



HAL
open science

Estimations et projections d'indicateurs de santé pour maladies chroniques et prise en compte de l'impact d'interventions

Mathilde Wanneveich

► To cite this version:

Mathilde Wanneveich. Estimations et projections d'indicateurs de santé pour maladies chroniques et prise en compte de l'impact d'interventions. Médecine humaine et pathologie. Université de Bordeaux, 2016. Français. NNT : 2016BORD0262 . tel-01432836

HAL Id: tel-01432836

<https://theses.hal.science/tel-01432836>

Submitted on 12 Jan 2017

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.

ÉCOLE DOCTORALE SP2
Société Politique, Santé Publique

T H È S E

pour obtenir le titre de

Docteur de l'Université de Bordeaux
Spécialité : BIostatistique

Présentée et soutenue par

Mathilde WANNEVEICH

le 04-11-2016

Estimations et projections
d'indicateurs de santé pour
maladies chroniques et prise en
compte de l'impact d'interventions

Thèse dirigée par Pierre JOLY

Jury :

<i>Président :</i>	--	-	-
<i>Rapporteur :</i>	M. Aurélien LATOUCHE	Professeur	Paris
<i>Rapporteur :</i>	M. Yohann FOUCHER	Maître de conférence	Nantes
<i>Examineur :</i>	Mme Hélène JACQMIN-GADDA	Directrice de Recherche	Bordeaux
<i>Examineur :</i>	M. Christophe BERGOUIGNAN	Professeur	Bordeaux
<i>Examineur :</i>	M. Olivier BOUAZIZ	Maître de conférence	Paris
<i>Directeur :</i>	M. Pierre JOLY	Maître de conférence	Bordeaux

Remerciements

Pendant ces trois ans de thèse, j'ai eu la chance d'être entourée par de nombreuses personnes qui m'ont permis, d'une manière ou d'une autre, d'avancer et d'arriver à mes fins.

La thèse est une véritable aventure, qui ne se parcourt pas en solitaire. La famille, les amis, les collègues, les chefs... tous sont indispensables et font partie d'un équilibre permettant d'aller au bout de l'aventure, de "bien-vivre" sa thèse, se sentir soutenu, entouré et donc de ne pas perdre le cap, et garder la motivation jusqu'à la fin du voyage.

C'est toutes ces personnes, ces gens formidables d'un point de vue professionnel et/ou personnel que je souhaite remercier.

Mes premiers remerciements vont à mon directeur de thèse, "mon chef", Pierre. Merci pour ces années à avancer avec moi sur ce travail, pour tout ce que j'ai pu apprendre à vos côtés, merci d'être toujours là pour discuter, de choses importantes...comme moins importantes! Je continuerai sûrement à vous vouvoyer, peut être par habitude, mais surtout en raison de tout le respect que j'ai pour vous, pour vos qualités de directeur de thèse, de chercheur,..., mais également pour votre grande bonté, votre gentillesse et pour toute l'attention que vous avez à mon égard.

Je tiens également à remercier Hélène Jacqmin-Gadda qui a suivi de près ces trois années. Vos compétences et vos conseils m'ont beaucoup apporté et m'ont permis d'avancer dans mes travaux. Vous êtes sollicitée de toute part, mais vous trouvez toujours un moment pour tous.

Merci pour votre expérience, votre disponibilité et vos qualités de chef d'équipe.

Je remercie aussi les autres membres du jury.

Aurélien Latouche et Yohann Foucher, les rapporteurs de cette thèse, qui m'ont fait l'honneur d'accepter de juger mon travail. Vos remarques et votre recul sur ce travail sont très constructifs pour moi. Merci également à Christophe Bergouignan et Olivier Bouaziz de prendre part à l'évaluation de cette thèse.

Un immense merci à tous ceux rencontrés à l'ISPED.

À Messieurs Daniel Commenges et Jean-François Dartigues, pour les réponses qu'ils ont apportées à mes questions.

À toute l'équipe Biostat (et SYSTM), pour son accueil, son dynamisme et la bonne ambiance, c'est un vrai plaisir de travailler et d'évoluer dans une telle équipe.

