
title2 ' DELAI, AGE' ;

proc nlmixed data=table ;

bounds 0 < epsy2-epsy4 <= 0.9 ;
parms beta1=0.5 beta2=-0.229 epsy2=0.1 epsy3=0.1 epsy4=0.1 ;

eta1 = alpha1 + (beta1*delai + beta2*age ) + u ;
eta2 = alpha2 + (1-epsy2)*(beta1*delai + beta2*age ) + u ;
eta3 = alpha3 + (1-epsy2-epsy3)*(beta1*delai + beta2*age) + u ;
eta4 = alpha4 + (1-epsy2-epsy3-epsy4)*(beta1*delai + beta2*age) + u ;

if (iadl2=0) then
z=exp(eta1) / (exp(eta1)+exp(eta2)+ exp(eta3)+exp(eta4) +1) ;
else if (iadl2=1) then
z=exp(eta2) / (exp(eta1)+exp(eta2)+ exp(eta3)+exp(eta4) +1) ;
else if (iadl2=2) then
z=exp(eta3) / (exp(eta1)+exp(eta2)+ exp(eta3)+exp(eta4) +1) ;
else if (iadl2=3) then
z=exp(eta4) / (exp(eta1)+exp(eta2)+ exp(eta3)+exp(eta4) +1) ;
else if (iadl2=4) then
z=1 / (exp(eta1)+exp(eta2)+ exp(eta3)+exp(eta4) +1) ;

ll= log(z) ;
model iadl2 ~ general(ll) ;
estimate 'phi2' (1-epsy2);
estimate 'phi3' (1-epsy2-epsy3);
estimate 'phi4' (1-epsy2-epsy3-epsy4);
random u ~ normal(0,s2u) subject=sujet;