À l'ensemble de l'équipe pédagogique, que j'ai eu la chance de côtoyer pendant ces 3 années, pour m'avoir initiée à l'enseignement en me confiant des monitorats.

Puis, merci aux amis plus proches, je pense à tous les doctorants, ingénieurs, chercheurs, stagiaires... qui se reconnaîtront, les anciens comme les nouveaux.

La liste est longue alors ne m'en voulez pas de ne pas tous vous citer ici. J'ai néanmoins le besoin et l'envie de remercier "noir sur blanc" Émilie, Anaïs, Perrine,

Chloé, Henri, Loïc, Boris, Robin, Aïssatou, Audrey, Alexandra...avec qui j'ai partagé tellement de choses : discussions, congrès, pauses, sorties, tennis, soirées, vacances... ; sans oublier Viviane et Agnieszka qui me supportent au bureau 45 !
Bref, merci pour tous ces moments extras passés ensemble, dans ou hors le contexte du travail !

Évidemment, je remercie aussi tous mes amis non Ispédiens ! Là encore, il y en a tant auxquels je pense.

Merci à "Binouze" pour être toujours partant pour tout et n'importe quoi : les filles Cam, Cha et Lulu ; les gars Axou, Bryce, John, Onno et Pablal. Vous êtes essentiels à mon équilibre (ou plutôt déséquilibre !) et tous les moments avec vous sont inoubliables.

Merci aux amis plus anciens que je choisis de citer par villes : ceux de Sarlat, Périgueux, Paris, Bordeaux, Bayonne, La Rochelle, Toulouse, Limoges... Vous ne cessez jamais de croire en moi et, grâce à vous, j'ai toujours l'occasion de m'échapper les week-ends.

Merci aussi aux autres copains de l'INRIA, les Italiens comme les moins Italiens, et de l'école doctorale, je rigole toujours beaucoup à vos côtés !

Enfin, un Merci spécial et éternel à mes proches, ma super famille.

Je pense tout spécialement à mon papa, faire en sorte que tu puisses être fier de moi a toujours été une grande source de motivation, mais en fait, c'est de toi qu'il faut être fier. Ainsi qu'à mes frères et soeurs Stéphane, Aude, Benoît, Félix, Clémence, Marine et Lancelot (petite liste pour vous chef !) vous êtes ma précieuse fratrie et je sais que je peux compter sur chacun d'entre vous quoi qu'il arrive.

J'ai une pensée aussi pour ma grand-mère et mes beaux-frère/soeur, là en toutes circonstances.

Je finirai par un merci tout particulier, le plus beau des mercis que je gardais pour toi Andrea. Tu es rentré dans mon monde et a décidé d'y rester, depuis tu prends soin de moi au jour le jour et encore plus en cette fin de thèse "triple" pour nous ! Tu es essentiel pour moi.

Merci de ton soutien, merci pour tout ce que tu m'apportes et ce que tu partages avec moi, merci pour tout ce bonheur, merci d'être là... un semplice grazie, che per me, significa molto più di quanto tu possa immaginare...

Table des matières

Introduction	5
1 État de l'art des principaux modèles de projections	11
1.1 Approche en démographie	12
1.1.1 Modèle de Lee Carter	12
1.1.2 Modèles âge-période-cohorte	15
1.2 Les modèles multi-états	18
1.2.1 Contexte	18
1.2.2 Principaux modèles	18
1.2.3 Définitions, notions	20
1.2.4 Le modèle illness-death	22
2 Estimations avec un modèle illness-death et Indicateurs de santé	25
2.1 Notions générales sur l'analyse de données d'évènements	26
2.1.1 Censures et troncatures	26
2.1.2 Les paramètres du modèle	27
2.1.3 Méthodes d'estimations	27
2.2 Estimations	31
2.2.1 Probabilités de transition	31
2.2.2 Méthode du maximum de vraisemblance	33
2.3 Indicateurs de santé	34
2.3.1 Motivations épidémiologiques	34
2.3.2 Espérances de vie	36
2.3.3 Autres indicateurs	37
2.4 Intervalles de confiance pour les projections	37
3 Méthodes de projections avec un modèle illness death	39
3.1 Principales sources de données	40
3.1.1 Données de cohorte	40
3.1.2 Données Nationales démographiques	40
3.1.3 Autres sources de données possibles	40
3.2 Modèle Illness death et hypothèses Makoviennes	41
3.2.1 Modèles	41
3.2.2 Méthodologie pour l'estimation des paramètres	42
3.2.3 Équations différentielles	43
3.3 Modèle Illness death et hypothèses semi-Makoviennes	45
3.3.1 Modèles	45
3.3.2 Méthodologie pour l'estimation des paramètres	46
3.3.3 Équations différentielles	47
3.4 Projections	48