run ;

```

### 1.2.5 Méthodes d'estimation des modèles logistiques ordinaux à effets aléatoires

Comme pour les réponses binaires corrélées, les méthodes numériques d'intégration de la vraisemblance décrite partie II, paragraphe 1.4 sont les plus efficaces.

## 1.3 Sélection des modèles et critères d'adéquation aux données ordinales

Concernant la sélection des modèles, les estimations basées sur la vraisemblance telles que celles utilisées dans NLMIXED, permettent de comparer aisément les modèles. On peut utiliser comme dans le cas simple des réponses non corrélées, le critère d'information de Akaike (AIC pour Akaike's Information Criterion) ou le critère bayésien de Schwarz. Pour des modèles emboîtés, le test du rapport de vraisemblance peut être utilisé pour comparer la qualité des modèles (cf. partie II, paragraphe 1.2).

Cependant, plus généralement, comme dans le cas non longitudinal, on peut évaluer l'adéquation du modèle aux données, en étudiant les écarts entre les valeurs prédites et les valeurs observées. Dans le modèle linéaire généralisé, les valeurs prédites sont calculées à partir des estimations du maximum de vraisemblance du vecteur des effets fixes ( $\hat{\beta}$ ) et des estimations individuelles des effets aléatoires qui sont réduits, dans notre exemple, aux estimations des ordonnées à l'origine aléatoires ( $\hat{u}_i$ ). Les effets aléatoires peuvent être estimés par les méthodes empiriques de Bayes [89]. La valeur prédite est alors une probabilité spécifique du sujet pour des valeurs particulières des covariables fixes et pour une ordonnée à l'origine aléatoire spécifique à un sujet. Les valeurs prédites peuvent être obtenues directement dans la procédure NLMIXED de SAS avec l'instruction PREDICT. La concordance entre les réponses ordinales prédites et observées peut être évaluée par une des mesures standards d'association : gamma, tau-b de Kendall [28]. Cependant Vonesh a suggéré

d'utiliser un coefficient de détermination et un coefficient de corrélation et de concordance [90]. Le coefficient de détermination est défini comme suit :

$$R^2 = 1 - \frac{\sum_i \sum_j (Y_{ij} - \hat{Y}_{ij})^2}{\sum_i \sum_j (Y_{ij} - \bar{Y})^2}.$$

Ce coefficient peut être interprété comme la portion de la variation totale de  $\mathbf{Y}$  qui est prise en compte par le modèle.

Le second coefficient s'écrit de la manière suivante :

$$r_c = 1 - \frac{\sum_i \sum_j (Y_{ij} - \hat{Y}_{ij})^2}{\sum_i \sum_j (Y_{ij} - \bar{Y})^2 + \sum_i \sum_j (\hat{Y}_{ij} - \bar{Y})^2 + N(\bar{Y} - \bar{\hat{Y}})^2}$$

Ce coefficient mesure, sur le graphique de  $Y_{ij}$  versus  $\hat{Y}_{ij}$ , la proximité des points avec la première bissectrice. Les valeurs possibles varient de  $-1$  à  $+1$ , une modélisation parfaite correspond à la valeur 1, tandis qu'une modélisation mal adaptée correspond à des valeurs inférieures à 0.

Le principal avantage de ces critères est qu'ils peuvent s'appliquer à des modèles non emboîtés et qu'il n'est pas nécessaire de définir un modèle complet pour évaluer si le modèle choisi s'ajuste aux données. Un autre avantage est que le coefficient de concordance n'a pas besoin que la fonction de vraisemblance soit spécifiée. Il peut donc s'appliquer à la comparaison d'estimations provenant de procédure utilisant par exemple des pseudo vraisemblances. Par contre, ces coefficients sont conçus pour des réponses continues ou ordinales et sont mal adaptés aux réponses binaires.

## 1.4 Conclusion

Ces modèles logistiques mixtes, pour réponses ordinales, peuvent être appliqués plus généralement à de nombreux types de données ordinales corrélées. Cependant, la plupart des



modèles utilisés font l'hypothèse d'odds proportionnels, c'est-à-dire d'un effet homogène des covariables sur toutes les catégories de l'échelle ordonnée considérée. Cette hypothèse n'est pas toujours licite et les exemples de covariables agissant de manière différenciée sur, par exemple, les stades d'une maladie sont assez fréquents [91-93]. Il est donc important d'utiliser dans un premier temps des modèles levant cette hypothèse et de tester la proportionnalité des odds. Cependant, les modèles autorisant des effets hétérogènes des covariables sur les niveaux de la réponse requièrent une taille d'échantillon assez grande. En effet, les observations sont à la fois partagées selon les K catégories de la réponse et selon les catégories des covariables. Il faut des effectifs suffisants dans chaque cellule et en particulier pour les catégories extrêmes de la réponse [91]. Notons aussi que les tests statistiques sur les paramètres formés à partir du quotient de l'estimation et de l'erreur standard sont basés, comme dans beaucoup d'autres techniques statistiques, sur des résultats asymptotiques ce qui implique aussi, de larges effectifs.

## 2 Exemple d'utilisation du GLMM pour réponses ordinales<sup>3</sup>

Dans cette partie, nous reprenons l'étude longitudinale des facteurs de risque d'incapacité mais en considérant maintenant la variable dépendante sous forme ordinale. Notre but est d'illustrer la recherche du type de modèle logistique mixte qui caractérise le mieux nos données sur quelques covariables typiques agissant de manière différente sur les niveaux de la réponse et non d'explorer l'ensemble de facteurs potentiels. Nous partons d'un modèle général avec de nombreux paramètres à estimer et nous montrons comment le réduire.

### 2.1 Matériel et méthodes

Notre variable réponse est constituée par le score AIVQ. Rappelons que cet indicateur d'incapacité varie de 0 à 8 en fonction du nombre d'activités que le sujet n'est plus en mesure d'effectuer sans assistance. Afin que les effectifs soient suffisants dans chaque catégorie, nous avons réuni les réponses comme suit : 0, 1-2, 3-4, 5-6, 7-8. Les réponses annuelles ont finalement cinq catégories variant de 0 (aucune incapacité) à 4 (presque entièrement dépendant).

L'analyse se limite aux femmes sans incapacité à l'inclusion et sans donnée manquante sur les covariables considérées, soit 929 femmes. Elle utilise les données d'incapacité évaluées par auto-questionnaire, sur la totalité des 7 ans de suivi.

Notre score réponse peut être considéré comme le reflet d'un phénomène continu sous-jacent : le déclin de l'état général des personnes âgées. Chaque activité composant le score, a été

---

<sup>3</sup> Ce chapitre a donné lieu à un article en révision dans la Revue d'Epidémiologie et de Santé Publique.

choisie pour rendre compte d'un état de dépendance plus ou moins avancé [58, 94]. Ces activités vont des plus complexes qui disparaissent en premier (téléphoner, remplir ses papiers, prendre ses médicaments) aux plus basiques (manger, s'habiller) en passant par des activités intermédiaires (préparer les repas, faire les courses). En général, chez les personnes âgées, un score d'AIVQ élevé signifie une perte d'autonomie pour les activités complexes et les activités de difficultés intermédiaires. Ces caractéristiques nous ont amené à considérer un modèle logistique cumulatif. Comme expliqué au paragraphe 1.4 de la partie III, nous commencerons par un modèle logistique ordonné généralisé et nous essayerons ensuite de le simplifier. Le délai depuis l'inclusion sera inclus dans le modèle sous forme de variable continue. Les covariables d'inclusion suivantes seront ajoutées au modèle multivarié en tant qu'effets fixes : âge à l'inclusion (en années au-delà de 74 ans), la vitesse de marche (en mètres/secondes, pour parcourir 6 mètres à allure normale), douleurs (douleurs fréquentes au bas du dos, aux hanches, aux genoux, aux chevilles ou aux jambes), temps pour se lever cinq fois d'une chaise (en secondes), acuité visuelle (échelle décimale). Toutes ces covariables sont considérées dans cette analyse comme quantitatives sauf la variable douleurs, qui est binaire. Comme dans le cas des réponses binaires, la procédure NLMIXED converge et donne des estimations fiables seulement lorsque l'initialisation des paramètres à estimer n'est pas trop éloignée de la solution finale. Nous avons donc calculé une succession de modèles en commençant par le plus simple comprenant seulement une ordonnée à l'origine aléatoire et en introduisant une à une les covariables. Les estimations calculées à chaque étape sont utilisées comme initialisation à l'étape suivante.

## 2.2 Sélection des modèles

L'évolution de l'incapacité est décrite dans le tableau 15. Le pourcentage de personnes sans incapacité décroît de 73,6% à la fin de la première année à 38,2% à la fin de la septième année. Inversement, le pourcentage de femmes incapables de faire seule 5 activités ou plus croît de 1,2% à 15,1%.

**Tableau 15 - Description de l'évolution des sujets par année de suivi**

	Année 1	Année 2	Année 3	Année 4	Année 5	Année 6	Année 7
Nombre de femmes avec une évaluation des IADL	927	907	879	820	702	665	683
Nombre de décès dans l'année précédente	1	11	20	30	22	33	26
Pourcentage d'incapacité							
IADL = 0	73,6	66,7	61,4	55,6	54,0	47,7	38,2
IADL = 1 or 2	21,5	23,0	27,0	26,5	26,2	25,2	29,7
IADL = 3 or 4	3,7	6,5	6,1	9,5	12,0	15,2	17,0
IADL = 5 or 6	0,6	2,2	2,7	4,3	5,0	5,7	8,4
IADL = 7 or 8	0,6	1,6	2,8	4,1	2,8	6,2	6,7
Délai depuis l'inclusion							
Médiane (ans)	0,99	1,98	2,97	3,95	4,97	6,10	7,00
Etendue (ans)	0,91; 1,48	1,95; 2,40	2,93; 3,36	3,87; 4,53	4,88; 5,89	5,88; 6,87	6,91; 8,19

Les résultats de l'estimation du modèle logistique ordonné généralisé complet sont donnés dans la première partie du tableau 16.

**Tableau 16 - Comparaison des estimations de trois modèles à effets aléatoires pour réponses ordinales.**

	Modèle logistique ordonné généralisé			Modèle logistique à odds partiellement proportionnel			Modèle stéréotype		
	Paramètre	ES*	p	Paramètre	ES*	p	Paramètre	ES*	p
Ordonnée à l'origine									
4	-6,89	1,08	< 0,0001	-6,93	0,83	< 0,0001	-5,02	0,83	< 0,0001
3	-5,51	0,93	< 0,0001	-6,04	0,80	< 0,0001	-4,55	0,76	< 0,0001
2	-4,82	0,83	< 0,0001	-4,82	0,78	< 0,0001	-3,40	0,70	< 0,0001
1	-2,05	0,77	0,008	-1,77	0,75	0,02	-1,71	0,46	0,02
Délai depuis l'inclusion (ans)									
4	0,58	0,05	< 0,0001	0,56	0,05	< 0,0001	0,80	0,04	< 0,0001
3	0,61	0,04	< 0,0001	0,61	0,04	< 0,0001			
2	0,65	0,03	< 0,0001	0,65	0,03	< 0,0001			
1	0,53	0,02	< 0,0001	0,52	0,02	< 0,0001			
Age (en années moins 74 ans)									
4	0,22	0,04	< 0,0001	0,22	0,04	< 0,0001	0,31	0,03	< 0,0001
3	0,24	0,03	< 0,0001	0,24	0,03	< 0,0001			
2	0,24	0,03	< 0,0001	"	"	"			
1	0,18	0,03	< 0,0001	0,18	0,03	< 0,0001			
Vitesse de marche (m/s)									
4	-3,32	0,81	< 0,0001	-2,92	0,59	< 0,0001	-4,10	0,63	< 0,0001
3	-3,32	0,68	< 0,0001	"	"	"			
2	-2,88	0,61	< 0,0001	"	"	"			
1	-2,36	0,56	< 0,0001	-2,57	0,56	< 0,0001			
Douleur									
4	-0,0004	0,25	0,99	0,24	0,20	0,23	0,66	0,19	0,0005
3	0,26	0,21	0,23	"	"	"			
2	0,63	0,19	0,0008	0,69	0,17	< 0,0001			
1	0,72	0,17	< 0,0001	"	"	"			
Temps pour 5 levers d'une chaise (sec)									
4	0,63	0,18	0,0005	0,62	0,14	< 0,0001	0,59	0,15	0,0001
3	0,43	0,17	0,01	"	"	"			
2	0,61	0,15	< 0,0001	"	"	"			
1	0,68	0,15	< 0,0001	"	"	"			
Acuité visuelle									
4	-0,09	0,05	0,06	-0,14	0,03	< 0,0001	-0,16	0,04	< 0,0001
3	-0,11	0,04	0,008	"	"	"			
2	-0,14	0,04	0,0003	"	"	"			
1	-0,15	0,03	< 0,0001	"	"	"			
Paramètres d'échelle									
$\phi_4$							1		
$\phi_3$							0,91	0,05	< 0,0001
$\phi_2$							0,83	0,04	< 0,0001
$\phi_1$							0,55	0,03	< 0,0001
Ecart type de l'effet aléatoire	2,19	0,08	< 0,0001	2,19	0,08	< 0,0001	2,24	0,10	< 0,0001

\* ES : Ecart Standard

Ce modèle général a été réduit en le comparant successivement avec des modèles faisant l'hypothèse d'odds proportionnels sur certaines covariables. Nous considérons chaque covariable séparément et si les estimateurs des paramètres pour certaines réponses adjacentes sont très proches (en prenant en compte l'erreur standard), nous remplaçons ces paramètres par un seul. Par exemple, les trois premiers paramètres de la covariable "vitesse de marche" peuvent être remplacés par un seul paramètre, mais le quatrième reste différent. A chaque étape, le test du rapport de vraisemblance est utilisé pour tester si le nouveau modèle restreint, est significativement différent du modèle complet initial. La comparaison du modèle final simplifié avec le modèle complet est donnée dans le tableau 17. La différence des déviations entre le modèle ordonné généralisé et le modèle à odds partiellement proportionnels est égale à 13. La différence du nombre de paramètres entre les deux modèles est de 11 (nombre de degrés de liberté). Le test du Chi2 correspondant donne une probabilité de 0,29 ce qui indique que les deux modèles ne sont pas significativement différents.

**Tableau 17 - Comparaison des performances des trois modèles logistiques, mixtes, ordonnés**

	Modèle logistique ordonné, généralisé	Modèle logistique à odds partiellement proportionnels	Modèle stéréotype
-2Log Vraisemblance	9756	9769	10208
AIC	9814	9805	10236
Coefficient de détermination	0,580	0,582	0,32
Coefficient de concordance et de corrélation	0,770	0,770	0,57

Pour les variables âge, vitesse de marche, douleurs, la proportionnalité des odds n'a été validée que pour une partie des catégories de la réponse. Par exemple, le log-odds ratio associé avec les douleurs est égal à 0,69 ( $p < 0,0001$ ) pour les deux premiers niveaux et à 0,24 ( $p = 0,23$ ) pour les deux suivants. Ce résultat montre que les douleurs ont un effet prononcé au

début de la perte des capacités, mais qu'ensuite elles n'ont pas d'effet sur le passage à des stades plus graves.

Pour les variables "temps pour se lever 5 fois d'une chaise" et acuité visuelle, le modèle simplifié n'a qu'un paramètre indiquant que les odds sont entièrement proportionnels.

Comme indiqué au paragraphe 1.2.3 de la partie III, nous avons vérifié la hiérarchie entre les prédicteurs linéaires estimés pour tous les sujets et tous les temps. Aucune inversion de l'ordre de ces prédicteurs n'a été trouvée. Le modèle final peut donc être considéré comme valide sur l'ensemble des valeurs observables des covariables.

Les deux modèles ont des coefficients de détermination (0,58) et des coefficients de concordance et de corrélation (0,77) identiques (tableau 17). Nous préférons donc le modèle à odds partiellement proportionnels en accord avec le principe de parcimonie sur les paramètres. Il est d'ailleurs plus intéressant puisqu'il peut indiquer à partir de quel niveau d'incapacité chaque covariable agit.

Nous avons aussi construit un modèle stéréotype avec effets aléatoires. Les estimations des paramètres de ce modèle sont données dans la dernière colonne du tableau 16. La catégorie de référence correspondait à une score IADL égal à zéro. Dans ce type de modèle, un seul paramètre caractérise chaque covariable. Le logarithme de l'odds ratio pour chaque niveau de réponse est obtenu en multipliant le paramètre estimé par le paramètre d'échelle correspondant. Par exemple le log-odds ratio pour le niveau 3 d'incapacité pour la covariable "douleur" est égal à 0,60 ( $0,66 \times 0,91$ ). Ce modèle paraît être moins performant que les modèles précédents : sa déviance ( $-2\text{Log vraisemblance}$ ) est de 10208 alors qu'elle est de 9769 pour le modèle à odds partiellement proportionnels (tableau 15). Cependant les modèles n'étant pas emboîtés, on ne peut les comparer avec le test du rapport de vraisemblance, par contre on peut utiliser le coefficient de détermination égal à 0,32 et le coefficient de concordance et de corrélation, égal à 0,57. Ces coefficients sont inférieurs aux précédents

(égaux respectivement à 0,58 et 0,77) ce qui permet de conclure que le modèle stéréotype est moins bon.

Le tableau 18 donne pour le modèle final à odds partiellement proportionnels, la comparaison entre les valeurs observées et les valeurs prédites sur l'ensemble des données recueillies pendant 7 ans soit 5583 observations. La proportion globale des paires concordantes, correspondant à la diagonale du tableau, est de 73,2%. Le Kappa pondéré (95%IC) est égal à 0,63 (0,61 ; 0,65). Ces résultats montrent une bonne concordance entre les réponses prédites et observées même si l'on constate une tendance à une légère sous estimation dans la prédiction du niveau d'incapacité.

**Tableau 18 - Comparaison entre les valeurs observées et prédites par le modèle à odds partiellement proportionnels\***

Niveau d'incapacité observé N=5583	Niveau d'incapacité estimé				
	0	1	2	3	4
0	<b>52,48</b>	5,39	0,16	0	0
1	8,60	<b>15,55</b>	1,18	0,02	0,02
2	0,64	5,16	<b>3,39</b>	0,09	0,14
3	0,16	1,43	1,63	<b>0,18</b>	0,45
4	0,05	0,45	1,13	0,11	<b>1,59</b>

\* Le tableau donne les pourcentages de sujets dans chacune des "cases"

Comme dans le cas des réponses binaires le logiciel WinBUGS peut être utilisé pour confirmer les estimations de NLMIXED. La comparaison des estimations données par les deux logiciels est présentée dans le tableau 19 pour le modèle à odds partiellement proportionnels.



**Tableau 19 - Comparaison des estimations obtenues avec WinBUGS et SAS pour le modèle à odds partiellement proportionnels**

	WinBUGS						SAS NLMIXED	
	moyenne	Écart type	MC erreur	2,5%	médiane	97,5%	Paramètre estimé	Erreur standard
Ordonnée à l'origine								
4	-7.077	0.5683	0.07536	-8.077	-7.076	-5.915	-6,93	0,83
3	-6,199	0,5068	0,07161	-7,016	-6,237	-5,105	-6,04	0,80
2	-4,961	0,4784	0,06887	-5,755	-4,982	-3,928	-4,82	0,78
1	-1,936	0,401	0,05699	-2,618	-1,956	-1,133	-1,77	0,75
Temps								
4	0,5691	0,04756	0,004757	0,4785	0,5697	0,6632	0,56	0,05
3	0,6177	0,03307	0,003044	0,5578	0,6162	0,6884	0,61	0,04
2	0,659	0,02644	0,002396	0,6066	0,6599	0,7148	0,65	0,03
1	0,529	0,02207	0,001786	0,4893	0,5287	0,5744	0,52	0,02
Age								
4	0,2169	0,03606	0,003943	0,148	0,2171	0,2829	0,22	0,04
3 et 2	0,2378	0,0293	0,003723	0,1791	0,2375	0,2962	0,24	0,03
1	0,1889	0,02782	0,003409	0,1397	0,1882	0,2443	0,18	0,03
Vitesse de marche								
4, 3, et 2	-2,92	0,3769	0,0518	-3,642	-2,872	-2,236	-2,92	0,59
1	-2,546	0,3575	0,04907	-3,294	-2,578	-1,886	-2,57	0,56
Douleur								
4 et 3	0,2735	0,1946	0,0166	-0,115	0,2739	0,6479	0,24	0,20
2 et 1	0,7245	0,1669	0,01601	0,41	0,7262	1,039	0,69	0,17
Temps pour 5 levers d'une chaise								
4, 3, 2, 1	0,609	0,125	0,01728	0,3278	0,6091	0,8717	0,62	0,14
Acuité visuelle								
4, 3, 2, 1	-0,1283	0,02956	0,003738	0,1882	0,1286	0,06584	-0,14	0,03
Précision pente	0,196	0,01476	9,618E-4	0,1687	0,1956	0,2263	2,19*	0,08

\* variance=1/précision

## 2.3 Qualité du modèle

Les coefficients de concordance donnés dans le paragraphe précédent ont été calculés en prenant en compte les estimations des effets fixes et les estimations des effets aléatoires. Ces coefficients peuvent être qualifiés de conditionnels dans la mesure où ils s'appliquent aux

réponses estimées conditionnellement aux effets aléatoires [90]. Cependant, on peut aussi calculer des coefficients de concordance moyennés en ne prenant en compte que les estimations issues de la partie fixe du modèle. Pour le modèle final à odds partiellement proportionnels, ce coefficient moyenné de concordance et de corrélation vaut 0,41. La comparaison avec le coefficient conditionnel (égal à 0,77) indique que les effets fixes n'expliquent qu'une partie de la variabilité des réponses. Une partie importante de l'hétérogénéité inter-individuelle est modélisée grâce à l'ordonnée à l'origine aléatoire.

## 2.4 Conclusion

Nous venons de démontrer que les modèles logistiques mixtes cumulatifs (modèle logistique ordonné généralisé ou modèle logistique à odds partiellement proportionnels) sont bien adaptés à l'analyse de l'évolution de l'incapacité évaluée en tant que réponse ordinaire répétée.

Par rapport à l'analyse de réponses binaires, la difficulté supplémentaire réside dans le choix du modèle logistique ordinal. En utilisant une méthode d'estimation basée sur la vraisemblance, nous avons pu comparer les divers modèles candidats et vérifier la supériorité des modèles cumulatifs pour nos données.

Cependant, cette méthode rend difficile l'introduction d'une structure de covariance complexe car le temps de calcul augmente rapidement avec le nombre d'effets aléatoires. Cette limitation tend néanmoins à devenir moins pénalisante au fur et à mesure que les capacités de calcul des micro-ordinateurs s'accroissent. Ainsi, dans notre exemple, le modèle logistique partiellement proportionnel final nécessitait 56 minutes de calcul sur un ordinateur PC, P4, 1,4GHz et seulement 12 minutes sur un PC, P4, 2,5MHz.

Dans la littérature, peu d'auteurs se sont intéressés aux critères d'évaluation et de diagnostic des modèles GLMM [95, 96]. C'est cependant un point très important et nous avons présenté dans le paragraphe précédent une méthode simple pour la comparaison des modèles ordinaux à effets aléatoires et pour l'évaluation de leur qualité d'ajustement.

Dans cette partie, nous nous sommes limités à l'étude de quelques covariables permettant d'illustrer notre propos. Mais, comme nous l'avons fait dans le cas de réponses binaires longitudinales (partie II, paragraphe 3), cette modélisation au moyen de modèles logistiques cumulatifs mixtes pourrait être appliquée à l'ensemble des facteurs de risque potentiels disponibles dans la cohorte Epidos. Ainsi, nous pourrions distinguer les facteurs selon qu'ils agissent préférentiellement sur les stades précoces ou sur les stades plus avancés de l'incapacité.

## Discussion générale

Dans notre étude des facteurs de risque conduisant à l'incapacité chez les personnes âgées, nous avons adopté les concepts définis dans la littérature anglo-saxonne dans les années 1980 et qui ont été présentés dans la première partie de cette thèse. En particulier, notre variable événement se réfère aux notions d'incapacité définies par Nagi ou de handicap ou désavantage social définis par Wood (cf. partie I, paragraphe 1). En effet, les indicateurs utilisés : aptitude à sortir de chez soi sans aide ou activités instrumentales de la vie quotidienne, s'inscrivent dans la définition donnée par Nagi de "limitations des performances dans les rôles sociaux et les tâches dans un environnement social et culturel", ou dans celle de Wood de "conséquences des incapacités fonctionnelles dans la vie sociale". Dans la partie II, paragraphe 3, nous avons mis en évidence de nombreux facteurs de risque menant à l'incapacité et construit un score prédictif. Une grande partie de ces facteurs de risque appartiennent à ce que Nagi a appelé déficiences et limitations fonctionnelles et à ce que Wood a appelé déficiences et incapacités fonctionnelles. Il s'agit des tests de mobilité, d'équilibre statique et dynamique, des mesures de force, de l'acuité visuelle. Nous avons aussi trouvé des facteurs de risque d'ordre personnel selon la définition de la nouvelle CIF comme le niveau d'éducation, des facteurs d'ordre comportementaux comme un indice de masse corporelle élevé ou le manque d'activité physique qui se situent plutôt dans la sphère "participation". Enfin, nous avons mis en évidence l'importance d'une santé subjective positive qui révèle là aussi, un état d'esprit optimiste, probablement plus tourné vers l'environnement social et moins vers les problèmes individuels de santé.

Comme nous l'avons indiqué dans la partie II, paragraphe 3.4, les facteurs de risque prédictifs de l'incapacité, trouvés dans notre étude peuvent être aussi interprétés à la lumière du concept plus récent de fragilité des personnes âgées.

La fragilité de la personne âgée est restée longtemps un concept peu précis, confondu souvent avec la morbidité, l'incapacité ou le très grand âge. Ces dernières années, plusieurs auteurs ont

cherché à clarifier ce concept et à définir des critères destinés à repérer les personnes âgées fragiles.

Tout le monde s'accorde pour définir la fragilité comme un syndrome biologique caractérisé par la diminution des réserves fonctionnelles, une plus grande susceptibilité à des stress extérieurs et une homéostasie instable provenant du déclin simultané de multiples systèmes physiologiques ce qui conduit à une plus grande vulnérabilité générale. En particulier, ce dysfonctionnement ne permet pas à la personne de recouvrer rapidement cet équilibre homéostatique après un événement déstabilisant tel qu'une maladie ou un traumatisme. Ce réseau hautement intégré implique de nombreux éléments : hormones, réponse immune, équilibre entre les radicaux libres et le mécanisme antioxydant, équilibre dynamique entre le tonus sympathique et parasympathique. Le rôle de l'inflammation chez les personnes très âgées semble très important et pourrait être impliqué dans la fragilité liée à l'âge.

Le concept de fragilité prend en considération l'ensemble des interactions entre les différents systèmes impliqués dans le fonctionnement du corps humain [61]. L'affaiblissement d'une fonction retentit en cascade sur plusieurs chaînes fonctionnelles. La fragilité n'est pas inévitable et elle peut être réversible grâce à des programmes d'interventions.

Fried et al [63] suggèrent deux trajectoires d'accès à la fragilité :