4	Étude de l'impact d'interventions	51
4.1	Introduction	52
4.2	Prise en compte de l'effet d'un facteur de risque	53
4.3	Schéma d'intervention	53
4.3.1	Description	53
4.3.2	Paramètres pour l'intervention	54
4.3.3	Scénarios d'intervention	55
4.3.4	Conséquences	56
4.4	Adaptation du modèle.	56
4.4.1	Estimations et projections	56
4.4.2	Indicateurs de santé	58
5	Application à la démence de type Alzheimer	65
5.1	Contexte et données	66
5.1.1	La démence : définition et état des lieux	66
5.1.2	Cohorte PAQUID	68
5.1.3	Données démographiques de l'INSEE	70
5.2	Modèles et hypothèses	71
5.2.1	Modèle Markovien	71
5.2.2	Modèle semi-Markovien	75
5.3	Intervention et scénarios	76
5.3.1	Facteurs de risque	76
5.3.2	Hypothèses	77
5.3.3	Scénarios d'intervention	77
5.4	Résultats	79
5.4.1	Estimations de l'incidence de la démence	79
5.4.2	Mortalité générale dans la population	80
5.4.3	Sous les modèles Markovien	81
5.4.4	Sous le modèle semi-Markovien	91
5.5	Choix du modèle	95
5.6	Conclusion sur l'application à la démence	97
5.6.1	Intervention	98
5.6.2	L'Hypothèse sur l'incidence de la démence	99
5.6.3	Les hypothèses sur la mortalité des déments et des non-déments	99
6	Application à la maladie de Parkinson	103
6.1	Contexte et données	104
6.1.1	Parkinson : définition et état des lieux	104
6.1.2	Données	105
6.2	Modèles et hypothèses	107
6.3	Adaptation de la méthode	109
6.3.1	Pour le modèle Markovien	110
6.3.2	Pour le modèle semi-Markovien	110
6.4	Résultats	112

6.4.1	Estimations avec le modèle Markovien	112
6.4.2	Estimations avec le modèle semi-Markovien	116
6.4.3	Projections	117
6.5	Conclusion sur le Parkinson	118
6.5.1	Incidence et Prévalence observées	119
6.5.2	Perspectives	120
7	Conclusions et perspectives	123
7.1	Résumé	124
7.2	Discussion	125
7.2.1	Modélisation d'une intervention	125
7.2.2	Développement d'un modèle semi-Markovien	127
7.2.3	Modèle Markovien ou semi-Markovien	128
7.3	Perspectives	129
7.3.1	Développements des modèles	129
7.3.2	Approche médico-économique	130
7.3.3	Généralisation et utilisation	131
A	Annexes	133
A.1	Contributions à la vraisemblance pour un modèle illness-death	134
A.2	Résolution des équations différentielles	136
A.2.1	Modèle Markovien M2 : méthode numérique, Runge-Kutta d'ordre 4	136
A.2.2	Modèle semi-Markovien S2	137
A.3	Résultats supplémentaires sur l'application à la démence	139
A.3.1	Mortalités estimées sur PAQUID	139
A.3.2	Impacts de l'intervention pour les hommes (modèle M2)	139
A.3.3	Comparaison des projections de prévalence : notre interven- tion <i>vs</i> l'intervention de Jacqmin <i>et al.</i> (modèle M2)	142
A.3.4	Autres résultats pour le modèle S2	144
A.4	Résultats supplémentaires sur l'application à la maladie de Parkinson	145
A.4.1	Estimation de l'incidence du Parkinson sur les données de la DCIR	145
A.4.2	Estimation du risque relatif, modèle M2	148
	Bibliographie	151