- des changements physiologiques dus à l'avancée en âge et n'ayant pas pour origine une maladie (sarcopénie (perte de masse maigre liée au vieillissement), anorexie, ...)
- une dégradation induite par une maladie sévère.

Cette distinction est reprise par Gill et al [97] et par Ferrucci et al [64, 98] qui décrivent deux types d'incapacité : l'incapacité insidieuse ou progressive et l'incapacité "catastrophique" induite par un événement aigu de santé tel qu'un AVC ou une fracture de la hanche. La première est plus fréquente chez les personnes très âgées. Ce déclin progressif passe par une phase de compensation au cours de laquelle la personne conserve les capacités de la vie

quotidienne malgré l'apparition de certaines limitations fonctionnelles ou handicaps [99]. Pour ces auteurs, la prévention de l'incapacité progressive est la plus intéressante en termes de santé publique. Elle s'adresse aux personnes fragiles pas encore en incapacité ou avec des incapacités très précoces qui sont à haut risque de progresser. La compréhension des mécanismes de compensation à travers lesquels les individus conservent une autonomie malgré des déficits physiologiques est un point crucial en recherche gériatrique. Au cours de cette phase, les réserves fonctionnelles sont presque entièrement mobilisées. Dans la vie courante l'individu utilise seulement 20 à 30% des capacités d'un système gardant 70 à 80% en réserve. Quand cette marge de sécurité devient très faible, l'apparition d'incapacités à la faveur d'un événement mineur, devient très probable. L'évaluation de ces réserves fonctionnelles nous paraît pouvoir permettre de détecter les personnes "asymptomatiques", apparemment sans problème d'autonomie mais qui seraient dans cette phase de compensation maximale et donc à haut risque de décompensation. Cette évaluation consisterait à faire passer aux personnes des tests du type tests d'effort, dans des conditions légèrement plus exigeantes que les conditions de leur vie quotidienne pour détecter celles ayant des difficultés.

Pour l'analyse statistique de cohortes, cette distinction entre incapacité progressive et incapacité "catastrophique" permet d'envisager plusieurs études chez les personnes vivant à domicile : comorbidité et fragilité comme facteurs de risque de l'incapacité en tenant compte des censures informatives telles que le décès ou l'entrée en incapacité catastrophique.

La caractérisation par des critères physiques de la fragilité pourrait permettre d'isoler une sous-population de personnes âgées fragiles indemnes de pathologies avérées pour mieux étudier les aspects biologiques du vieillissement comparativement à des personnes non fragiles. Dans cette optique, le score de risque d'incapacité chez des personnes sans incapacité au moment des mesures que nous avons proposé peut être une première étape pour repérer ces personnes fragiles.

Comme nous l'avons vu, l'incapacité est un processus dynamique avec de multiples épisodes récurrents. Lorsque les personnes sont en incapacité, la réversibilité et la durée de la nouvelle période d'indépendance physique dépendent des épisodes d'incapacité précédents et donc de leur chronicité [100]. Cette nature discontinue de l'incapacité explique les difficultés rencontrées pour mesurer sa prévalence dans une population. Une évaluation ponctuelle ou bien périodique mais avec des intervalles de temps entre évaluations trop larges aboutit à une sous estimation des incapacités brèves et de la réversibilité.

Pour réduire ces inconvénients, l'on pourrait recommander, à l'image des questions que l'on pose pour les hospitalisations ou les chutes, de faire des évaluations assez rapprochées, par exemple tous les six mois, et de demander à la personne âgée de décrire ses états d'incapacité au cours des 6 derniers mois. Au-delà de cet intervalle de temps, on peut craindre un biais de mémoire trop important surtout dans ces tranches d'âge. Ainsi, un cumul du type (personnes x années) en incapacité pourrait être calculé rendant mieux compte des besoins en assistance de vie et services de santé.

La nature de l'incapacité (progressive ou "catastrophique") joue aussi sur le taux de réversibilité. En effet, chez les personnes âgées fragiles, l'incapacité débute souvent par une période d'instabilité au cours de laquelle les sujets connaissent de brèves périodes d'incapacité faisant suite à des événements mineurs extérieurs. Dans cette période initiale, la réversibilité est élevée [101]. Ensuite, les épisodes d'incapacité deviennent de plus en plus longs. Le déclin peut être très lent jusqu'au décès. Inversement, chez les personnes non fragiles entrant en incapacité brutalement, la durée du premier épisode est plus longue mais si elles sortent de cet état d'incapacité, elles peuvent rester indemnes pendant plusieurs années.

La mise en évidence simultanée des facteurs de risque d'incapacité a été rendue possible grâce à l'outil puissant que constitue le modèle logistique à effets aléatoires. A notre connaissance,



c'est la première fois que ce modèle a été utilisé dans l'analyse des facteurs de risque de l'incapacité chez les personnes âgées. Plusieurs caractéristiques font de ce modèle, un outil très utile pour les études longitudinales. D'abord, le nombre de mesures peut ne pas être le même pour tous les sujets et les sujets avec des données incomplètes au cours du temps peuvent être conservés dans l'analyse. La possibilité d'inclure ces sujets est un avantage important comparativement aux méthodes nécessitant des données complètes au cours du temps car la puissance de l'analyse statistique en est augmentée et l'on évite des biais de sélection, les sujets avec des données complètes pouvant ne pas être comparables aux autres. De plus, comme le temps est traité en variable continue dans ces modèles, il n'est pas nécessaire d'avoir les mêmes dates ni le même nombre d'évaluations par sujet. Ceci est très utile dans les études de cohorte où les temps de mesure ne sont souvent pas uniformes d'un sujet à l'autre. Ces modèles permettent d'inclure à la fois des variables fixes dans le temps et des variables dépendantes du temps. Les évolutions au cours du temps sur la variable réponse peuvent ainsi provenir de caractéristiques stables du sujet (sexe, niveau d'étude) comme de caractéristiques changeant au cours du temps (pathologies). Enfin, alors que les méthodes traditionnelles estiment des évolutions moyennes dans le temps, les modèles à effets aléatoires estiment aussi les changements par sujet. Ceci peut être intéressant dans les études longitudinales où une partie des sujets ont des profils d'évolution qui s'écartent de la tendance moyenne. Les modèles à effets aléatoires modélisent des trajectoires individuelles tandis que les modèles marginaux ne tiennent pas compte du lien unissant les données d'un même individu.

Concernant les études longitudinales, une importante question est l'existence de sorties d'étude au cours du temps. Le modèle marginal GEE n'est valide que dans le cas de sorties d'étude complètement aléatoires. Le modèle à effets aléatoires fait l'hypothèse moins restrictive de

sorties d'étude aléatoires. Mais, lorsqu'il existe des sorties d'étude informatives, les inférences basées sur la vraisemblance du modèle mixte sont biaisées. Dans l'exemple de l'étude de l'incapacité, les décès posent un problème car ils sont souvent précédés d'une période d'incapacité, parfois brève, et une partie d'entre eux sont vraisemblablement informatifs. La solution qui consiste à introduire le décès comme un stade ultime de l'incapacité, n'est pas satisfaisante car ce qui nous intéresse est de prédire la vie en incapacité et donc de rechercher les facteurs spécifiques de l'incapacité. Cette question a été étudiée par Ten Have [102-104] qui a proposé d'utiliser des modèles partagés. Cela consiste dans la forme la plus simple à modéliser deux variables à la fois: la dépendance fonctionnelle par un modèle ordinal et la sortie d'étude (décès et perdus du vue) par un modèle discret de survie (modèle de ratios de continuation). Les deux modèles partagent un vecteur commun d'effets aléatoires, ce qui a conduit au terme de modèle partagé. Bien que délicates, les simulations semblent montrer qu'utiliser un modèle simple à effets mixtes en présence de sorties d'études informatives tend à induire des biais plus grands pour les effets intra-sujets que pour les effets inter-sujets. Ces méthodes sont encore du domaine expérimental, mais présentent un grand intérêt dans notre problématique.

La méthode de modélisation décrite dans ce travail pour la dépendance est facilement transposable à d'autres pathologies chroniques comportant des épisodes avec des périodes de rémission et dont la gravité est mesurée au moyen d'indicateurs binaires ou d'échelles ordinales. Nous pensons par exemple à l'étude longitudinale des migraines, de la polyarthrite. Plus largement, les modèles étudiés peuvent être appliqués à d'autres structures de données groupées. Nous nous sommes par exemple intéressés à l'étude des pathologies oculaires. L'unité statistique observée est alors l'œil et les résultats concernant les yeux d'une même personne ont tendance à être corrélés. Nos analyses ont porté sur l'enquête POLA (Pathologies

Oculaires Liées à l'Age) [105] qui a recruté entre mai 1995 et juillet 1997, 2584 personnes (1451 femmes et 1133 hommes) de 60 ans et plus. Un examen de suivi a été proposé aux participants survivants, trois ans après l'inclusion. Parmi les 2436 survivants, 1947 sujets (80%) ont été alors revus. Nous avons étudié plus particulièrement, l'analyse des facteurs de risque de l'incidence de la cataracte. Les résultats ont été publiés dans l'article de Delcourt et al [106]. Dans ce type de structure de données, lors des analyses en transversal à l'inclusion, la plupart des auteurs ont travaillé au niveau du sujet en considérant l'œil le plus gravement atteint et en utilisant des modèles logistiques simples. Par contre, en longitudinal, l'utilisation de modèles pour données groupées travaillant au niveau de l'œil devient indispensable. En effet, l'œil le plus gravement atteint à l'inclusion n'est pas forcément le plus gravement atteint à l'examen de suivi. De plus, dans l'analyse des facteurs de risque d'aggravation des pathologies oculaires, ces modèles permettent de prendre en compte des facteurs mesurés au niveau de l'œil à l'inclusion tels que le degré de myopie ou d'hypermétropie, ces facteurs étant souvent asymétriques chez un même sujet.

L'utilisation du modèle logistique à effets aléatoires a été limitée en partie à cause de la complexité des méthodes d'estimation et de la lourdeur des calculs. Mais ces modèles commencent à être implémentés dans des logiciels populaires comme SAS avec la procédure NLMIXED ou STATA avec la procédure GLLAMM ce qui va accélérer leur diffusion. Nous espérons que ce travail de thèse contribuera à cette diffusion, tout aidant à mieux cerner les limites et les contraintes qu'imposent ces modèles.

## Bibliographie

1. Diggle P., Liang K. Zeger S. Analysis of longitudinal data, Oxford, Clarendon Press: Oxford Science Publications, 1994.
2. Borrel C. Durr J. Enquêtes annuelles de recensement: premiers résultats de la collecte 2004. Insee Première 2005(1001).
3. Brutel C. Omalek L. Projection démographiques pour la France, ses régions et ses départements (horizon 2030/2050). Insee Résultats Société 2003(16).
4. Bontout O., Colin C. Kerjosse R. Personnes âgées dépendantes et aidants potentiels: une projection à l'horizon 2040. Etudes et Résultats 2002(160): 1-10.
5. Nagi S., *Disability Concepts Revisited: Implications for Prevention*, in *Disability in America: Toward a National Agenda for Prevention*, A.M. Pope and A.R. Tarlov, Editors, Editors. 1991, National Academy Press: Washington, D.C.
6. Wood P.H. Appreciating the consequences of disease: the international classification of impairments, disabilities, and handicaps. WHO Chron 1980; **34**(10): 376-80.
7. Wood P.H. The language of disablement: a glossary relating to disease and its consequences. Int Rehabil Med 1980; **2**(2): 86-92.
8. Wood P.H. Maladies imaginaires: some common misconceptions about the ICIDH. Int Disabil Stud 1987; **9**(3): 125-8.
9. *Toward a common language for functioning, disability and health: The international classification of functioning, disability and health.*, WHO, Editor. 2002: Geneva.
10. Katz S., Downs T.D., Cash H.R., et al. Progress in development of the index of ADL. Gerontologist 1970; **10**(1): 20-30.
11. Lawton M. Brody E. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969; **9**(3): 179-186.
12. Colin C. Coutton V. Le nombre de personnes âgées dépendantes d'après l'enquête Handicaps-incapacités-dépendance. Etudes et Résultats 2000(94): 1-8.
13. ENSP, ed. *RAI-domicile version 2.0. Méthode d'évaluation pour les clients des services à domicile*. 2003: Rennes.
14. Hebert R., Carrier R. Bilodeau A. Le système de mesure de l'autonomie fonctionnelle (SMAF). Rev Gériatr 1988; **13**: 161-167.

15. Desrosiers J., Bravo G., Hebert R., et al. Reliability of the revised functional autonomy measurement system (SMAF) for epidemiological research. *Age Ageing* 1995; **24**(5): 402-6.
16. Guralnik J.M. Ferrucci L. Underestimation of disability occurrence in epidemiological studies of older people: is research on disability still alive? *J Am Geriatr Soc* 2002; **50**(9): 1599-601.
17. Colvez A., Castex A. Carriere I. Recovery from disability in an older population. Results of long-term follow-up in Upper-Normandy (France). *Rev Epidemiol Sante Publique* 2003; **51**(6): 565-73.
18. Gill T.M., Robison J.T. Tinetti M.E. Predictors of recovery in activities of daily living among disabled older persons living in the community. *J Gen Intern Med* 1997; **12**(12): 757-62.
19. Hardy S.E. Gill T.M. Recovery from disability among community-dwelling older persons. *Jama* 2004; **291**(13): 1596-602.
20. Gill T.M., Hardy S.E. Williams C.S. Underestimation of disability in community-living older persons. *J Am Geriatr Soc* 2002; **50**(9): 1492-7.
21. Beckett L., Brock D., Lemke J., et al. Analysis of change in self-reported physical function among older persons in four population studies. *Am J Epidemiol* 1996; **143**(8): 766-778.
22. Mendes de Leon C.F., Glass T.A., Beckett L.A., et al. Social networks and disability transitions across eight intervals of yearly data in the New Haven EPESE. *J Gerontol B Psychol Sci Soc Sci* 1999; **54**(3): S162-72.
23. Leveille S.G., Penninx B.W., Melzer D., et al. Sex differences in the prevalence of mobility disability in old age: the dynamics of incidence, recovery, and mortality. *J Gerontol B Psychol Sci Soc Sci* 2000; **55**(1): S41-50.
24. Gauzère F., Commenges D., Barberger-Gateau P., et al. Maladie et dépendance. Description des évolutions par modèles multi-états. *Population* 1999; **54**(2): 205-222.
25. Albert P. Longitudinal data analysis (repeated measures) in clinical trials. *Stat Med* 1999; **18**(13): 1707-1732.
26. Dargent-molina P., Favier F., Grandjean H., et al. Fall-related factors and risk of hip fracture: the EPIDOS prospective study. *Lancet* 1996; **348**(9021): 145-149.
27. Mc Cullagh P. Nelder J. *Generalized Linear Models - Second Edition*, Boca Raton, London, New-York, Washington, D.C.: Chapman & Hall/CRC, 1989.

28. Agresti A. Categorical data analysis, New York, Chichester, Brisbane, Toronto, Singapore: Wiley -Interscience Publication, 1990.
29. Liang K. Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**(1): 13-22.
30. Zeger S., Liang K. Albert P. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988; **44**: 1049-1060.
31. Pinheiro J. Bates D. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J Comput Graph Stat* 1995; **4**: 12-35.
32. Gilks W., Richardson S. Spiegelhalter D. *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, 1996.
33. Zeger S. Karim M. Generalized linear models with random effects: A Gibbs sampling approach. *J Am Stat Assoc* 1991; **86**: 79-86.
34. Albert I. Jais J. Gibbs sampler for the logistic model in the analysis of longitudinal binary data. *Stat Med* 1998; **17**(24): 2905-2921.
35. Breslow N. Clayton D. Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 1993; **88**(421): 9-25.
36. Wolfinger R. O'connell M. Generalized linear mixed models: a pseudo-likelihood approach. *J Statist Comput Simul* 1993; **48**: 233-243.
37. Breslow N. Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* 1995; **82**(1): 81-91.
38. Rodriguez G. Goldman N. An assessment of estimation procedures for multilevel models with binary responses. *J R Statist Soc A* 1995; **158**: 73-89.
39. Lindsey J. *Models for Repeated Measurements Second Edition*, ed. O.S.S. Series: Oxford University Press, 1999.
40. Littell R., Pendergast J. Natarajan R. Modelling covariance structure in the analysis of repeated measures data. *Stat Med* 2000; **19**(13): 1793-1819.
41. Agresti A., Booth J., Hobert J., et al., *Random effects modelling of categorical response data*, in *Technical report*. 2000, University of Florida.
42. Little R. Rubin D. *Statistical analysis with missing data*, New-York: Wiley and Sons, 1987.
43. Diggle P. Kenward M. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C* 1994; **43**: 49-93.
44. Rotnitzky A. Wypij D. A note on the bias of estimators with missing data. *Biometrics* 1994; **50**(4): 1163-70.

45. Robins J., Rotnitzky A. Zhao L. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J Am Stat Assoc* 1995; **90**(429): 106-121.
46. Collett D. *Modelling Survival Data in Medical Research. Second Edition.*, ed. T.i.S.S. Series, Boca Raton, London, New York, Washington DC: Chapman, & Hall/CRC, 2003.
47. Lindsey J. Lambert P. On the appropriateness of marginal models for repeated measurements in clinical trials. *Stat Med* 1998; **17**(4): 447-469.
48. Neuhaus J., Kalbfleisch J. Hauck W. A comparison of cluster-specific and population-averaged approaches for analysing correlated binary data. *Int Stat Rev* 1991; **59**: 25-36.
49. Seeman T., Charpentier P., Berkman L., et al. Predicting changes in physical performance in a high-functioning elderly cohort: MacArthur studies of successful aging. *J Gerontol* 1994; **49**(3): M97-108.
50. Molenberghs G., Burzykowski T., Michiels B., et al. Analysis of incomplete public health data. *Rev Epidemiol Santé Publique* 1999; **47**(6): 499-514.
51. Gill T., Baker D., Gottschalk M., et al. A program to prevent functional decline in physically frail, elderly persons who live at home. *N Engl J Med* 2002; **347**(14): 1068-1074.
52. Cole T. Scaling and rounding regression coefficients to integers. *Appl Stat* 1993; **42**: 261-268.
53. Harrell Fe J., Lee K., Califf R., et al. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984; **3**(2): 143-152.
54. Oman D., Reed D. Ferrara A. Do elderly women have more physical disability than men do? *Am J Epidemiol* 1999; **150**(8): 834-842.
55. Von Strauss E., Aguero Torres H., Kareholt I., et al. Women are more disabled in basic activities of daily living than men only in very advanced ages: a study on disability, morbidity, and mortality from the Kungsholmen Project. *J Clin Epidemiol* 2003; **56**(7): 669-677.
56. Wolinsky F. Johnson R. The use of health services by older adults. *J Gerontol* 1991; **46**(6): S345-S357.
57. Kovar M. Lawton P. Activities and instrumental activities of daily living. *Ann Rev Geriat Gerontol* 1994; **14**: 57-75.

58. Thomas V., Rockwood K. McDowell I. Multidimensionality in instrumental and basic activities of daily living. *J Clin Epidemiol* 1998; **51**(4): 315-321.
59. Barberger-Gateau P., Commenges D., Gagnon M., et al. Instrumental activities of daily living as a screening tool for cognitive impairment and dementia in elderly community dwellers. *J Am Geriatr Soc* 1992; **40**(11): 1129-34.
60. Fried L., Ferrucci L., Darer J., et al. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. *J Gerontol A Biol Sci Med Sci* 2004; **59**(3): 255-263.
61. Bortz Wm N. A conceptual framework of frailty: a review. *J Gerontol A Biol Sci Med Sci* 2002; **57**(5): M283-M288.
62. Campbell A. Buchner D. Unstable disability and the fluctuations of frailty. *Age Ageing* 1997; **26**(4): 315-318.
63. Fried L., Tangen C., Walston J., et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 2001; **56**(3): M146-M156.
64. Ferrucci L., Guralnik J., Studenski S., et al. Designing randomized, controlled trials aimed at preventing or delaying functional decline and disability in frail, older persons: a consensus report. *J Am Geriatr Soc* 2004; **52**(4): 625-634.
65. Studenski S., Hayes R., Leibowitz R., et al. Clinical Global Impression of Change in Physical Frailty: development of a measure based on clinical judgment. *J Am Geriatr Soc* 2004; **52**(9): 1560-1566.
66. Yu E., Kean Y., Slymen D., et al. Self-perceived health and 5-year mortality risks among the elderly in Shanghai, China. *Am J Epidemiol* 1998; **147**(9): 880-890.
67. Hoeymans N., Feskens E., Kromhout D., et al. The contribution of chronic conditions and disabilities to poor self-rated health in elderly men. *J Gerontol A Biol Sci Med Sci* 1999; **54**(10): M501-M506.
68. Jylha M., Guralnik J., Balfour J., et al. Walking difficulty, walking speed, and age as predictors of self-rated health: the women's health and aging study. *J Gerontol A Biol Sci Med Sci* 2001; **56**(10): M609-M617.
69. Chin A Paw M., Dekker J., Feskens E., et al. How to select a frail elderly population? A comparison of three working definitions. *J Clin Epidemiol* 1999; **52**(11): 1015-1021.
70. Manton K., Stallard E. Corder L. Education-specific estimates of life expectancy and age-specific disability in the U.S. elderly population: 1982 to 1991. *J Aging Health* 1997; **9**(4): 419-450.



71. Launer L., Harris T., Rumpel C., et al. Body mass index, weight change, and risk of mobility disability in middle-aged and older women. The epidemiologic follow-up study of NHANES I. *JAMA* 1994; **271**(14): 1093-1098.
72. Calle E., Thun M., Petrelli J., et al. Body-mass index and mortality in a prospective cohort of U.S. adults. *N Engl J Med* 1999; **341**(15): 1097-1105.
73. Nourhashemi F., Andrieu S., Gillette Guyonnet S., et al. Instrumental activities of daily living as a potential marker of frailty: a study of 7364 community-dwelling elderly women (the EPIDOS study). *J Gerontol A Biol Sci Med Sci* 2001; **56**(7): M448-M453.
74. Ananth C. Kleinbaum D. Regression models for ordinal responses: a review of methods and applications. *Int J Epidemiol* 1997; **26**(6): 1323-1333.
75. Lall R., Campbell M., Walters S., et al. A review of ordinal regression models applied on health-related quality of life assessments. *Stat Methods Med Res* 2002; **11**(1): 49-67.
76. Agresti A. Natarajan R. Modeling clustered ordered categorical data: a survey. *Int Stat Rev* 2001; **69**(3): 345-371.
77. Engel J. Polytomous logistic regression. *Stat Neerlandica* 1988; **42**: 233-252.
78. Walker S. Duncan D. Estimation of the probability of an event as a function of several independent variables. *Biometrika* 1967; **54**: 167-179.
79. Mc Cullagh P. Regression models for ordinal data (with discussion). *J R Statist Soc B* 1980; **42**: 109-142.
80. Peterson B. Harrell F. Partial proportional odds models for ordinal response variables. *Applied Statistics* 1990; **39**: 205-217.
81. Anderson J. Regression and ordered categorical variables. *J R Statist Soc B* 1984; **46**: 1-30.
82. Lunt M. Prediction of ordinal outcomes when the association between predictors and outcome differs between outcome levels. *Stat Med* 2004.
83. Armstrong B. Sloan M. Ordinal regression models for epidemiologic data. *Am J Epidemiol* 1989; **129**(1): 191-204.
84. Cole S. Ananth C. Regression models for unconstrained, partially or fully constrained continuation odds ratios. *Int J Epidemiol* 2001; **30**(6): 1379-1382.
85. Greenland S. Alternative models for ordinal logistic regression. *Stat Med* 1994; **13**(16): 1665-1677.

86. Scott S., Goldberg M. Mayo N. Statistical assessment of ordinal outcomes in comparative studies. *J Clin Epidemiol* 1997; **50**(1): 45-55.
87. Hartzel J., Agresti A. Caffo B. Multinomial logit random effects models. *Statistical Modelling* 2001; **1**(2): 81-102.
88. Hedeker D. Mermelstein R. A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research* 1998; **33**(4): 455.
89. Carlin B. Louis T. Bayes and empirical Bayes methods for data analysis- Second Edition, ed. H.C. Chapman, London, 2000.
90. Vonesh E., Chinchilli V. Pu K. Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics* 1996; **52**(2): 572-587.
91. Hedeker D. Mermelstein R. Analysis of longitudinal substance use outcomes using ordinal random-effects regression models. *Addiction* 2000; **95**: S381-S394.
92. Hedeker D., Mermelstein R. Weeks K. The thresholds of change model: an approach to analysing stages of change data. *Ann Behav Med* 1999; **21**(1): 61-70.
93. Hedeker D. A mixed-effects multinomial logistic regression model. *Stat Med* 2003; **22**(9): 1433-1446.
94. Kempen G., Myers A. Powell L. Hierarchical structure in ADL and IADL: analytical assumptions and applications for clinicians and researchers. *J Clin Epidemiol* 1995; **48**(11): 1299-1305.
95. Zheng B. Summarizing the goodness of fit of generalized linear models for longitudinal data. *Stat Med* 2000; **19**(10): 1265-1275.
96. Zheng B. Agresti A. Summarizing the predictive power of a generalized linear model. *Stat Med* 2000; **19**(13): 1771-1781.
97. Gill T.M., Allore H., Holford T.R., et al. The development of insidious disability in activities of daily living among community-living older persons. *Am J Med* 2004; **117**(7): 484-91.
98. Ferrucci L., Guralnik J.M., Simonsick E., et al. Progressive versus catastrophic disability: a longitudinal view of the disablement process. *J Gerontol A Biol Sci Med Sci* 1996; **51**(3): M123-30.
99. Ferrucci L., Guralnik J.M., Cavazzini C., et al. The frailty syndrome: a critical issue in geriatric oncology. *Crit Rev Oncol Hematol* 2003; **46**(2): 127-37.
100. Hardy S.E. Gill T.M. Factors associated with recovery of independence among newly disabled older persons. *Arch Intern Med* 2005; **165**(1): 106-12.

101. Hardy S.E., Dubin J.A., Holford T.R., et al. Transitions between states of disability and independence among older persons. *Am J Epidemiol* 2005; **161**(6): 575-84.
102. Ten Have T., Kunselman A., Pulkstenis E., et al. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* 1998; **54**: 367-383.
103. Ten Have T., Miller M., Reboussin B., et al. Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the longitudinal study of aging. *Biometrics* 2000; **56**(1): 279-287.
104. Ten Have T., Reboussin B., Miller M., et al. Mixed effects logistic regression models for multiple longitudinal binary functional limitation responses with informative drop-out and confounding by baseline outcomes. *Biometrics* 2002; **58**: 137-144.
105. Delcourt C., Cristol J.P., Tessier F., et al. Risk factors for cortical, nuclear, and posterior subcapsular cataracts: the POLA study. *Pathologies Oculaires Liees a l'Age. Am J Epidemiol* 2000; **151**(5): 497-504.
106. Delcourt C., Carriere I., Delage M., et al. Associations of cataract with antioxidant enzymes and other risk factors: the French Age-Related Eye Diseases (POLA) Prospective Study. *Ophthalmology* 2003; **110**(12): 2318-26.

# Annexe 1 : Articles originaux

Choosing marginal or random effects models for longitudinal binary responses: application to self-reported disability among older persons

**BMC Medical Research  
Methodology**



Research article

Open Access

## Choosing marginal or random-effects models for longitudinal binary responses: application to self-reported disability among older persons

Isabelle Carrière\*<sup>1</sup> and Jean Bouyer<sup>2</sup>

Address: <sup>1</sup>INSERM Unité 500, 39 avenue Charles Flahault, 34093 MONTPELLIER, FRANCE and <sup>2</sup>INSERM Unité 569-IFR69, 82 rue du Général Leclerc, 94276 LE KREMLIN BICETRE, FRANCE

Email: Isabelle Carrière\* - [carriere@montp.inserm.fr](mailto:carriere@montp.inserm.fr); Jean Bouyer - [bouyer@vjf.inserm.fr](mailto:bouyer@vjf.inserm.fr)

\* Corresponding author

Published: 5 December 2002

Received: 29 July 2002

*BMC Medical Research Methodology* 2002, **2**:15

Accepted: 5 December 2002

This article is available from: <http://www.biomedcentral.com/1471-2288/2/15>

© 2002 Carrière and Bouyer; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Longitudinal studies with binary repeated outcomes are now widespread in epidemiology. The statistical analysis of these studies presents difficulties and standard methods are inadequate.

**Methods:** We consider strategies for modelling binary repeated responses and focus on two specific issues: the choice between marginal and random-effects models, and the choice of the time point origin. These issues are addressed using the example of self-reported disability in older women assessed annually for 6 years. The indicator of disability "needing help to go outdoors or home-confined" is used.

**Results:** In view of the observed associations between the responses for consecutive years, the baseline response was considered as a covariate. We compared the marginal and random-effects models first when only the influence of time and age is analysed and second when individual risk factors are studied in an aetiological perspective. There were substantial differences between the parameter estimates. They were due to differences between specific concepts related to the two models and the large between-individual heterogeneity revealed by the analysis.

**Conclusions:** A random-effects model appears to be most suitable for the analysis of self-reported disability in older women.

### Background

In developed countries, disability in older persons is a major concern for public health authorities. Progress in medical care together with improved living conditions have led to longer life spans. One adverse consequence is that the number of very old people who are disabled is also increasing. The study of the succession of the different stages of disability, and of the ability to recover, together with the analysis of the risk factors for disability, are important issues for research in gerontology.

Longitudinal studies are more appropriate than cross-sectional studies which suffer various limitations due to biases such as the selective removal of disabled persons by institutionalisation, or differences in the proportions of disabled persons between age groups, also confounded with changes across generations [1]. Several authors [2] have analysed repeated measures of disability by comparing two by two time points. This method involves a very large number of tests and leads to partial results for fixed periods. The statistical analysis must take four main char-

acteristics of the longitudinal data into account: 1) time may be an explanatory variable, 2) repeated observations for a subject are likely to be correlated, 3) the covariables may be time-dependent (they may vary through time for a subject), 4) missing data in the successive responses may induce a bias. Several statistical models developed for longitudinal data have become more popular and there is software available for some of them. The problem of the choice of which to use remains.

Here, we present a strategy to model binary repeated responses and to explain how to choose between marginal and random-effects models. The evolution of disability in older persons is used as an example to illustrate each step.

## Methods

### Example

The French EPIDOS study is a prospective multicentre study of the risk factors for hip fractures in 7575 women aged 75 years or older, included in 1992–1993 and recruited by mailing based on large population-based listings including electoral rolls [3]. Once in the study, these women were contacted by mail or telephone every four months to collect information about any falls and new fracture events. They had to complete a mail questionnaire annually investigating hospitalisation, new health events, changes in weight, type of housing, ability to go outside, activities of daily life (ADL), instrumental activities of daily life (IADL), medications used, and subjective health. This follow-up was initially planned for 4 years and then was extended to 6 years. In this paper we analyse the data from Montpellier (southern France), one of the 5 participating centres. We used the data from all the subjects included at this centre (1548 women) to analyse the evolution of disability.

The annual questionnaire was in some cases missing. This was mainly due to illness or to family events such as the death of the spouse. The women then postponed their answer.

Various indicators have been proposed to assess disability [4]. Needing help for going outdoors represents a first evident level of functional limitation. It is easy to identify and concerns both men and women. Therefore we chose "needing help to go outdoors or home-confined" as an indicator of disability. The annual response variable is binary.

### Marginal and random-effects models

In longitudinal studies, multiple assessments of the same subject at different time points are used and the within subject responses are then correlated. This correlation must be accounted for by analysis methods appropriate to the data [5]. Several models have been proposed for the

analysis of such data. Most of them are extensions of the well-known logistic regression that is a particular case of the generalized linear models with a logistic link function [6]. They are usually classified into marginal or random-effects models. Random-effects models are also called generalized linear mixed models or multilevel models or conditional models. However, this last term is ambiguous as several authors use it only for conditional maximum likelihood estimation [5] that we do not consider in this paper. Unlike linear models, the interpretation of the coefficients of these two types of model differs (see below). The choice of one or the other depends on the objectives of the study.

Let  $Y_{ij}$  denote a binary outcome (in our example: needing help to go outdoors or home confined) corresponding to the  $j$ th response ( $j$ th year in the study in our example,  $j = 1$  to  $n_i$ ) of the  $i$ th subject ( $i = 1$  to  $K$ ). Let also  $X_{ij}$  be a design matrix of covariates ( $1 \times p$  vector, with first element being 1 for the intercept). The covariates may be fixed (for example age at baseline), or take different values at each year of the study (for example time or hospitalised for the last year). The marginal model, also called the population-averaged model [7], estimates the model thus:

$$\text{logit}(E(Y_{ij} | X_{ij})) = \text{logit}(P(Y_{ij} = 1 | X_{ij})) = X_{ij}'\beta$$

whereas the random-effects model, also called the "subject-specified models" [7], estimates the model as follows:

$$\text{logit}(P(Y_{ij} = 1 | X_{ij})) = X_{ij}'\beta_i$$

Thus, the marginal model supposes that the relationship between the outcome  $Y$  and the covariate  $X$  is the same for all the subjects, and the random-effects model allows this relationship to differ between subjects. To highlight this point, the random-effects model may also be written:

$$\text{logit}(E(Y_{ij} | X_{ij}, U_i)) = X_{ij}'(\beta^* + U_i) = X_{ij}'\beta^* + X_{ij}'U_i$$

or:

$$\text{logit}(E(Y_{ij} | X_{ij}, Z_{ij}, U_i)) = X_{ij}'\beta^* + Z_{ij}'U_i$$

where the random effects  $U_i$  are assumed to vary independently from one subject to another according to a common distribution. This distribution is often supposed to be normal with mean 0 and variance  $D$ .  $Z_{ij}$  is often a subvector of  $X_{ij}$ , which means that random effects apply only to a part of the covariates and the intercept. The variance,  $D$ , has to be estimated and represents the extent of the unexplained between-individual variability.

It is important to note that the  $p \times 1$  vectors  $\beta$  in the marginal model and  $\beta^*$  in the random effects model are not

equal. Hence, the estimators estimate different things [8] and the magnitude of the difference between  $\beta$  and  $\beta^*$  is function of the variance,  $D$  [9].

Moreover, dependencies between observations over the time are handled differently in the two models. In the marginal model, it is popular to fit the vector of parameters,  $\beta$ , using the Generalized Estimating Equations (GEE) proposed by Liang and Zeger [10] wherein the covariance matrix is structured by using a working correlation matrix  $R(\alpha)$  fully specified by the vector of parameters,  $\alpha$ . This working correlation is assumed to be the same for all the subjects, reflecting an average dependence among the repeated observations for all subjects. In contrast, the random-effects model allows this within-subjects dependency to vary from one subject to another, by the means of the random part of the covariable linear combination. In the simplest case, only a random intercept is introduced

$$\text{logit}(E(Y_{ij} | X_{ij}, U_i)) = X_{ij}'\beta^* + U_i0$$

$U_i0$  being an individual parameter of propensity to become disabled, constant through the time. For a given subject, whatever the interval of time between two responses, the strength of dependency is then the same. If a random slope for the time covariable is added

$$\text{logit}(E(Y_{ij} | X_{ij}, U_i)) = X_{ij}'\beta^* + U_i0 + U_i1 \text{ time}$$

this individual strength of association increases or decreases with the width of the interval.

In the marginal model, several specific choices of the structure of the working correlation matrix  $R(\alpha)$  are possible. For example,  $R(\alpha)$  is called  $m$ -dependent if  $\text{corr}(Y_{ij}, Y_{ik}) = \alpha^t$  for  $t = 1, 2, \dots, m$  and  $\text{corr}(Y_{ij}, Y_{ik}) = 0$  for  $t > m$ ;  $R(\alpha)$  is exchangeable if  $\text{corr}(Y_{ij}, Y_{ik}) = \alpha$  for  $j \neq k$ , and it is unstructured if  $\text{corr}(Y_{ij}, Y_{ik}) = \alpha_{ik}$ . An advantage of the marginal model, demonstrated by Liang and Zeger, is that  $\beta$  and their robust variance are consistent (the estimator converges towards the parameter being estimated as the sample size increases) even when the correlation structure is misspecified. However, choosing the working correlation structure closest to the true structure increases the statistical efficiency of the parameter estimator. Consequently, it is recommended to specify the working correlation as accurately as possible, based on the knowledge of the longitudinal process [11].

Concerning the estimating procedures, the GEE method for marginal models is not difficult to implement and is now available in the major statistical analysis packages. The procedures are more complex for the random-effects model. The most attractive of them directly maximizes an

approximate integrated likelihood. With non-linear models computational-intensive integration methods, such as Gauss-Hermite quadrature, are necessary to evaluate the likelihood [12].

Different assumptions are required for the two models regarding missing data. The marginal model using the GEE requires a missing data process completely at random (MCAR). Under this assumption, missingness does not depend on individual characteristics (observed or not). In contrast, random models only need the less stringent assumption of missing at random (MAR). In this process, the probability of missingness depends only on observed variables (previous covariates or outcomes) [13]. All marginal models do not require the MCAR assumption. For example, Robins et al [14] proposed methods to allow for data that are MAR in marginal models, but these methods are more complicated to implement.

The interpretation of the coefficients  $\beta$  and  $\beta^*$  also differs. Consider, for instance, the covariate  $X$  "living alone". The odds ratio  $OR^* = \exp(\beta^*)$  represents the odds of the outcome (needing help to go outdoors or home confined) for a person living alone compared to the same person supposed not to live alone. It can be seen as an odds ratio adjusted on unobserved individual characteristics. Under the marginal model,  $OR = \exp(\beta)$  represents the averaged odds of the sub-group living alone compared to the sub-group not living alone.

In the following sections we will first describe the data and then focus on the specific problems inherent to longitudinal analyses: how to choose the time point origin and how to take into account the influence of time on responses. Then, we will present an example of an estimation of risk factors including covariates fixed across time for a single subject, and time-dependent covariates.

## Results

### Description of the data

The mean (SD) age at inclusion was 80 years (3.7) and only 10 percent of the included women were aged 90 years or over. The proportion of women reporting disability (Table 1) increased only moderately with time except for the first year during which it jumped from 29 percent to 39 percent. The sequence of the seven successive states of disability (path) differs between the women. Only 268 women (17 percent) were able to go outside without help from the beginning to the end of the study, 77 (5 percent) were disabled throughout the 6 years. The frequencies of each of the other paths are very small: 1 or 2 percent. During the 6 years, recovery from disability was observed at least once in 498 women (32 percent). This relatively high proportion of women recovering from a disabled state is

Table 1: Description of the data by year of assessment

	Inclusion	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Total number of women with disability assessment	1548	1503	1450	1409	1294	1020	970
Percent of disabled women	29%	39%	41%	46%	47%	48%	52%
Time since inclusion (yrs)							
Median		0.99	1.98	2.97	3.95	4.96	6.12
Range		0.91, 1.67	1.95, 2.44	2.93, 3.37	3.87, 4.53	4.87, 6.35	5.88, 6.94
Number of deaths for the preceding year		23	34	40	63	58	76

why we chose to perform a repeated measures analysis and not to use statistical methods analysing time before entering disability, such as Cox regression.

The study was initially planned to last 4 years, some women refused to continue in years 5 and 6, and the number of those returning the completed questionnaire decreased most quickly between years 4 and 5 (Table 1). Some women gave intermittent answers. For example, of the 970 women still providing a response in year 6, at least one observation in the previous years was missing for 150 (15 percent). The missingness partly depended on unobserved individual events (illness or family events) and is unlikely to be MCAR. The use of the marginal model is therefore questionable because an averaged effect is only poorly meaningful. The true time between yearly assessments and inclusion varied from one woman to another and the range tended to be higher at the end of the follow-up. Consequently, time is considered as a continuous covariable. Vital status was known throughout the 6 years for all the participants, even those who no longer sent back the questionnaires. The total number of deaths was 294 (19 percent) by the end of the study.

**Choice of the time point origin**

The response  $Y_{i0}$  given by the  $i$ th subject at baseline may be handled in two ways. It can be considered either as a part of the longitudinal response, and thus the response vector has 7 elements corresponding to the times 0, 1, ..., 6 years, or as a baseline covariate, in which case the response vector has 6 elements. These two possibilities correspond to different structures for the raw data and are displayed in Tables 2 and 3. In the models studied (marginal or random-effects model), the association between two successive responses is supposed to have the same structure through the whole survey. If the association between baseline disability and the other responses differs from the association between the follow up responses, then it is better to consider the baseline response as a covariable.

In our example, as in many cohort studies, the baseline examination was exhaustive. The women were proposed

Table 2: Structure of the raw data when the baseline response is taken as a part of the response

Response	Time	Covariates $l = 1, \dots, p$	Time dependent covariates $l = 1, \dots, q$
$y_{i0}$	$t_{i0}$	$x_{i0l}$	$z_{i0l}$
$y_{i1}$	$t_{i1}$	$x_{i1l}$	$z_{i1l}$
$y_{i2}$	$t_{i2}$	$x_{i2l}$	$z_{i2l}$
$y_{i3}$	$t_{i3}$	$x_{i3l}$	$z_{i3l}$
$y_{i4}$	$t_{i4}$	$x_{i4l}$	$z_{i4l}$
$y_{i5}$	$t_{i5}$	$x_{i5l}$	$z_{i5l}$
$y_{i6}$	$t_{i6}$	$x_{i6l}$	$z_{i6l}$

clinical and functional examinations (series of standard tests of physical performance) and a bone densitometry at the hospital. However, only those who felt well enough to undergo these tests were examined, probably introducing a selection bias at inclusion. During the follow-up, the women were only asked to complete a mailed questionnaire that even physically dependent persons could answer. The profile of our sample changed through time, with the women becoming more and more physically dependent. At the beginning, many of the disabled women were probably only slightly disabled and thus there was the possibility of recovery. As time passed more women were severely disabled, and stayed so until the end of the study. Thus, the association between the responses for consecutive years is likely to be stronger in the last years. Also, the baseline response may be considered as a special case, less related to the following disabled states.

To check this, we examined the table of the odds ratios of being disabled for the year  $j$  ( $Y_{ij}$ ) according to disability the previous years ( $Y_{i(j-t)}$ ). The association between two

Table 3: Structure of the raw data when the baseline response is taken as a baseline covariate

Response	Time	Response at t 0	Covariates l = 1, ... p	Baseline values of the time dependent covariates l = 1, ...q	Time dependent covariates l = 1, ...q
y11	t1	y10	...x01...	..zi01..	...zi11..
y12	t2	y10	...x01...	..zi01..	...zi21..
y13	t3	y10	...x01...	..zi01..	...zi31..
y14	t4	y10	...x01...	..zi01..	...zi41..
y15	t5	y10	...x01...	..zi01..	...zi51..
y16	t6	y10	...x01...	..zi01..	...zi61..

Table 4: Odds ratios of being disabled at each year according to disability at the previous years

Inclusion	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
inclusion	5.3	5.1	5.7	4.1	4.3	4.8
Year 1		16.3	12.2	11.1	10.0	9.5
Year 2			18.2	14.3	13.1	11.4
Year 3				16.7	17.3	13.1
Year 4					20.2	15.8
Year 5						20.7

successive time points beyond baseline was stronger than that with the baseline response (Table 4). For instance, in the lower diagonal giving the association between two observations separated by one year, the odds ratio between year 1 and the baseline is 5.3 whereas it is greater than 16 between year 2 and 1, and between the following pairs of years. In view of these results, the baseline response was considered as a covariate and the data were structured as shown in Table 3. There was also a tendency for correlations to decrease with increasing time differences (Table 4). For example, the odds ratio is 16.3 between year 1 and year 2 and 10.0 between year 1 and year 5. This observation allows a better selection of the correlation structure, avoiding the use of an exchangeable working correlation matrix in marginal models, and introducing a random slope in random effect models.

**Analysis of time and age evolution**

The objective here is to evaluate the impact of time on the proportion of disabled women. We chose to use the age at entry (fixed covariable) and the time since baseline to characterize this phenomenon. Another solution would be to use the age at every response (as a time varying cov-

ariable) but our option allows the introduction of an interaction between the age at entry and the time since baseline to test whether the effect of time is more pronounced in the oldest women.

In view of the missingness process resulting in the sample differing at every assessment, and the possible selection bias at entry, the search for individual relationships using the random-effects model is more suitable than using population averaged associations. However the two models are interesting to compare to show that the estimate parameters can be different, and to explain these differences.

We used two methods to characterize the changes in disability over time: the marginal model with the GEE approach and the random model with likelihood integration. The random model had Gaussian random effects and errors. We used the SAS procedures GENMOD with the REPEATED statement and an unstructured working correlation matrix and also NLMIXED with the Gauss-Hermite quadrature integration method [12]. For the SAS code refer to Appendix (see Additional file 1). In the ran-

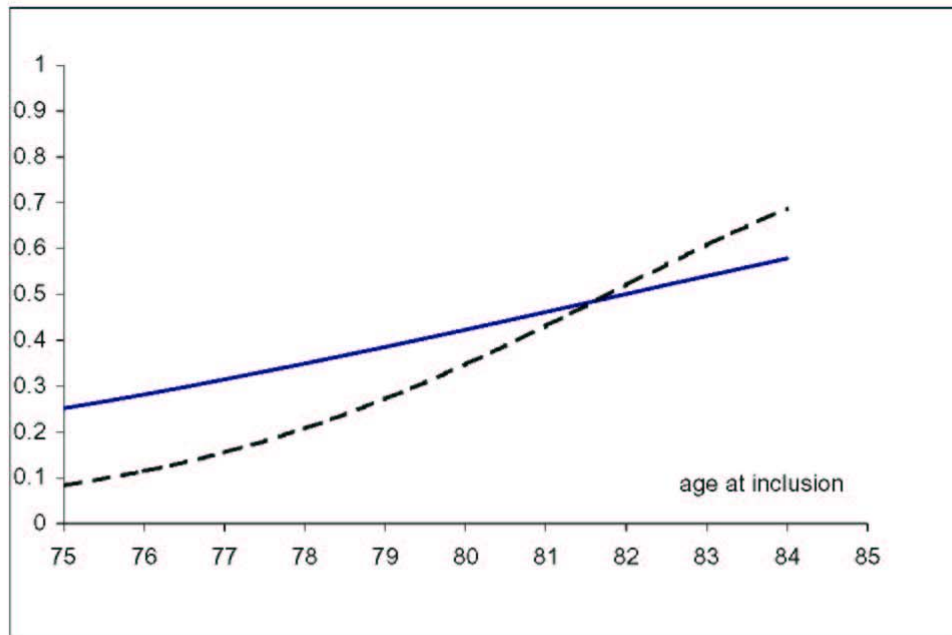


dom model, we determined the structure of the covariance introducing successively two random effects: random intercept and random slope for the time covariate.

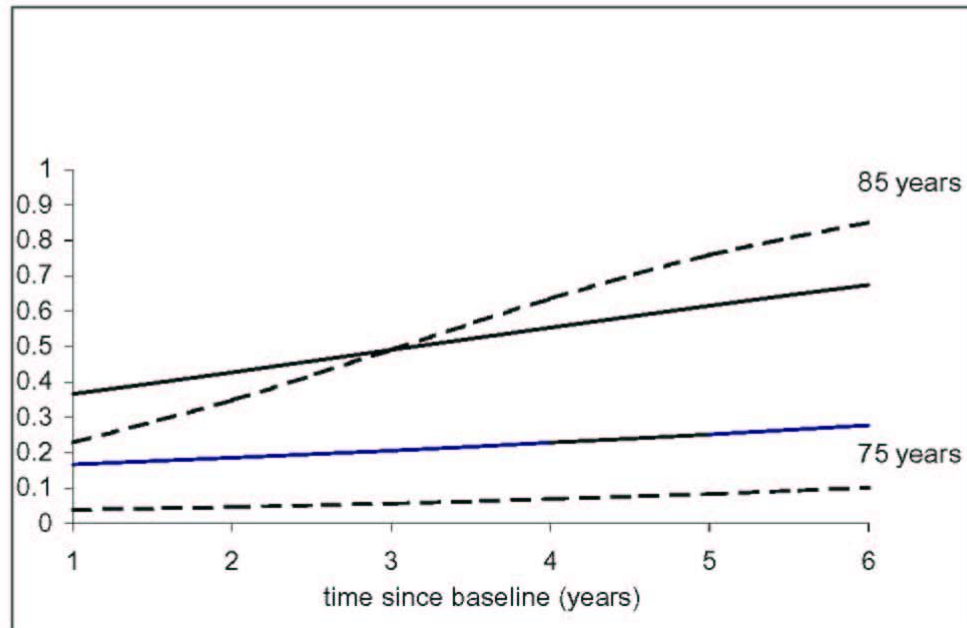
In all the models, the following fixed effects were then tested: time since baseline (continuous variable), time square, age at entry (in years exceeding 74), interactions age  $\times$  time and baseline inability to go outside without help. Only the time square effect was not significant, and was removed from the model. We also introduced an indicator variable for women dying within the 6-year period. This is a simple way to take into account dropouts due to death, a difficulty often encountered in longitudinal studies in the elderly. The use of this time-fixed indicator was possible since information about death was known for the six years for all the women. The NLMIXED procedure converged and provided estimations only when the initial parameters were close to the final solution. We calculated successive models, beginning with a simple model with only a random intercept, and at each step we used the

previous estimate parameters as initialisation for the next estimation.

The results of the comparison between the 2 modelling strategies are shown in Table 5. The general conclusion is the same for the 2 models but the estimated parameters are different. For instance, the most important factor, baseline inability, is very significant in both, but parameter estimates differ from 1.54 to 3.18. The significant increase in risk of being disabled with age and time since baseline is illustrated in figure 1, presenting the averaged probability of disability, 5 years after the inclusion, in women able to go outside without help at baseline and who did not die during the study. In our cohort, recruited in 1992–1993, the probability of restricted mobility, 5 years later, was high, especially after the age of 80 years. Figure 2 shows the changes with time in two groups aged 75 and 85 years at entry. There was a significant interaction between age at entry and time, and consequently the risk of disability is accelerated in the oldest women.



**Figure 1**  
Age at baseline and probability of disability at 5 years in women without disability at entry and surviving to the end of the study, GEE marginal model —, random model - - - - -



**Figure 2**  
Changes through time evolution of the probability of disability in women without disability at entry and surviving to the end of the study, aged 75 and 85 years at entry, GEE marginal model —, random model ----

**Table 5: Marginal model and random-effects models analysing the influence of time and age**

	Marginal model GEE			Random model likelihood integration		
	Estimate	SE	p value	Estimate	SE	p value
Intercept	-1.83	0.12	< 0.0001	-3.61	0.25	< 0.0001
Time (years)	0.12	0.03	< 0.0001	0.17	0.06	0.002
Age at baseline*	0.09	0.02	< 0.0001	0.16	0.03	< 0.0001
Age* × time	0.01	0.004	0.003	0.04	0.009	< 0.0001
Baseline disability	1.54	0.10	< 0.0001	3.18	0.22	< 0.0001
Death during the study	0.68	0.12	< 0.0001	1.36	0.26	< 0.0001
Random intercept variance	-	-	-	7.25	0.68	< 0.0001
Random slope variance	-	-	-	0.18	0.04	< 0.0001

Differences between the estimators of the marginal and random-effects parameters are expected (see above). The marginal model expresses averaged relationships without taking into account the fact that the same subjects are considered at each time point, whereas the random-effects model gives relationships conditionally on having certain

individual characteristics modelled by the random effects. In the case of only a random intercept, Nehaus et al [9] demonstrated that the estimates from the marginal model are systematically lower than those from the random model. This characteristic is shown in figures 1 and 2, where the curves from the marginal model are flatter than

the others. The differences between the estimates of the two approaches are largely dependent on the inter-individual heterogeneity. This heterogeneity can be assessed in the random models by looking at the intercept and slope variances.

In the random model, the random intercept variance is high. The estimate is 7.25, indicating that the variability of this individual additional intercept has a 95 percent width of 10.6 ( $\sqrt{7.25} \times 1.96 \times 2$ ). If we consider the variability given by the fixed covariates (age varying from 75 to 85 years, baseline disability and death each varying from 0 to 1) the width is 6.14 ( $0.16 \times 10 + 3.18 + 1.36$ ). Hence the variability explained by these three fixed variables is lower than the unexplained between-individual variability. Similarly, the random slope variability has a 95 percent width of 1.66, whereas the variability explained by the age has a 95 percent width of 0.40 ( $0.04 \times 10$ ).

The probability of being disabled is therefore much more due to the woman's uncharacterised "frailty" than to age or to baseline disability. This wide between-individual heterogeneity explains the differences between the parameter estimates of the marginal and random-effects models.

#### Analysis of risk factors

Changes with time expressed as age at entry and the time since baseline were modelled in the previous section. The

marginal and random-effects models can also be used to identify other risk factors for disability. It would also be interesting to test how the estimating algorithms behave when numerous covariables are jointly introduced.

We compared two multivariate models adjusted on the covariates presented in the previous section (age, time, interactions between age and time, baseline disability, still alive at the 6-year period) (see table 6). When numerous covariables are put in the model, the NLMIXED procedure may fail to integrate the likelihood or give non-stationary estimations. This non-optimal estimation can be diagnosed by checking the gradient (vector of first derivative) of the negative log likelihood function for each parameter. These gradients are systematically provided by SAS in the results. If one of them is not close to zero, then the solution cannot be considered to be valid. These problems of convergence were not encountered with the set of covariables shown in table 6. Two categories of covariables were tested: variables collected at baseline (living at home alone, body mass index (BMI), visual acuity measured with Snellen letter test chart on a decimal scale, and perceived health) and variables that vary with time (hospitalised at least once during last year, temporarily bed-confined during last year and number of falls during last year). Adjusted on all the others, all these factors were significant.

Table 6: Marginal model and random-effects models analysing time-fixed and time varying covariates

	Marginal model* GEE			Random model* likelihood integration		
	Estimate	SE	p value	Estimate	SE	p value
<b>Baseline</b>						
Living alone	-0.31	0.09	0.0004	-0.66	0.18	0.0002
BMI (kg/m <sup>2</sup> )						
< 25 (ref)	0			0		
25 - 29	0.30	0.10	0.002	0.69	0.20	0.0005
≥ 29	0.75	0.12	< 0.0001	1.46	0.23	< 0.0001
Visual acuity						
≤ 2/10	0.64	0.20	0.001	1.21	0.35	0.0006
> 2/10 (ref)	0			0		
Perceived health						
Bad or very bad	0.76	0.15	< 0.0001	1.43	0.28	< 0.0001
<b>Follow-up</b>						
Hospitalised (past year)	0.26	0.05	< 0.0001	0.49	0.11	< 0.0001
Temporarily bed confined (past year)	0.39	0.07	< 0.0001	0.73	0.15	< 0.0001
Number of falls ≥ 2 (past year)	0.23	0.08	0.004	0.49	0.15	0.0013

\*adjusted on time since inclusion, age at inclusion, age × time, baseline disability and death

## Discussion

We present the steps for choosing a model able to characterize disability taken as a binary response.

Several important points have to be considered in the analysis of repeated binary data. First, the choice between the marginal and random-effects models depends mainly on the aims of the study. If the goal is to predict a mean prevalence of disability over time in elderly people by sex or age group, the marginal model is suitable. In contrast, if the goal is to study the individual risk factors for aetiological considerations, the random-effects model is more suitable because it allows adjustment on non-observed individual characteristics, and a better understanding of the underlying mechanism.

Second, the missing data process has to be examined. The marginal model using GEE assumes that the sample is representative of the whole population at each time point and the missing data is MCAR. This is unlikely to be the case in our example. The random-effects model assuming an MAR process is more appropriate. By the end of the study, half of the 37% of missing data were due to deaths and most were predictable from the previously collected data. The same probably applies to missing data due to chronic illnesses. Nevertheless, we cannot exclude the possibility that unobserved level of disability in our example may influence a part of the missing process. The magnitude of the potential bias introduced by non-random missingness should therefore be examined. A sensitivity analysis to assess the impact of missing data on subsequent statistical inference would be worthwhile [15], but is beyond the scope of this paper.

Another important point is to determine whether the disability indicator at baseline should be considered as a response or as a covariate. In epidemiological cohorts, the conditions in which responses are collected at baseline and during the follow-up are often different. The first response is often considered as a covariate but this choice has to be confirmed in view of the analysis of dependency between the responses.

The structure of the covariance between the repeated responses has also to be chosen. In the marginal model, the inferences on the parameter estimates are asymptotically valid under any assumed structure but it is better to choose a structure corresponding to the data. In contrast, in the random model, the fixed and random parameters are simultaneously estimated and the choice of the covariance structure influences the final results. For the mixed model applied to gaussian responses, it is recommended [16] (i) to consider first the more general model with all the relevant covariables, (ii) to specify a model for the covariance structure (i.e. to specify the random effects), and

to estimate the parameters, and (iii) to try to reduce the fixed effect portion. In practice, the estimating procedures that we used do not allow introduction of a complex covariance structure due to computational time and unstable estimations. Other more flexible methods allowing multiple levels of clustering, such as Markov chain Monte Carlo methods, can be used but are more complex to handle [17].

The calculation procedures for random models need to be improved. The method of estimation with likelihood integration requires excessive computation. An alternative strategy to fit mixed models is the penalized quasi-likelihood (PQL) approach [18,19] (GLIMMIX macro in SAS) but this method provides highly biased estimates of mixed-effects parameters with binary responses [20,21].

The analysis of disability evolution with age at entry and time, as well as the study of other risk factors, show that the differences between the estimates for the two models can be large. However, the interpretation of the estimate parameters is different. In the marginal model, the exponential of an estimate parameter is a population-averaged odds ratio for disability and concerns the sub-population that shares a characteristic relative to the sub-population not sharing this characteristic. In the random model, the exponential of an estimate is an odds ratio for a woman that has a characteristic relative to this same woman if she were free of this characteristic. The random model takes into account the underlying dependence relationship. Furthermore, the assumptions about the distributions are different, and the fact that the conditional distribution is binomial does not imply that the marginal distribution is also binomial [22].

In our example, the search for risk factors for disability prevention, and the characteristics of our sample, such as the missingness process and the number of drop-outs after 4 years, render the random-effects models more appropriate. The marginal model has a tendency to waste information and does not measure the association of within-subject covariate change with change in the response, the associations typically of particular scientific value in longitudinal studies.

## Conclusions

In epidemiology, many rich databases are now available from longitudinal studies with binary repeated events. Traditional analyses comparing time points two by two have serious limitations. Marginal models are easy to implement and represent a first solution, but the random models, although more complex, use all available data and are more suitable for explicative studies.

**Competing interests**

None declared

**Authors' contributions**

IC wrote an initial draft and JB made important improvements. Both authors read and approved the final manuscript.

**Additional material****Additional File 1**

SAS code

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2288-2-15-S1.doc>]

**Acknowledgements**

The authors are grateful to Dr François Favier and the EPIDOS group who provided the data for this work and to Annie Lacroux for editorial assistance.

**References**

- Beckett LA, Brock DB and Lemke JH Analysis of change in self-reported physical function among older persons in four population studies. *Am J Epidemiol* 1996, **143**:766-78
- Hébert R, Brayne C and Spiegelhalter D Factors associated with functional decline and improvement in a very elderly community-dwelling population. *Am J Epidemiol* 1999, **150**:501-10
- Dargent-Molina P, Favier F and Grandjean H Fall-related factors and risk of hip fracture: the EPIDOS prospective study. *Lancet* 1996, **348**:145-49
- WHO/EUROPE, NCBS Third consultation to develop common methods and instruments for Health interviews surveys. Voorburg, the Netherlands: WHO and Netherlands Central Bureau of Statistics 1992.
- Diggle PJ, Liang KY and Zeger SL Analysis of longitudinal data. Oxford: Clarendon Press, Oxford Science Publications 1994.
- McCullagh P and Nelder JA Generalized Linear Models – Second Edition. Boca Raton, London, New-York, Washington, D.C., Chapman & Hall/CRC 1989.
- Zeger SL, Liang KY and Albert PS Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988, **44**:1049-60
- Pendergast JF, Gange SJ and Newton MA A survey of methods for analyzing clustered binary response data. *Int Stat Rev* 1996, **64**:89-118
- Neuhauss JM, Kalbfleisch JD and Hauck WW A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int Stat Rev* 1991, **59**:25-36
- Liang KY and Zeger SL Longitudinal data analysis using generalized linear models. *Biometrika* 1986, **73**:13-22
- Albert PS Longitudinal data analysis (repeated measures) in clinical trials. *Stat Med* 1999, **18**:1707-32
- Pinheiro JC and Bates DM Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* 1995, **4**:12-35
- Little RJA and Rubin DB Statistical analysis with missing data. New-York: Wiley and Sons 1987.
- Robins JM, Rotnitzky A and Zhao LP Analysis of semiparametric regression models for repeated outcomes in presence of missing data. *J Am Stat Assoc* 1995, **90**:106-121
- Molenberghs G, Burzykowski T and Michiels B Analysis of incomplete public health data. *Rev Epidemiol Santé Publique* 1999, **47**:499-14
- Littell RC, Pendergast J and Natarajan R Modelling covariance structure in the analysis of repeated measures data. *Stat Med* 2000, **19**:1793-1819
- Zeger SL and Karim MR Generalized linear models with random effects: A Gibbs sampling approach. *J Am Stat Assoc* 1991, **86**:79-86
- Breslow NE and Clayton DG Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 1993, **88**:9-25
- Wolffinger R and O'Connell M Generalized linear mixed models: a pseudo-likelihood approach. *J Statist Comput Simul* 1993, **48**:233-43
- Breslow NE and Lin X Bias correction in generalised linear models with a single component of dispersion. *Biometrika* 1995, **82**:81-91
- Rodriguez G and Goldman N An assessment of estimation procedures for multilevel models with binary responses. *Journal of Royal Statistical Society* 1995, **A158**:73-89
- Lindsay JK and Lambert P On the appropriateness of marginal models for repeated measurements in clinical trials. *Stat Med* 1998, **17**:447-69

**Pre-publication history**

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/2/15/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)



Page 10 of 10

(page number not for citation purposes)

# Hierarchical components of physical frailty predict incidence of dependency in a cohort of elderly women<sup>4</sup>.

Isabelle Carrière<sup>1</sup>, Alain Colvez<sup>1</sup>, Francois Favier<sup>2</sup>, Claude Jeandel<sup>3</sup>, Hubert Blain<sup>3</sup>  
for the EPIDOS study group\*

## Corresponding author/reprint request:

Isabelle Carrière, INSERM Unité 500, 39 avenue Charles Flahault, 34093 MONTPELLIER CEDEX 5, FRANCE. Tel: (33) 4 99 61 46 91; Fax: (33) 4 99 61 45 79; e-mail: [carriere@montp.inserm.fr](mailto:carriere@montp.inserm.fr)

<sup>1</sup> INSERM U500, 39 avenue Charles Flahault, 34093 Montpellier Cedex, France

<sup>2</sup> INSERM CIC-EC, Groupe Hospitalier Sud Réunion, 97410 Saint Pierre, France

<sup>3</sup> Department of Internal Medicine and Geriatrics, Montpellier University Hospital, 34295 Montpellier Cedex 5, France

Members of the EPIDOS study group: *Coordinators*: G. Bréart, P. Dargent-Molina (epidemiology), P.J. Meunier, A.M. Schott (clinical aspects), D. Hans (bone densitometry and ultrasound quality control), P.D. Delmas (biochemistry). *Principal investigators (center)*: C. Baudoin, J.L. Sebert (Amiens), M.C. Chapuy, A.M. Schott (Lyon); F. Favier, C. Marcelli (Montpellier); E. Hausherr, C.J. Menkes, C. Cormier (Paris); H. Grandjean, C. Ribot (Toulouse)

## Abstract

**Objective:** to identify the most important predictors of early disability incidence and devise a simple score of physical frailty.

**Methods:** a cohort of 545 high-functioning women aged 75 years and older was followed for seven years. Every year, the self-reported loss of at least one instrumental activity of daily living was chosen as definition of disability. An extension of the logistic regression for repeated responses, the random-effect model, was used to assess the effects of baseline predictors. The regression coefficients of the final multivariate model were scaled and rounded to create a practical score.

**Results:** the proportion of women reporting disability increased from 22.1 percent to 52.1 percent throughout the follow-up. In the multivariate model increasing age, lower performances in mobility and balance tests, bad perceived health, lower muscle strength, higher body mass index, lower educational level, lower reported physical activity were strong predictors of disability. Evaluating the predictive value of the simplified predictive score on an independent cohort gave a c-statistic equal to 0.71.

**Conclusion:** the use of a powerful fitting method allows to establish a hierarchy between the components of physical frailty and to provide a predictive score with substantial practical value for clinicians and public health professionals.

**Key-words:** longitudinal studies; logistic models; random-effects models; disability evaluation; frail elderly

---

<sup>4</sup> *Accepté par Journal of Clinical Epidemiology*

One important public health goal for an aging society is to minimize the impact of chronic diseases and impairments on the health status of older adults, to prevent dependency, and to improve their quality of life. If progress in medical care, together with better life conditions, account for a longer life span, the counterpart is that the number of very old people at risk of becoming disabled is increasing. Several studies suggest that the prevalence of moderate disability has declined in elderly population, but conflicting results exist about the trend for more severe stages [1] and the overall prevalence remains very high: 20 percent of older US adults have chronic disabilities.

Fortunately, older age is not uniformly associated with decline of performance [2], indicating the potential for effective interventions to promote more successful aging. The results of the few intervention programs designed to prevent functional decline in older people show that participants with relatively good functional status or moderate frailty are those who benefit the most of these programs [3], suggesting that interventions targeted early in the process of functional decline is potentially very effective. A concept of "frail subjects" is sometime used to define this target population [4]. Until now only minor attention has been paid to primary prevention of progressive disability among non-disabled old people. It seems useful to better characterize this early state of physical frailty associated with later disability.

However, many risk factors are involved in the disability process [5] and thus, a valid estimation of the weight of each factor is not straightforward. Most of the results are issued from cross-sectional studies whereas longitudinal studies are far more informative. But the analysis of repeated outcomes with possible recovery throughout time requires the use of non classical models that have only recently become available in statistical packages.

The first aim of this study is to identify the most important baseline predictors of the decline in instrumental activities of daily living (IADL) during a seven-year follow-up in a group of non-disabled women aged 75 years and over.

The second aim is to devise an indicator of physical frailty in order either to select healthy women at risk of becoming disabled within the next few years and liable to benefit from an intervention program, or to evaluate the global risks in populations.

## **MATERIALS AND METHODS**

### **Subjects**

This article is based on a cohort of 545 high-functioning women aged 75 years and older, selected from 1547 women recruited in Montpellier, one of the five participating centers of the French EPIDOS study on the risk of hip fracture [6]. In this center the follow-up was extended to seven years. To ensure that the selected women have not already begun the IADL disability process, we considered only those independent in IADL both at inclusion and at the first year of follow-up (i.e.: performing light or heavy housework, laundry, shopping for groceries or personal items, preparing meals, using a telephone, taking public transport, managing money and taking medication without assistance) [7].

In addition the data from the EPIDOS center of Toulouse were used to test the predictive value of our indicator. This independent sample consisted of 807 women free of disability at inclusion and followed-up for four years.

## **Baseline Evaluation**

The baseline evaluation was done in the local clinical center by trained doctors and nurses, and included a questionnaire, a physical examination, a functional assessment, and measurements of whole body bone mineral density (BMD) and body composition with lean mass (bone mineral excluded) and fat mass.

Educational level was assessed by a dichotomous variable: obtaining or not the old French Certificate of Elementary Education generally taken at the age of 14 years.

Current physical activity was estimated by questions concerning physical exercise (walking, gymnastics, cycling, swimming, gardening), and heavy housework. The time spent for each activity was converted into metabolic equivalent (MET) by week and then the overall total was calculated.

Physical performance was assessed by a series of standard tests: quantitative measurements of grip and quadriceps strength and calf circumference; time taken to stand up and sit down five times with arms across the chest (chair stands); time taken to tap one foot back and forth ten times between two circles placed 30 cm apart in a sitting position (foot tapping); subjects were asked twice to walk six meters at normal pace to determine gait speed and step length.

Standing balance was assessed by timed maintenance of the balance in progressively more challenging positions: side-by-side position, semi tandem position, and tandem position. Dynamic balance was assessed by testing the ability to walk with the heel of the front foot touching the big toe of the rear foot (tandem walk).

Corrected binocular visual acuity was measured at a distance of five meters with a Snellen letter test chart (decimal scale).

## **Outcome Variable**

Every year, women filled out a mail questionnaire in which they were asked whether they were able to do the eight basic IADL without assistance. Women were scored on a dependence scale equivalent to the number of activities for which they needed assistance. Our variable of interest is a repeated binary outcome taking the value one if the yearly IADL score is positive and otherwise zero. Other cut-off points (score  $\geq$  two and score  $\geq$  three) were also investigated to check if the same risk factors were found again.

## **Statistical Analysis**

The analysis includes six steps: 1) the sample was described and compared to the group of women with no IADL evaluation during the follow-up or missing data for at least one baseline covariable, using Student's t-test or Wilcoxon's test for quantitative variables and Chi square test for qualitative variables; 2) a comprehensive analysis was carried out in order to identify the baseline factors linked to the outcome variable of IADL decline. Every potential factor was tested in a univariate model, adjusted on time since inclusion and age at inclusion. 3) Predictors that were significant in the previous step were grouped into categories according to their source and availability in a typical clinical setting. Five categories of factors were used in our analysis: mobility, standing and dynamic balance, grip and quadriceps strength, body composition, perceived health. Several factors were not grouped: self-reported physical activity, visual acuity, educational level and body mass index (BMI) change since the age of 30 years. The predictors within each category were entered group-wise into a multivariate model. 4) The remaining significant predictors from each model were entered



into a global multivariate model and the final selection of the covariables was done using a descending step by step method. 5) The regression coefficients of the final model were scaled and rounded to integers to create a score easy to implement. Optimal scaled and rounded coefficients were calculated using the algorithm proposed by Cole [8]. This algorithm consists in finding the smallest common multiplier,  $k$ , which permits each estimated coefficient to be transformed into an integer without too much loss of precision. 6) The predictive value of this new scoring system was evaluated using the set of data provided by the EPIDOS center of Toulouse. The performance of this prognostic score was assessed using the c-statistic that is identical to the area under the receiver operator characteristic curve in the case of a binary outcome [9]. This statistic was calculated on the overall repeated observations throughout the 7 years of follow-up.

In longitudinal studies, the within subject responses (IADL) are correlated. This correlation was accounted for by using an extension of the logistic regression, the random-effect model [10]. The justification for applying this model to our data is extensively discussed in [11]. Briefly, this model has two basic characteristics: (i) subjects with incomplete responses across time are included in the analysis, (ii) subjects do not have to be measured at the same time points. In particular, the subjects who died during the seven years of follow-up are included in the analysis until the date of their death. This model allows the within-subjects dependency to vary from a subject to another, by the means of the random part ( $u_i$ ) of the covariables linear combination.

Let  $Y_{ij}$  denote the binary outcome (IADL > 0) corresponding to the  $j$ th response ( $j$ th year in the study,  $j=1$  to 7), evaluated at the time  $t_{ij}$ , of the  $i$ th subject ( $i=1$  to 545). Let also  $X_{ik}$  be the baseline covariates ( $k=1$  to  $p$ ). The estimated model is :

The parameters  $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_k$ 's are common to all the subjects, while the  $u_i$ 's are

$$\text{Logit}\left(P(Y_{ij} = 1 / t_{ij}, X_{ik}, u_i)\right) = \alpha_0 + \alpha_1 t_{ij} + \alpha_2 \text{age} + \alpha_3 (\text{age} * t_{ij}) + \sum_{k=1}^p \beta_k X_{ik} + u_i$$

individual parameters of propensity to become disabled, constant through time. The scaled rounded coefficients of the final scoring system (step 5 of the analysis) are provided by multiplying the estimated parameters,  $\alpha_1, \alpha_2, \alpha_3, \beta_k$ 's, by the optimal scaling multiplier,  $k$ , and rounding them to the nearest integers. The SAS procedure NLMIXED was used to estimate the parameters. The final multivariate model was validated with the WinBUGS statistical software [12]. The quartile values of each quantitative potential risk factor were determined, thus forming three groups: lower quartile, middle quartiles and higher quartile. This categorization makes it possible to compare the odds ratio between the predictive factors to be compared.

## RESULTS

Among the 691 women free of disability at inclusion and year one, data for at least one baseline factor were missing for 146 subjects; the final analysis thus includes 545 women. Subjects are described in table 1. A significant difference in age was found between the 545 selected women and the 146 excluded women. The women included in the analysis were younger (median age of 79 years compared to 80 years in the excluded group,  $p < 0.005$ ). Most of them (92.7 percent) lived at home.

The proportion of women reporting disability (table 2) increased from 22.1 percent at the end of the second year to 52.1 percent at the end of the seventh year. During the seven years,

recovery from disability – that is an IADL score going back to zero - was observed at least once in 155 women (28.4 percent). Some women gave intermittent answers. For example, among the 432 women still providing a response at year seven, 83 women (19.2 percent) had at least one observation missing in the previous years. The actual time between yearly assessments and inclusion varied from one woman to another and the range tended to be higher at the end of the follow-up. Consequently, time is considered as a continuous covariable in the subsequent models. Vital status was known throughout the seven years for all the participants, even for those who no longer sent back the questionnaires. The total number of deaths was 63 (11.6 percent) by the end of the study.

Tables 3a and 3b give the univariate odds ratios (OR) and the 95 percent confidence intervals (95%CI) for all the predictive factors significantly related to disability defined by a positive IADL score. The predictors are presented according to the groups of pre-selection for the final model. Twenty-one factors were found to be significantly associated with disability.

The final multivariate model is presented in table 4 with the risk factors in the order of importance. The parameters and the odds ratios corresponding to the time since inclusion, interaction age \*time and nine remaining significant risk factors are given. For three covariables (gait speed, physical activity and BMI), the middle category was removed from the final model since its parameter was very close to zero.

Concerning the predictors measured at baseline, the strongest belonged to the "mobility" group. Compared to the quartile of best performances, the odds ratio (95%CI) corresponding to lower performances was equal to 7.62 (3.86, 15.04), 6.31 (3.26, 12.19) and 5.31 (2.43, 11.59) for, respectively, time to complete five chair stands, gait speed and number of steps to walk six meters. In the multivariate model the first two remained significant. In second place, the group of balance tests (ten foot taps, standing in tandem position and tandem walk) was also very important but in the multivariate model the ability to stand in tandem position was the most discriminating. At last, the group "body composition" was also predictive, with measures such as BMI, relative fat mass being very significant in the univariate model. In the final model, BMI seems the best measure of body composition for predicting disability. Both the grip and quadriceps strength were significant in the univariate analysis but only the grip strength remains in the final model.

Concerning the factors collected by the questionnaire, the group "perceived health" was particularly noteworthy, with two variables remaining in the final model: self-rated health and fear of falling. Declared physical activity and educational level were also very significant.

The algorithm calculating a simpler score from the final model, leads to an optimal scaling multiplier (k), unique and equal to 16.44 (table 4). For a given woman, the score predicting functional dependence is easily provided by summing the scaled rounded coefficients for her observed risk factors (Cf. appendix).

The predictive value of this score was evaluated using the same sample of 545 women observed for seven years and the independent sample of 807 women included by the Epidos center of Toulouse. The c-statistic was equal respectively to 0.72 and 0.71, indicating a good level of prediction.

## **DISCUSSION**

### **Relevant Targeted Population**

Our study is limited to women aged 75 years or over. This segment of the population represents an interesting target for intervention. Women represent the largest proportion of the elderly and their patterns of functional decline seem to be different from the men. Concerning light disability measured with the IADL indicator and lower body physical ability score, it is suggested that if incidence rates are roughly equal in both genders, the prevalence rates are higher in women, indicating a longer time spent in disability. This is consistent with several studies demonstrating that both recovery and death rates are lower for women than for men [13]. Women with light disability seem to outlive men from 75 to 85 years to the oldest ages, and for that reason they are at higher risk of developing severe disability in the very old ages [14].

In the age group of our study the life expectancy is still high (13 and 6.5 years for respectively the 75 and 85 year old French women). However as functional decline had begun among some subjects we needed to select a sample of women free of disability for two consecutive years to ensure that we studied incident disability only. Our high-functioning sample is therefore closer to a younger age group. Further studies are necessary to define the age class in which the screening of at risk subjects would be the most cost-effective. This optimal age interval is likely to be between 65 and 75 years but it will tend to shift upwards with increase in life expectancy.

### **Definition of disability onset**

Several instruments are currently used to assess functional limitation. Since our purpose was to predict the early stage of dependency, the IADL scale was preferred. Wolinsky et al. [15] showed that deterioration in IADL functioning is a strong predictor of future decline. This outcome seems more suitable to our goal than the Katz's ADL scale that was demonstrated to only capture disability at the extreme end of the process and to be thus unable to discriminate levels of disability in community populations where the prevalence of ADL disability tends to be low [16].

The loss of IADL has both a cognitive and motor origin [17]. In particular complex activities (handling money, phone use, self-medication) were demonstrated to be correlated with cognitive impairment [18]. Our score is therefore predictive of global disability including cognitive impairment.

### **Powerful Fitting Method**

One of the strengths of this study is the longitudinal design with outcomes measured yearly over a seven-year period with minimal loss to follow-up. Even though our sample is not very large (n=545), the number of repeated evaluations by subject and the use of a powerful model allow us to detect numerous risk factors. To our knowledge it is the first time that a random effect model for binary responses is used to predict disability in elderly people. The statistical methods (Cox models) analyzing the time to disability onset cannot be used, as the proportion of women who reverse from disability (28 percent) is very high. Several authors [19, 20] used random effect models or generalized estimating equations (GEE) models but they considered the outcome as a normally distributed response. In general the scores measuring disability are finite ordinal scales, positively skewed, and they cannot be considered as quantitative. Compared to GEE, we advocated [11] that the random effect model for binary response is more suitable and has interesting advantages.

### **Hierarchy Between The Frailty Components**

Firstly, our statistical analysis modeled the impact of time on disability occurrence. Every additional year brings a marked increase in the probability of being dependent. Moreover we found an interaction between age at entry and time: the risk of disability is accelerated in the oldest women.

Our final model also gives the relative weight of baseline risk factors in predicting disability during the following seven years. This combination of predictors may be seen as a physical frailty score. Physical frailty was defined as a physiological state of increased vulnerability to stressors that results from decreased physiological reserves of multiple physiological systems [21,22]. In community dwelling people this frailty state is often asymptomatic because the individuals use high compensatory strategies. This instable equilibrium can be easily broken by minimal environmental challenges.

Clinical experts and geriatricians have proposed various criteria for defining frailty. Campbell et al [23] defined frailty as a loss of the person's capability to withstand minor environmental stresses and suggested four essential reserve capacities: musculoskeletal function, aerobic capacity, cognitive function and nutritional state. Fried et al [24] introduced the concept of "frailty syndrome" based on five domains: nutrition (loss of weight), muscle strength (grip test), endurance (exhaustion), mobility (walking speed) and physical activity. During the last decade, three other domains have often been cited in the literature: balance, motor processing (speed of movement, coordination) and cognition [4,25].

Our large screening of potential factors in a prospective study confirms the important role of five of these dimensions and establishes the following hierarchy: mobility (gait speed and chair stand), balance (tandem position test), nutrition (BMI), muscle strength (grip test) and physical activity. We also found a marked impact of perceived health (self-rated health and fear of falling) on the probability of becoming dependent. Several authors also found an association of perceived health with mortality [26], or chronic conditions [27], or functional capacity loss [28], or physical inactivity [29]. But, except for mortality, this association was demonstrated in cross-sectional studies and interpreted as a consequence of physical deterioration. In contrast, our findings suggest that poor perceived health increases the risk of later physical limitations. At last, women with low educational level were found to be significantly more at risk of becoming disabled. This finding is supported by other results [30]. Low educational level may be interpreted as a proxy for characteristics such as low income or adverse life conditions or limited access to prevention programs.

The nutritional dimension of the frailty syndrome is usually defined as a shrinking process and identified by a low body weight or a lean mass loss [24]. By contrast, in the univariate analysis we found a positive association between the risk of disability and body weight, BMI, fat body mass and gain of weight from the age of 30 years. The prospective work of Launer et al [31] suggests that high BMI and weight loss are risk factors for decreased mobility in elderly women. In our cohort there are few women with low BMI. Further studies are needed to identify the mechanism through which BMI may affect disability and to test whether the J shaped risk curve found for the relation between BMI and mortality [32] is also relevant here, particularly in later functional dependence.

In the Epidos cohort cognitive impairment was not evaluated at inclusion and therefore this domain is not present in our final score. However in this highly selected group, no diagnosis of moderate or severe dementia was reported at baseline and cognitive impairment if present is likely to be mild.

In our analysis, risk factors such as history of depression, heart disease, cancer or body mass density, were not found to be predictive of disability in contradiction with some cross-

sectional analyses [33]. This finding suggests that functional performance tests are more relevant for evaluating reserve capacities than screening of chronic diseases. Similarly variables concerning baseline food intakes were also non significant in our cohort but the questionnaire focused mainly on calcium intake and probably lacked exhaustiveness.

### **Applicability of the Predictive Score**

We present a predictive score using an integer-based linear combination of risk factors. The good correspondence between the predicted and observed disability status in an independent sample suggests that the scale can provide accurate risk evaluation. However it may be worthwhile to validate our score by measuring the misclassification rate in new cohorts of subjects.

An important requirement for the practical use of a predictive score is its simplicity in order to ensure its applicability in various settings. Fortunately the selected predictors in the final model are very simple to collect. They are a combination of self-reported data and objective measures particularly easy to implement, even by non-clinical health care workers. Furthermore scaling and rounding equation coefficients to integers contributes to this simplicity.

Our study attempts to quantify physical frailty viewed as a physiological precursor of disability. The predictive score computed thus has considerable practical value for clinicians and public health professionals: it can be used to select subjects apparently in good health but who are at risk of becoming disabled and to which can then be recommended a well-designed physical training program in order to postpone dependence.

### **Acknowledgements:**

The authors are grateful to Dr Hélène Grandjean and the unit INSERM U558, who provided the data from the Toulouse EPIDOS center and to Annie Lacroux for editorial assistance.

## REFERENCES

1. Freedman VA, Martin LG, Schoeni RF. Recent trends in disability and functioning among older adults in the United States: a systematic review. *JAMA* 2002;288:3137-3146.
2. Seeman TE, Charpentier PA, Berkman LF, et al. Predicting changes in physical performance in a high-functioning elderly cohort: MacArthur studies of successful aging. *J Gerontol* 1994;49:M97-108.
3. Gill TM, Baker DI, Gottschalk M, et al. A program to prevent functional decline in physically frail, elderly persons who live at home. *N Engl J Med* 2002;347:1068-1074.
4. Ferrucci L, Guralnik JM, Studenski S, et al. Designing randomized, controlled trials aimed at preventing or delaying functional decline and disability in frail, older persons: a consensus report. *J Am Geriatr Soc* 2004;52:625-634.
5. Stuck AE, Walthert JM, Nikolaus T, et al. Risk factors for functional status decline in community-living elderly people: a systematic literature review. *Soc Sci Med* 1999;48:445-469.
6. Dargent-Molina P, Favier F, Grandjean H, et al. Fall-related factors and risk of hip fracture: the EPIDOS prospective study. *Lancet* 1996;348:145-149.
7. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969;9:179-186.
8. Cole TJ. Scaling and rounding regression coefficients to integers. *Appl Stat* 1993;42:261-268.
9. Harrell Fe JR, Lee KL, Califf RM, et al. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;3:143-152.
10. Diggle PJ, Liang KY, Zeger SL. *Analysis of longitudinal data*. Oxford: Oxford Science Publications;1994.
11. Carriere I, Bouyer J. Choosing marginal or random-effects models for longitudinal binary responses: application to self-reported disability among older persons. *BMC Med Res Methodol* 2002;2:15.
12. Spiegelhalter DJ, Thomas A, Best NG, et al. *BUGS: Bayesian Inference Using Gibbs Sampling, Version 5.0, (version ii)*. MRC Biostatistics Unit, Cambridge, 1996.
13. Oman D, Reed D, Ferrara A. Do elderly women have more physical disability than men do? *Am J Epidemiol* 1999;150:834-842.
14. Von Strauss E, Agüero Torres H, Kareholt I, et al. Women are more disabled in basic activities of daily living than men only in very advanced ages: a study on disability, morbidity, and mortality from the Kungsholmen Project. *J Clin Epidemiol* 2003;56:669-677.
15. Wolinsky FD, Johnson RJ. The use of health services by older adults. *J Gerontol* 1991;46:S345-S357.
16. Kovar MG, Lawton P. Activities and instrumental activities of daily living. *Ann Rev Geriatr Gerontol* 1994;14:57-75.
17. Thomas VS, Rockwood K, McDowell I. Multidimensionality in instrumental and basic activities of daily living. *J Clin Epidemiol* 1998;51:315-321.
18. Barberger-Gateau P, Commenges D, Gagnon M, et al. Instrumental activities of daily living as a screening tool for cognitive impairment and dementia in elderly community dwellers. *J Am Geriatr Soc* 1992;40:1129-34.
19. Beckett LA, Brock DB, Lemke JH, et al. Analysis of change in self-reported physical function among older persons in four population studies. *Am J Epidemiol* 1996;143:766-778.
20. Dunlop DD, Hughes SL, Edelman P, et al. Impact of joint impairment on disability-specific domains at four years. *J Clin Epidemiol* 1998;51:1253-1261.

21. Fried LP, Ferrucci L, Darer J, et al. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. *J Gerontol A Biol Sci Med Sci* 2004;59:255-263.
22. Bortz Wm 2ND. A conceptual framework of frailty: a review. *J Gerontol A Biol Sci Med Sci* 2002;57:M283-M288.
23. Campbell AJ, Buchner DM. Unstable disability and the fluctuations of frailty. *Age Ageing* 1997;26:315-318.
24. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 2001;56:M146-M156.
25. Studenski S, Hayes RP, Leibowitz RQ, et al. Clinical Global Impression of Change in Physical Frailty: development of a measure based on clinical judgment. *J Am Geriatr Soc* 2004;52:1560-1566.
26. Yu ES, Kean YM, Slymen DJ, et al. Self-perceived health and 5-year mortality risks among the elderly in Shanghai, China. *Am J Epidemiol* 1998;147:880-890.
27. Hoeymans N, Feskens EJ, Kromhout D, et al. The contribution of chronic conditions and disabilities to poor self-rated health in elderly men. *J Gerontol A Biol Sci Med Sci* 1999;54:M501-M506.
28. Jylha M, Guralnik JM, Balfour J, et al. Walking difficulty, walking speed, and age as predictors of self-rated health: the women's health and aging study. *J Gerontol A Biol Sci Med Sci* 2001;56:M609-M617.
29. Chin A Paw MJ, Dekker JM, Feskens EJ, et al. How to select a frail elderly population? A comparison of three working definitions. *J Clin Epidemiol* 1999;52:1015-1021.
30. Manton KG, Stallard E, Corder L. Education-specific estimates of life expectancy and age-specific disability in the U.S. elderly population: 1982 to 1991. *J Aging Health* 1997;9:419-450.
31. Launer LJ, Harris T, Rumpel C, et al. Body mass index, weight change, and risk of mobility disability in middle-aged and older women. The epidemiologic follow-up study of NHANES I. *JAMA* 1994;271:1093-1098.
32. Calle EE, Thun MJ, Petrelli JM, et al. Body-mass index and mortality in a prospective cohort of U.S. adults. *N Engl J Med* 1999;341:1097-1105.
33. Nourhashemi F, Andrieu S, Gillette Guyonnet S, et al. Instrumental activities of daily living as a potential marker of frailty: a study of 7364 community-dwelling elderly women (the EPIDOS study). *J Gerontol A Biol Sci Med Sci* 2001;56:M448-M453.

TABLE 1: Comparison of women included in the analysis versus those excluded because of missing data

Parameters	Included in the analysis (n = 545)	With missing data (n = 146)	p value*
Age in years [median (IQR)]	79 (76-81)	80 (77-82)	< 0.005
Body mass index in kg/m <sup>2</sup> [mean (SD)]	25.2 (3.8)	25.9 (4.0)	0.05
Comparative health (%)			
Healthier	43.3	34.1	
As healthy	55.0	65.9	
Less healthy	1.7	0.0	0.04
Living at home (%)	92.7	95.2	0.22
Educational level: primary school (%)	46.8	43.0	0.41
Hospitalized during the previous year (%)	15.6	18.6	0.38
History of cancer (%)	19.8	21.9	0.57
History of cardiovascular disease (%)	26.4	23.3	0.44
History of depression (%)	0.9	2.1	0.38

\*Wilcoxon's test, or Student *t*-test, or Chi 2 test, or Fisher's exact test.

TABLE 2: Description of the data by year of assessment (n=545)

	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
Number of women with disability assessment	542	525	498	438	413	432
Number of deaths for the preceding year	1	8	15	13	15	11
Number of women without disability assessment	2	11	23	70	80	50
Percent of disabled women						
IADL > 0	22.1	27.4	33.7	36.3	42.4	52.1
IADL > 1	10.0	10.7	18.1	17.8	27.9	33.8
IADL > 2	5.4	5.7	10.2	12.1	19.4	24.1
Time since inclusion (years)						
Median	1.98	2.97	3.94	4.96	6.06	6.99
Range	1.95-2.37	2.95-3.26	3.90-4.26	4.89-5.89	5.92-6.87	6.91-7.59



TABLE 3a: Univariate models, time and age-adjusted odds-ratios

	Odds-ratio	95% Confidence Interval
<b>MOBILITY</b>		
Gait speed (m/sec)		
< 0.78	6.31****	3.26, 12.19
0.78 – 1.0	2.32**	1.32, 4.08
≥ 1.0	1	
Number of steps to walk 6m		
< 11	1	
11 – 13	1.01	0.50, 2.05
≥ 13	5.31****	2.43, 11.59
Time to complete five chair stands (sec)		
< 9.6	1	
9.6 – 13	2.35**	1.30, 4.24
≥ 13	7.62****	3.86, 15.04
<b>BALANCE</b>		
Time to complete 10 foot taps (sec)		
< 4.25	1	
4.25 – 6.6	2.05*	1.14, 3.67
≥ 6.6	5.35****	2.75, 10.43
Time to stand in tandem position (sec)		
0	6.07****	2.86, 12.88
0.1 – 5	1.74	1.00, 3.01
≥ 5	1	
Difficulty in performing a tandem walk		
Able to do 4 consecutive tandem steps	1	
Unable without stepping off	1.27	0.74, 2.17
Unable	8.36***	2.70, 25.91
<b>STRENGTH</b>		
Grip strength (Kpa)		
< 47	2.81**	1.43, 5.52
47 - 61	1.36	0.77, 2.40
≥ 61	1	
Quadriceps strength (Newton/lower leg length in cm)		
3.52	2.69**	1.38, 5.24
3.52 – 4.95	1.50	0.85, 2.66
≥ 4.95	1	
Visual acuity		
≤ 4	1.77	0.97, 3.26
> 4	1	

\* p < 0.05

\*\* p < 0.01

\*\*\* p < 0.001

\*\*\*\* p < 0.0001

TABLE 3b: Univariate models, time and age-adjusted odds-ratios

	Odds-ratio	95% Confidence Interval
<b>BODY COMPOSITION</b>		
Weight (kg)		
< 53	1	
53 – 65	0.94	0.52, 1.70
≥ 65	2.02*	1.02, 3.98
Lean mass/body mass		
< 0.54	1	
0.54 – 0.63	0.45**	0.25, 0.80
≥ 0.63	0.35**	0.18, 0.69
Fat mass/body mass		
< 0.31	1	
0.31 – 0.40	1.09	0.61, 1.93
≥ 0.40	2.87**	1.52, 5.42
Boby mass index (kg/m <sup>2</sup> )		
< 22.8	1	
22.8 – 27.6	0.87	0.49, 1.54
≥ 27.6	2.97**	1.55, 5.70
Waist/hip circumference		
< 0.81	1	
0.81 – 0.90	1.40	0.79, 2.48
≥ 0.90	1.82	0.93, 3.58
Weight change from the age of 30 yrs(kg)		
< 0	1	
0 – 12	1.02	0.56, 1.85
≥ 12	2.02*	1.02, 4.02
BMI change from the age of 30 yrs(kg/m <sup>2</sup> )		
< 2	1	
2 – 7	1.10	0.62, 1.95
≥ 7	2.15*	1.09, 4.24
Physical activity (MET/week)		
< 6.9	3.53***	1.78, 7.01
6.9– 36.6	1.58	0.90, 2.79
≥ 36.6	1	
Educational level		
Primary school	2.33***	1.46, 3.72
> primary school	1	
<b>PERCEIVED HEALTH</b>		
Self-rated health		
Bad or very bad	7.26**	2.01, 26.25
Good or very good	1	
Compared health / women of same age		
As or less healthy	2.42***	1.51, 3.90
Healthier	1	
Fear of falling		
Yes	2.57****	1.61, 4.12
No	1	

\* p < 0.05

\*\* p < 0.01

\*\*\* p < 0.001

\*\*\*\* p < 0.0001

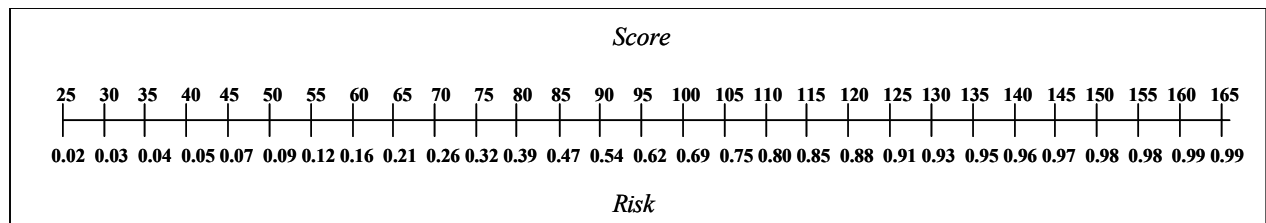
TABLE 4: Multivariate model and the derived simplified model (with rounded/scaled coefficients)

	Final model			Scaled and rounded coefficient (k=16.44)*
	Parameter	Odds-ratio	95% Confidence Interval	
Time since baseline evaluation (years)	0.38	1.46	1.33, 1.60	6
Age† * Time since baseline evaluation	0.03	1.03	1.02, 1.05	1
<b>MOBILITY</b>				
Gait speed (m/sec)				
< 0.78	0.57	1.76	1.04, 2.98	9
Time to complete five chair stand (sec)				
9.6 – 13	0.48	1.61	0.92, 2.82	8
≥ 13	1.23	3.41	1.74, 6.67	20
<b>PERCEIVED HEALTH</b>				
Self-rated health				
Bad or very bad	1.25	3.47	1.07, 11.33	20
Fear of falling	0.46	1.58	1.01, 2.46	7
<b>BALANCE</b>				
Time to stand in tandem position (sec)				
0	1.11	3.04	1.49, 6.21	18
0.1 – 5	0.33	1.39	0.83, 2.34	5
<b>NUTRITION</b>				
Boby mass index (kg/m <sup>2</sup> )				
≥ 27.6	0.89	2.43	1.48, 3.98	15
<b>MUSCLE STRENGTH</b>				
Grip strength (Kpa)				
< 47	0.58	1.78	1.07, 2.95	9
≥ 47	1			
Physical activity (MET/week)				
< 6.9	0.51	1.67	0.99, 2.82	8
Educational level				
Primary school	0.45	1.57	1.02, 2.44	7

\* k: scaling coefficient

† years over 74

## Appendix



For example, a 78 years old woman, free of disability, with a grip strength < 47 Kpa, a BMI  $\geq$  27.6, bad self-rated health and who has fear of falling, will have, 3 years later, a score =  $6 \times (3 \text{ years since baseline}) + (4 \text{ years over } 74 \text{ years old} \times 3 \text{ years since baseline}) + 9 \text{ (grip strength)} + 15 \text{ (BMI)} + 20 \text{ (self-rated health)} + 7 \text{ (fear of falling)} = 81$  which corresponds to a risk for dependency = 0.39.

Random-effect models for ordinal responses: Application to self-reported disability among older persons<sup>5</sup>.

I. Carrière<sup>1</sup> and J. Bouyer<sup>2</sup>

<sup>1</sup>INSERM Unité 500, 39 avenue Charles Flahault, 34093 MONTPELLIER, FRANCE

<sup>2</sup>INSERM Unité 569, 82 rue du Général Leclerc, 94276 LE KREMLIN BICETRE, France ; INED ; IFR69

Corresponding author:

Isabelle Carrière, INSERM Unité 500, 39 avenue Charles Flahault, 34093 MONTPELLIER, France

Email : [carriere@montp.inserm.fr](mailto:carriere@montp.inserm.fr)

Tel : 04 99 23 23 72

Fax: 04 99 61 45 79

Short title: Random-effect models for ordinal data

---

<sup>5</sup> Article en révision pour la Revue d'Epidémiologie et de Santé Publique

## **Abstract**

### Background

Longitudinal studies with ordinal repeated outcomes are now widespread in epidemiology and clinical research. The statistical analysis of these studies combines two difficulties: the choice of the best ordinal model and taking into account correlations for within-subject responses.

### Methods

Random-effect models are of particular value in this context and we propose here a fitting strategy. The various ordinal models extended to the case of repeated responses are detailed. We explain how the choice of model imposes constraints on the random effect structure. Model selection criteria and goodness-of-fit measures are also presented.

These issues are dealt with for the example of self-reported disability in older women assessed annually over a period of seven years.

### Results

The proportionality of the odds ratios was validated for the covariables "physical capacity" and "visual acuity". In contrast the impact of the covariables "pain" and "gait speed" differs according to the levels of disability. The restricted partial proportional odds model was found to be as efficient as the full generalized ordered logit model while the stereotype model appeared to give poorer fit.

### Conclusions

The random-effects models presented in this paper allow to analyze longitudinal studies that are of particular interest in epidemiology and to take into account the ordinal nature of the outcome. Furthermore the impact of the risk factors can be modeled according to the response levels. This approach can be useful for a better understanding of complex processes of evolution.

*Longitudinal studies. Random-effects models. Ordinal response. Disability evaluation*

## **Résumé**

### Position du problème

Les études longitudinales avec des réponses ordinales répétées sont maintenant fréquentes en épidémiologie et recherche clinique. L'analyse statistique de ces études comporte deux écueils : le choix du meilleur modèle ordinal et la prise en compte des corrélations intra-sujets des réponses.

### Méthodes

Les modèles à effets aléatoires sont particulièrement bien adaptés dans ce contexte et nous proposons dans cet article une stratégie de modélisation. Les différents modèles ordinaux étendus au cas des réponses répétées sont détaillés. Nous expliquons comment le choix du modèle impose des contraintes sur la structure des effets aléatoires. Nous présentons aussi des critères de sélection et des mesures d'adéquation des modèles aux données.

Ces questions sont illustrées par un exemple d'étude de l'incapacité mesurées chaque année, chez des femmes âgées, sur une période de sept ans.

## Résultats

La proportionnalité des odds ratios a été validée pour les covariables "force de relevé" et "acuité visuelle". Par contre l'impact des covariables "douleurs" et "vitesse de marche" est différent selon le niveau d'incapacité. Le modèle restreint à odds partiellement proportionnels était aussi efficace que le modèle logistique ordonné complet tandis que le modèle stéréotype était moins performant.

## Conclusion

Les modèles à effets aléatoires présentés dans cet article permettent d'analyser les études longitudinales qui sont particulièrement intéressantes en épidémiologie en prenant en compte le caractère ordinal de la réponse. De plus, l'impact des facteurs de risque peut être modélisé selon les niveaux de la variable réponse. Cette approche peut être utile pour mieux comprendre les profils d'évolution complexes.

*Etudes longitudinales. Modèles à effets aléatoires. Réponse ordinale. Evaluation de l'incapacité.*

## Background

In epidemiology and clinical research, dimensions such as quality of life, pain, fatigue, depression, cognitive impairment, mood and physical disability are increasingly studied as outcomes of interest. In elderly people, for example, measurements of functional limitation are of key importance. In recent years, a large number of scales have been developed for the evaluation of these dimensions, most of them being ordinal indicators. Moreover, the study of the succession of stages of disability, together with analyses of the risk factors for future deterioration is a major concern in gerontology. Cross-sectional studies are inappropriate as they present various limitations due to biases such as selective removal or differences in the proportions of impaired individuals between age groups, which may also be confounded with changes over generations [1].

Therefore, studies on risk factors for disability among elderly people require statistical tools able to take into account at the same time the ordinal nature of the outcome variable, the correlation between repeated observations for a subject and that the covariables may be time-dependent (they may vary over time for a subject). These issues may be addressed by means of random-effect models. However, few studies have used models of this type for the analysis of ordinal responses and, indeed, their application raises several problems.

In this paper, we describe a strategy for selecting and fitting multivariate random-effect models to longitudinal ordinal responses. We firstly present the models and their constraints coming from both the random-effect and the ordinal aspects of the models. Then, we provide tools for comparing their goodness of fit. Finally, this strategy is applied and the models are compared in an epidemiological study of changes over time of a disability indicator among a cohort of elderly women.

## Methods

In longitudinal studies, each subject is assessed at different time points and these intra-individual responses are correlated. These correlations between individual responses must be accounted for by appropriate analysis methods [2] such as the generalized linear mixed model (GLMM), also known as the multilevel model or the conditional model [3]. The use of random effects in linear models for normal responses is well established. By contrast, random effects have only recently been incorporated into models for categorical data, due to complexities of implementation.

A basic characteristic of these models is the introduction of random subject effects into the linear predictor, which has led to random-effects models being described as "subject-specified models" [4]. The random part of the linear covariable combination describes and explains the structure of the longitudinal correlation and indicates the magnitude of between-individual heterogeneity.

Let  $Y_{ij}$  denote an outcome corresponding to the  $j$ th response (ie the response at time  $j$ ) of the  $i$ th subject ( $i=1$  to  $N$ ). Let  $\mathbf{X}_{ij}$  be a design matrix of time dependent covariates and  $\mathbf{Z}_{ij}$  a submatrix of  $\mathbf{X}_{ij}$ . Let  $\mathbf{U}_i$  denote the vector of random effect values for the subject  $i$ . Conditional on  $\mathbf{U}_i$ , a GLMM resembles an ordinary GLM. Let  $\mu_{ij}=E(Y_{ij}/\mathbf{U}_i)$  denote the mean of the conditional distribution of  $Y_{ij}$  given  $\mathbf{U}_i$ . The linear predictor for a GLMM therefore takes the form:

$$g(\mu_{ij}) = \mathbf{X}_{ij}'\beta + \mathbf{Z}_{ij}'\mathbf{U}_i$$

where  $g(\cdot)$  is a link function and  $\mathbf{U}_i$  is generally assumed to be normally distributed.



For the fixed parameter  $\beta$ , estimates are obtained conditionally to individual random effects. Various estimation procedures have been proposed. The most attractive of these procedures directly maximizes an approximate likelihood obtained by integration and is implemented for instance, in the SAS NLMIXED procedure. The likelihood is the integral, over the unobserved random effects, of the joint distribution of the data and the random effects. With non-linear models, computational intensive integration methods, such as the Gauss-Hermite quadrature [5], are required to evaluate this likelihood. Having approximated the likelihood, standard maximization methods, such as the Newton-Raphson method, are used to obtain the estimates.

Several models have been proposed for the analysis of ordinal responses in the case of non correlated data [6-7]. These models can be extended to the case of repeated responses. However, the choice of the model imposes constraints on the random effect structure. We present here two types of ordinal models : the proportional odds models and the stereotype model.

### The proportional odds mixed models

Let  $Y_{ij}$  now denote an ordinal outcome with K categories.

The cumulative logit model was the first model proposed for ordinal responses. The ordinal nature of the response was taken into account by considering the cumulative probabilities  $P(Y_{ij} \geq c)$ , ordered as follows  $1=P(Y_{ij} \geq 0) \geq P(Y_{ij} \geq 1) \geq \dots P(Y_{ij} \geq c) \geq \dots P(Y_{ij} \geq K)$ .

The model can be represented as follows:

$$\log\left(\frac{P(Y_{ij} \geq c | \mathbf{X}_{ij}, \mathbf{U}_i)}{P(Y_{ij} < c | \mathbf{X}_{ij}, \mathbf{U}_i)}\right) = \alpha_c + \mathbf{X}'_{ij}\beta + \mathbf{Z}'_{ij}\mathbf{U}_i \quad (1)$$

where  $\mathbf{U}_i$  are the random effects specific to subject  $i$  and  $\alpha_c$  is the intercept for category  $c$ . In this model, the random effects ( $\mathbf{U}_i$ ) and the fixed effects ( $\beta$ ) are independent of the response level ( $c$ ). Definition of the model is completed by specifying a distribution for  $\mathbf{U}_i$ . This distribution is often assumed to be independent, multinormal, with a mean of 0 and a covariance matrix  $\Sigma$ .

In practice, the computer time required increases substantially as random effects are introduced, and the procedure may fail if more than a few random effects are introduced into the model.

Let us now consider a simpler case, in which only a random intercept is introduced into the model:

$$\log\left(\frac{P(Y_{ij} \geq c | \mathbf{X}_{ij}, u_i)}{P(Y_{ij} < c | \mathbf{X}_{ij}, u_i)}\right) = \alpha_c + \mathbf{X}'_{ij}\beta + u_i \quad (2)$$

The assumption that  $\beta$  is independent of response level implies that the log-odds ratio for a given covariable is the same for all levels of the response. This led McCullagh to refer to this model as the "proportional odds model" [8].

The consideration of cumulative properties confers several useful properties on the model: (1) The proportional odds model is invariant if the codes for the response  $\mathbf{Y}$  are reversed, resulting only in a change in the sign of the regression parameters, (2) the proportional odds model is invariant if the categories of the ordinal response are collapsed [6]. The parameter vector,  $\beta$ , remains unchanged if categories are combined, with only the intercepts  $\alpha_c$  affected.

The proportional odds model is often presented as designed for ordinal response categories monotonically related to an underlying continuous latent variable  $\mathbf{Z}$ . The relationship between  $\mathbf{Y}$  and  $\mathbf{Z}$  is such that the parameters  $\alpha_c$  are the cut-off points on the continuum  $\mathbf{Z}$  and that  $Y_{ij}$  equals  $c$  if  $Z_{ij}$  lies between  $\alpha_c$  and  $\alpha_{c+1}$ . However, as pointed out by McCullagh [9], the existence of  $\mathbf{Z}$  is usually unverifiable in practice.

However, the use of a random effect,  $u_i$ , independent of response category is based on the notion that a unique unknown continuous phenomenon underlies the ordinal response. It is possible to consider a model including differential random effects,  $u_{ic}$ , by category of  $\mathbf{Y}$ . However, estimation is more complicated in this extended model because constraints must be added to ensure that intercept ordering is not violated. For a given response category,  $c$ , the intercept is then:  $\alpha_c + u_{ic}$  and therefore the following must be true:  $\alpha_1 + u_{i1} \geq \alpha_2 + u_{i2} \geq \dots \geq \alpha_K + u_{iK}$ .

The proportional odds assumption ( $\beta$  independent of the response level  $c$ ) may be too strict and should, in any case, be tested. For this purpose the proportional odds model may be compared to more general ordinal models that permit the effects,  $\beta$ , to vary according to the categories of  $\mathbf{Y}$ . The most general model of this type is called the generalized ordered logit model and can be expressed as follows:

$$\log \left( \frac{P(Y_{ij} \geq c | \mathbf{X}_{ij}, u_i)}{P(Y_{ij} < c | \mathbf{X}_{ij}, u_i)} \right) = \alpha_c + \mathbf{X}_{ij}' \beta_c + u_i \quad (3)$$

In the proportional odds model (2), the ordered nature of the response is expressed exclusively by the mean of the ordered intercepts,  $\alpha_c$ . In contrast, in model (3), the covariates are involved in this ordering [10]:

The hierarchy of cumulative probabilities implies the same ordering for the logits:  $\text{logit}(P(Y_{ij} \geq 1)) \geq \dots \geq \text{logit}(P(Y_{ij} \geq c)) \geq \dots \geq \text{logit}(P(Y_{ij} \geq K-1))$ . Thus, the linear predictors must also be identically ordered. For a single covariable, and as the random effect is thought to be independent of the response level, this relationship becomes:  $\alpha_1 + \beta_1 x \geq \alpha_2 + \beta_2 x \geq \dots \geq \alpha_{K-1} + \beta_{K-1} x$ . These  $K-1$  regression lines are not necessarily parallel and they may cross for some values of  $x$ , resulting in violation of the previous hierarchy, although this crossing of regression lines generally occurs outside the range of observed data. It may nonetheless be problematic if extrapolation of the model beyond the range of observed covariables is required, and it is necessary to check the order of the estimated linear predictors.

If only some of the covariables are consistent with the assumption of proportional log-odds, the unconstrained partial proportional odds models can be used [11]:

$$\log \left( \frac{P(Y_{ij} \geq c | \mathbf{X}_{ij}, u_i)}{P(Y_{ij} < c | \mathbf{X}_{ij}, u_i)} \right) = \alpha_c + \mathbf{X}_{ij}^1 \beta_c^1 + \mathbf{X}_{ij}^2 \beta_c^2 + u_i \quad (4)$$

$\mathbf{X}_{ij}^1$ : the submatrix of  $\mathbf{X}_{ij}$  of the  $q$  covariables with non-proportional odds ratio

$\mathbf{X}_{ij}^2$ : the submatrix of  $\mathbf{X}_{ij}$  of the  $p-q$  covariables with proportional odds ratio

### The stereotype mixed model

In cases in which the ordinal categories of  $\mathbf{Y}$  cannot be considered as a discrete version of a continuous variable, an extension of the polytomous logistic model [12] that accounts for the ordering of the response categories can be used. This model is called the stereotype model [13].

$$\log \left( \frac{P(Y_{ij} = c | \mathbf{X}_{ij}, u_i)}{P(Y_{ij} = 0 | \mathbf{X}_{ij}, u_i)} \right) = \alpha_c + \phi_c \mathbf{X}'_{ij} \boldsymbol{\beta} + u_i \quad (5)$$

with  $0 < \phi_1 < \phi_2 < \dots < \phi_k = 1$

The constraints on the  $\phi$ 's completely ensure response level ordering. The random effects ( $u_i$ ) and the fixed effects ( $\boldsymbol{\beta}$ ) are then independent of the response level. It is assumed that a single linear combination of covariables is adequate to distinguish between all outcome levels. More complex stereotype models have been recently introduced using between 2 and (K-2) linear combinations [13]. If two linear combinations are used the model is written:

$$\log \left( \frac{P(Y_{ij} = c | \mathbf{X}_{ij}, u_i)}{P(Y_{ij} = 0 | \mathbf{X}_{ij}, u_i)} \right) = \alpha_c + \phi_c^1 \mathbf{X}'_{ij} \boldsymbol{\beta}^1 + \phi_c^2 \mathbf{X}'_{ij} \boldsymbol{\beta}^2 + u_i$$

The polytomous logistic model is equivalent to a stereotype model with (K-1) linear combination.

Several other models have been used for ordinal responses, including the adjacent category logistic model and the continuation ratio model, but these models are designed for situations in which the subject must 'pass through' one category to reach the next [15-16]. They will not be described in this paper.

### Model selection and goodness of fit

Model selection includes the choice of the type of model and variable selection within a model type. In this framework, the parameters estimating method with numerical integration has the advantage of being based on likelihood statistics. Thus, models can be ordered according to likelihood-based measures, such as Akaike's information criterion or Schwarz's Bayesian criterion. Moreover, for variable selection, nested models can be compared with the likelihood ratio test.

However, as in non-longitudinal cases, a more general way of evaluating goodness-of-fit is to calculate the degree to which predicted values agree with observed values. In the generalized linear mixed model, predicted values are calculated from parameter estimates ( $\hat{\boldsymbol{\beta}}$ ) and empirical Bayes estimates of random effects ( $\hat{u}_i$ ) [17]. These predicted values can be obtained directly by means of the PREDICT instruction of the SAS NLMIXED procedure. The concordance between the observed and estimated ordinal responses can be assessed by one of the standard measures of ordinal association: gamma, Kendall's tau-b, for example [18]. However, Vonesh suggested that the coefficient of determination and the concordance correlation coefficient could be used [19]. The first one of these coefficients, defined as,

$$R^2 = 1 - \frac{\sum_i \sum_j (y_{ij} - \hat{y}_{ij})^2}{\sum_i \sum_j (y_{ij} - \bar{y})^2}$$

can be interpreted as the proportion of total variation of  $\mathbf{Y}$  that is accounted for by the fitted model. The second coefficient is expressed as follows:

$$r_c = 1 - \frac{\sum_i \sum_j (y_{ij} - \hat{y}_{ij})^2}{\sum_i \sum_j (y_{ij} - \bar{y})^2 + \sum_i \sum_j (\hat{y}_{ij} - \bar{y})^2 + N(\bar{y} - \bar{\bar{y}})^2}$$

This coefficient measures how close the points on a scatter plot of  $y_{ij}$  versus  $\hat{y}_{ij}$  are to the identity line. Its possible values extend from -1 to 1, with a perfect fit corresponding to a value of one and a lack of fit corresponding to values  $\leq 0$ .

The main advantage is that these criteria are independent of the estimating procedure, making it possible to compare non-nested models or models with different covariance structure or models estimated by the means of non-likelihood based methods.

### **Application to the EPIDOS study**

The French EPIDOS study is a prospective multicenter study of risk factors for hip fractures in 7575 women aged 75 years or older, included in 1992-1993 and recruited by mailing, based on large population-based listings including electoral rolls[20]. These women were asked to complete an annual postal questionnaire, investigating hospitalization, new health events, changes in weight, type of housing, ability to go outside, activities of daily life (ADL), instrumental activities of daily life (IADL), medication used, and subjective health evaluation. We initially planned to follow up the patients for four years, but the follow-up period was then extended to 7 years. The present analysis was carried out on the data from Montpellier (Southern France), one of the five participating centers, where 1548 women were recruited.

Some annual questionnaires were missing, mainly due to illness or to family events, such as the death of the spouse. These women sent in their responses late.

We investigated the risk factors related to changes in IADL score. This score is designed to evaluate independent living and deals with the following activities: performing light or heavy housework, laundry, shopping for groceries or personal items, preparing meals, using a telephone, taking public transport, managing money and taking medication without assistance. This indicator of disability varies from zero to 8 according to the number of activities that the woman is not capable of performing on her own. To ensure that each category contained sufficient numbers, we collapsed the responses as follows: 0, 1-2, 3-4, 5-6, 7-8. The final annual response therefore had five ordered categories and varied from 0 (not disabled) to 4 (almost entirely dependent).

The analysis was restricted to the women free of disability at inclusion and without missing values for the baseline covariables (929 women).

### **Model fitting with SAS NLMIXED**

The SAS NLMIXED procedure is very flexible, enabling the user to code the likelihood directly. This makes it possible to program all the models described above, as described in this section.

As the response reflects an underlying continuous phenomenon - the decline in the women's general state of health - we used the proportional odds model. As already explained, we began with the generalized ordered logit model (4). In the case of only two covariables (age and time) the SAS code is:

```
1      title1 ' Generalized ordered logit model' ;
2      title2 ' TIME, AGE' ;

3      proc nlmixed data=table qmax=5000 ;
4          bounds i1>0, i2>0, i3>0 ;
5          parms beta11=0.56 beta12=0.61 beta13=0.64 beta14=0.64
6              beta0=-10 beta24=0.3 beta23=0.3 beta22=0.3 beta21=0.3 ;

7          eta4 = beta0 + beta14*time + beta24*age+ u ;
8          eta3 = beta0 + i1 + beta13*time + beta23*age + u ;
```

```

9          eta2 = beta0 + i1 + i2 + beta12*time + beta22*age + u ;
10         eta1 = beta0 + i1 + i2 + i3 + beta11*time + beta21*age + u ;

11        if (iadl=4) then z=1/(1+exp(-eta4)) ;
12        else if (iadl=3) then z= 1/(1+exp(-eta3)) - 1/(1+exp(-eta4)) ;
13        else if (iadl=2) then z= 1/(1+exp(-eta2)) - 1/(1+exp(-eta3)) ;
14        else if (iadl=1) then z= 1/(1+exp(-eta1)) - 1/(1+exp(-eta2)) ;
15        else if (iadl=0) then z= 1 - 1/(1+exp(-eta1));

16        ll=log(z) ;
17        model iadl ~ general(ll);
18        random u ~ normal(0,su**2) subject=ident;

19        estimate 'int4' beta0 ;
20        estimate 'int3' beta0+i1 ;
21        estimate 'int2' beta0+i1+i2 ;
22        estimate 'int1' beta0+i1+i2+i3 ;

23        predict eta1 out=eta1;
24        predict eta2 out=eta2;
25        predict eta3 out=eta3;
26        predict eta4 out=eta4;
27        run;

```

In this code, "iadl" is the outcome variable (with values from 0 to 4) and "ident" is the variable that indicates the subject ID number (line 18). Lines 7 to 18 define the model according to equation (4). If there are more than two variables, eta1 to eta4 should be modified and more parameters, beta(i, j), should be used.

We included time since baseline as a continuous variable. We also included the following fixed effects, recorded at inclusion: age at entry (in years exceeding 74), gait speed (in m/s, for the completion of 6 m at a normal pace), pain (frequent pain in the lower back, hip, knee, ankle or leg), time to complete five chair stands (standing up and sitting down five times with arms crossed, in seconds) and visual acuity (corrected binocular distant acuity, decimal scale). All these covariables except "pain" were considered to be quantitative.

The NLMIXED procedure converged and provided estimations only when the initial parameters, given in the PARMS instruction (line 5), were close to the final solution. We then calculated successive models, beginning with a simple model with only a random intercept, introducing the covariables one by one and using the previously estimated parameters as the starting point for the next estimation at each step.

## Results

### Model selection

The parameters of the full generalized ordered logit model are provided in the first part of Table 1. We reduced the number of parameters by comparing this model with successive restricted models assuming proportional odds ratios for some covariables. We considered each covariable separately and if parameter estimates for adjacent response categories were equal or very similar (taking into account standard error) we replaced these parameters by a single parameter. For instance, the first three parameters of the variable "gait speed" can be replaced by a single parameter, but the fourth parameter is different (see Table 1). In this case, lines 7 to 10 of the SAS code, which define the linear predictors, are modified and a single parameter is used to replace the corresponding beta's (beta(i, 1), beta(i, 2), beta(i, 3)). At every step we used the likelihood ratio to test whether the new restricted model differed significantly from the initial full model. The final comparison is shown in Table 2. The p-

value associated with the likelihood ratio was 0.29 ( $\chi^2=13$  with 11 df), indicating that the two models have an equivalent goodness-of-fit. For the covariables age, gait speed and pain, proportionality could be validated for only some of the response levels. For instance the log-odds ratio associated with pain was 0.69 for the first two levels and 0.24 for the next two levels. Pain was found to have a more pronounced effect early in the loss of ability. In contrast, for the covariables "time to complete 5 chair stands", and "visual acuity", the model was reduced to only one parameter, indicating that the odds are entirely proportional.

The hierarchy of linear predictor estimates was checked for all subjects and all time points. The linear predictors are given by the PREDICT instruction in the SAS program and we checked if the inequality  $\eta_1 > \eta_2 > \eta_3 > \eta_4$  was verified for all the observations. None presented a reversed order, and the final model was therefore considered valid in the range of the observed values for the covariables.

The two models had identical coefficients of determination (0.58) and concordance correlation coefficients (0.77). We therefore retained the partial proportional logit model, in accordance with the principle of parsimony.

We also fitted a random-effect stereotype model (5) including the same covariables (SAS code given in the appendix). The parameter estimates are given in the last column of the Table 1. The reference category was an IADL score equal to zero. In this model each covariable is characterized by only one parameter. This model appeared to give a poorer fit than the previous models: the -2Log Likelihood was 10208, whereas that for the partial proportional odds model was 9769, the coefficient of determination was 0.32 and the concordance correlation coefficient was 0.57 (Table 2).

### **Goodness of fit**

The concordance correlation coefficients reported above were calculated taking into account both the estimates of the fixed parameters and the empirical Bayes estimates of the random effects. They can be seen as conditional model concordance correlation coefficients [19]. The average model concordance correlation coefficient can also be calculated, based entirely on the non-random part of the model. For the final partial proportional odds model, this average model concordance correlation coefficient was 0.41. This coefficient differed from the corresponding conditional model concordance correlation coefficient (equal to 0.77), indicating that the fixed covariables explained only a part of the response variation. A high level of between-individual heterogeneity is modeled by the random intercept. The probability of being at a given level of disability is therefore partially due to the uncharacterized "frailty" of the woman concerned.

## **Discussion**

The mixed models presented in this paper are interesting in two ways.

Firstly, they allow to analyze ordinal response and to evaluate the influence of each risk factor according to the levels of the response. Compared to the methods which dichotomize the response using an arbitrary cut-off and model it with a simple logistic regression our models are more worthwhile. We demonstrated in our example that the optimal cut-off changes with the risk factors. For example the optimal cut-off is an IADL score equal to two for the covariable "pain" and equal to one for the age or the gait speed.

Secondly, the random-effect models for non-gaussian responses are very powerful tools for the explanation of longitudinal processes. They can be calculated by means of standard statistical packages. The numerical procedure, which directly approximates the likelihood, does not suffer the drawbacks of previously developed methods. For example, the penalized

quasi-likelihood approach [21-22] has been shown to provide highly biased estimates of mixed-effects parameters with binary responses [23-24].

But, it is difficult to introduce a complex covariance structure as the computational time greatly increases with the number of random effects. However, this limitation is likely to become less severe in the future as computer capacity is continually increasing. In our example, the partial proportional logit model presented in table 1, took 56 minutes on a P4, 1.4GHz PC, and only 12 minutes on a P4, 2.5GHz PC.

Other more flexible methods allowing multiple levels of clustering, such as Markov chain Monte Carlo methods [25], can be used but are not based on likelihood statistics and model comparison is less straightforward. However, they can be used to cross-validate the estimations generated by procedures with numerical integration, such as NLMIXED.

Little work has been done on the model checking and the model diagnosis for GLMMs. This is nevertheless an important area in practice and we have presented here a simple method for the comparison of ordinal random-effect models and assessment of their goodness of fit.

Efforts are still required to improve the numerical approximations of likelihood in random-effect modeling, but these methods appear promising.

## **Conclusions**

The models presented in this paper allow to analyze longitudinal studies that are of particular interest in epidemiology and to take into account the ordinal nature of the outcome. Furthermore the impact of the risk factors can be modeled according to the response levels. This presents an important perspective in better explaining complex processes of evolution.

## References

1. Beckett L A, Brock D B, Lemke J H, et al. Analysis of change in self-reported physical function among older persons in four population studies. *Am J Epidemiol* 1996; 143(8): 766-778.
2. Diggle P J, Liang K Y, Zeger S L. *Analysis of Longitudinal Data*. Oxford: Clarendon Press, 1994
3. Lindsey J K. A review of some extensions to generalized linear models. *Stat Med* 1999; 18(17-18): 2223-2236.
4. Zeger S L, Liang K Y, Albert P S. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988; 44: 1049-1060.
5. Pinheiro J C, Bates D M. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J Comput Graph Stat* 1995; 4: 12-35.
6. Ananth C V, Kleinbaum D G. Regression models for ordinal responses: a review of methods and applications. *Int J Epidemiol* 1997; 26(6): 1323-1333.
7. Lall R, Campbell M J, Walters S J, Morgan K. A review of ordinal regression models applied on health-related quality of life assessments. *Stat Methods Med Res* 2002; 11(1): 49-67.
8. Mc Cullagh P. Regression models for ordinal data (with discussion). *J R Statist Soc B* 1980; 42: 109-142.
9. Mc Cullagh P, Nelder J A. *Generalized Linear Models*. 2nd ed. Boca Raton, London, New-York, Washington, D.C.: Chapman & Hall/CRC, 1989
10. Hedeker D, Mermelstein R J. A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research* 1998; 33(4): 455.
11. Peterson B L, Harrell F E. Partial proportional odds models for ordinal response variables. *Applied Statistics* 1990; 39: 205-217.
12. Engel J. Polytomous logistic regression. *Stat Neerlandica* 1988, 42:233-252.
13. Lunt M. Prediction of ordinal outcomes when the association between predictors and outcome differs between outcome levels. *Stat Med* 2005; in press.
14. Anderson J A. Regression and ordered categorical variables. *J R Statist Soc B* 1984; 46: 1-30.
15. Armstrong B G, Sloan M. Ordinal regression models for epidemiologic data. *Am J Epidemiol* 1989; 129(1): 191-204.
16. Greenland S. Alternative models for ordinal logistic regression. *Stat Med* 1994; 13(16): 1665-1677.
17. Carlin B P, Louis T A. *Bayes and empirical Bayes methods for data analysis*. 2nd ed. London: Chapman & Hall/CRC, 2000
18. Agresti A. *Categorical data analysis*. New York, Chichester, Brisbane, Toronto, Singapore: Wiley - Interscience Publication, 1990
19. Vonesh E F, Chinchilli V M, Pu K. Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics* 1996; 52(2): 572-587.



20. Dargent-Molina P, Favier F, Grandjean H, et al. Fall-related factors and risk of hip fracture: the EPIDOS prospective study. *Lancet* 1996; 348(9021): 145-149.
21. Breslow N E, Clayton D G. Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 1993; 88(421): 9-25.
22. Wolfinger R, O'Connell M. Generalized linear mixed models: a pseudo-likelihood approach. *J Statist Comput Simul* 1993; 48: 233-243.
23. Breslow N E. Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* 1995; 82(1): 81-91.
24. Rodriguez G, Goldman N. An assessment of estimation procedures for multilevel models with binary responses. *J R Statist Soc A* 1995; 158: 73-89.
25. Zeger S L, Karim M R. Generalized linear models with random effects: A Gibbs sampling approach. *J Am Stat Assoc* 1991; 86: 79-86.

## **Acknowledgements**

We would like to thank Dr François Favier and the EPIDOS group who provided the data for this work.

Table I. - Parameter estimates of three random-effect models for ordinal responses

	Generalized ordered logit model			Unconstrained partial proportional odds models			Stereotype model		
	Parameter	SE	p value	Parameter	SE	p value	Parameter	SE	p value
<b>Intercept</b>									
4	-6.89	1.08	< 0.0001	-6.93	0.83	< 0.0001	-5.02	0.83	< 0.0001
3	-5.51	0.93	< 0.0001	-6.04	0.80	< 0.0001	-4.55	0.76	< 0.0001
2	-4.82	0.83	< 0.0001	-4.82	0.78	< 0.0001	-3.40	0.70	< 0.0001
1	-2.05	0.77	0.008	-1.77	0.75	0.02	-1.71	0.46	0.02
<b>Time since inclusion (years)</b>									
4	0.58	0.05	< 0.0001	0.56	0.05	< 0.0001	0.80	0.04	< 0.0001
3	0.61	0.04	< 0.0001	0.61	0.04	< 0.0001			
2	0.65	0.03	< 0.0001	0.65	0.03	< 0.0001			
1	0.53	0.02	< 0.0001	0.52	0.02	< 0.0001			
<b>Age (in years minus 74)</b>									
4	0.22	0.04	< 0.0001	0.22	0.04	< 0.0001	0.31	0.03	< 0.0001
3	0.24	0.03	< 0.0001	0.24	0.03	< 0.0001			
2	0.24	0.03	< 0.0001	"	"	"			
1	0.18	0.03	< 0.0001	0.18	0.03	< 0.0001			
<b>Gait speed (m/s)</b>									
4	-3.32	0.81	< 0.0001	-2.92	0.59	< 0.0001	-4.10	0.63	< 0.0001
3	-3.32	0.68	< 0.0001	"	"	"			
2	-2.88	0.61	< 0.0001	"	"	"			
1	-2.36	0.56	< 0.0001	-2.57	0.56	< 0.0001			
<b>Pain</b>									
4	-0.0004	0.25	0.99	0.24	0.20	0.23	0.66	0.19	0.0005
3	0.26	0.21	0.23	"	"	"			
2	0.63	0.19	0.0008	0.69	0.17	< 0.0001			
1	0.72	0.17	< 0.0001	"	"	"			
<b>Time to complete 5 chair stands(sec)</b>									
4	0.63	0.18	0.0005	0.62	0.14	< 0.0001	0.59	0.15	0.0001
3	0.43	0.17	0.01	"	"	"			
2	0.61	0.15	< 0.0001	"	"	"			
1	0.68	0.15	< 0.0001	"	"	"			
<b>Visual acuity</b>									
4	-0.09	0.05	0.06	-0.14	0.03	< 0.0001	-0.16	0.04	< 0.0001
3	-0.11	0.04	0.008	"	"	"			
2	-0.14	0.04	0.0003	"	"	"			
1	-0.15	0.03	< 0.0001	"	"	"			
<b>Scale parameters</b>									
$\phi_4$							1		
$\phi_3$							0.91	0.05	< 0.0001
$\phi_2$							0.83	0.04	< 0.0001
$\phi_1$							0.55	0.03	< 0.0001
<b>Random effect SD</b>	<b>2.19</b>	<b>0.08</b>	<b>&lt; 0.0001</b>	<b>2.19</b>	<b>0.08</b>	<b>&lt; 0.0001</b>	<b>2.24</b>	<b>0.10</b>	<b>&lt; 0.0001</b>

Table II - Goodness of fit of three random-effect models for ordinal responses

	Generalized ordered logit model	Unconstrained partial proportional odds models	Stereotype model
-2LL	9756	9769	10208
AIC	9814	9805	10236
Coefficient of determination	0.580	0.582	0.32
Concordance correlation coefficient	0.770	0.770	0.57

## Appendix :

### SAS code for the stereotype model:

```
title1 ' Stereotype' ;
title2 ' AGE, TIME' ;

proc nlmixed data= table qmax=5000 ;

    bounds 0 < epsy1-epsy3 <= 0.99 ;

    parms alpha4=-8 alpha3=-8 alpha2=-6 alpha1=-4
    beta1=0.8 beta2=0.4
    epsy3=0.1 epsy2=0.1 epsy1=0.2 ;

    eta4 = alpha4 + beta1*time + beta2*age + u ;
    eta3 = alpha3 + (1-epsy3)*(beta1*time + beta2*age) + u ;
    eta2 = alpha2 + (1-epsy3-epsy2)*(beta1*time + beta2*age) + u ;
    eta1 = alpha1 + (1-epsy3-epsy2-epsy1)*(beta1*time + beta2*age)+u ;

    if (iadl=4) then
    z=exp(eta4) / (exp(eta1)+exp(eta2)+exp(eta3)+exp(eta4)+1) ;
    else if (iadl=3) then z=exp(eta3) / (exp(eta1)+exp(eta2)+
    exp(eta3)+exp(eta4)+1) ;
    else if (iadl=2) then z=exp(eta2) / (exp(eta1)+exp(eta2)+
    exp(eta3)+exp(eta4)+1) ;
    else if (iadl=1) then z=exp(eta1) / (exp(eta1)+exp(eta2)+
    exp(eta3)+exp(eta4)+1) ;
    else if (iadl=0) then z=1 / (exp(eta1)+exp(eta2)+
    exp(eta3)+exp(eta4)+1) ;

    ll= log(z) ;
    model iadl ~ general(ll) ;
    random u ~ normal(0,su**2) subject=ident;

    estimate 'phi3' (1-epsy3);
    estimate 'phi2' (1-epsy3-epsy2);
    estimate 'phi1' (1-epsy3-epsy2-epsy1);

    predict eta1 out=eta1;
    predict eta2 out=eta2;
    predict eta3 out=eta3;
    predict eta4 out=eta4;

run ;
```

## Annexe 2 : Notations

Sujets :  $i=1, \dots, n$

Covariables :  $X_{ijl}$ ,  $j^{\text{ème}}$  mesure de la covariable  $l$ , pour le sujet  $i$ , avec  $l=1, \dots, p$

Réponses longitudinales :  $Y_{ij}$ ,  $j^{\text{ème}}$  réponse du sujet  $i$  au temps  $t$ ,  $j=1, \dots, n_i$

Catégorie de la réponse ordinale :  $c=0, \dots, K$  ( $K=1$  pour réponse binaire)

## Annexe 3 : liste des abréviations

AGGIR : Autonomie Gérontologique – Groupes Iso-Ressources

AIC : Akaike Information Criterion

AIVQ : Activités Instrumentale de la Vie Quotidienne

APA : Allocation Personnalisée d'Autonomie

AVQ : Activités de la Vie Quotidienne

BIC : Bayesian Information Criterion

CIF : Classification Internationale du Fonctionnement

CRD : Completely Random Drop-out

GEE : Generalized Estimating Equations

GLM : Generalized Linear Model

GLMM : Generalized Linear Mixed Model

ID : Informative Drop-out

IMC : Indice de Masse Corporelle

LMM : Linear Mixed Model

MAR : Missing At Random

MCAR : Missing Completely At Random

MCMC : Markov Chain Monte Carlo

MNAR : Missing Not At Random

MSD : Minimum de Données Standardisées

OMS : Organisation Mondiale de la Santé

PSD : Prestation Spécifique Dépendance

RAI : Resident Assessment Instrument

RD : Random Drop-out

SC : Schwarz Criterion

SMAF : Système de Mesure de l'Autonomie Fonctionnelle

**RESUME : Comparaisons des méthodes d'analyse des données binaires ou ordinales corrélées. Application à l'étude longitudinale de l'incapacité des personnes âgées**

La modélisation de réponses binaires ou ordinales corrélées est un domaine de développement important en épidémiologie. L'étude longitudinale de l'incapacité des personnes âgées et la recherche des facteurs de risque de la vie en incapacité représente un enjeu crucial de santé publique. Dans ce contexte nous comparons les modèles logistiques marginaux et les modèles à effets aléatoires en prenant comme réponse l'incapacité considérée comme variable binaire afin d'illustrer les aspects suivants : choix de la structure de covariance, importance de données manquantes et des covariables dépendantes du temps, interprétation des résultats. Le modèle à effets aléatoires est utilisé pour construire un score prédictif de l'incapacité issu d'une large analyse des facteurs de risque disponibles dans la cohorte Epidos. Les modèles logistiques ordonnés mixtes sont ensuite décrits et comparés et nous montrons comment ils permettent la recherche d'effets différenciés des facteurs sur les stades d'incapacité.

**MOTS CLES :** Etudes longitudinales ; modèles logistiques ; équations d'estimation généralisées ; modèles à effets aléatoires ; évaluation de l'incapacité

**ABSTRACT: Comparison of statistical methods for correlated binary or ordinal data. Application to the longitudinal study of disability in elderly people.**

Fitting binary and ordinal correlated responses is presently an important field of development in epidemiology. The longitudinal study of disability in elderly people and the search for risk factors of disability is a crucial issue for public health. In this context, we compared marginal logistic models and random effects logistic models considering disability as a binary response to illustrate the following aspects: choice of the covariance structure, impact of missing data, importance of time-dependent covariates, and interpretation of the results. The random effects model was used to compute a predictive score from a large number of risk factors available in the Epidos cohort. The mixed ordinal logistic models were then described and compared. We showed how they allow identifying differential effects of the factors on the stages of disability.

**KEY WORDS:** longitudinal studies; logistic models; generalized estimating equations; random-effects models; disability evaluation

**LABORATOIRES :**

Unité INSERM U500, 39 avenue Charles Flahault, 34093 Montpellier

Equipe INSERM E361, 39 avenue Charles Flahault, BP 34493, 34093 Montpellier