Listes des abréviations et des notations

Abréviations

OMS	organisation mondiale de la santé
SVD	méthode de la décomposition en valeurs singulières
ARIMA	modèle "AutoRegressive Integrated Moving Average"
APC	modèle âge-période-cohorte
GLM	modèle linéaire généralisé
VIH	virus de l'immunodéficience humaine
MLE	estimateur du maximum de vraisemblance
INSEE	institut national de la statistique et des études économiques
INED	institut national d'études démographiques
IC	intervalle de confiance
APA	american psychiatric association
DSM IV	manuel diagnostique et statistique des troubles mentaux 4 ^{ème} édition
DSM-III-R	manuel diagnostique et statistique des troubles mentaux 3 ^{ème} édition révisée
ICM-10	classification internationale des maladies 10 ^{ème} version
PAQUID	cohorte personnes âgées QUID
HBP	hypertension artérielle
ApoE4	Apolipoprotéine E4
INSERM	institut national de la santé et de la recherche médicale
INVS	institut de veille sanitaire
SNIIRAM	système national d'information inter-régimes de l'Assurance maladie
DCIR	données de consommations inter-régimes
ALD	affection de longue durée
PMSI	programme de médicalisation des systèmes d'information
PIB	produit intérieur brut
QALY	"quality adjusted life years"
3C	étude des 3 Cités
ATC	classification anatomique, thérapeutique et chimique des médicaments
<i>vs</i>	<i>versus</i>
<i>cf.</i>	<i>confer</i>

Principales notations

Analyse de données d'évènement

$f(\cdot)$	fonction de densité
$S(\cdot)$	fonction de survie
$\lambda(\cdot)$	fonction de risque
$\Lambda(\cdot)$	fonction de risque cumulé
L	vraisemblance
C	temps de censure à droite
T	temps de décès

Modèle illness-death

$\alpha_{ij}(\cdot)$	intensité de transition de l'état i à j
$A_{ij}(\cdot)$	intensité de transition cumulée de l'état i à j
$P_{ij}(\cdot)$	probabilité de transition de l'état i à j
$P_{02}^1(\cdot)$	probabilité de transition de l'état 0 à l'état 2 en étant passé par l'état 1
$P_v(\cdot)$	probabilité d'être vivant
$g(\cdot)$	risque relatif de décès des malades par rapport aux non-malades
$\alpha^d(\cdot)$	sur-risque de décès des malades par rapport aux non-malades
$\alpha_2(\cdot)$	mortalité générale dans la population
$\nu(\cdot)$	taille de la population à risque de développer la maladie
a_0	âge à partir duquel les sujets sont à risque de développer la maladie

Indicateur de santé

$LE_{00}(\cdot)$	espérance de vie sans la maladie
$LE_{11}(\cdot)$	espérance de vie d'un malade
$LE_0(\cdot)$	espérance de vie d'un non-malade
$LE_{..}(\cdot)$	espérance de vie générale
$T_{11}(\cdot)$	nombre moyen d'années passées avec la maladie
$F_{01}(\cdot)$	probabilité de développer la maladie
$AV_{01}(\cdot)$	âge moyen de début de la maladie
$Prev(\cdot)$	prévalence de la maladie

Intervention et facteur de risque

Z	variable d'exposition à un facteur de risque
$\alpha_{ij}^0(\cdot)$	intensité de transition de base de l'état i à j
θ_{ij}	risque relatif associé au facteur de risque pour la transition de l'état i à j
$p(\cdot)$	prévalence du facteur de risque
$p_0(\cdot)$	prévalence du facteur de risque chez les non-malades
$p_1(\cdot)$	prévalence du facteur de risque chez les malades
$\omega(\cdot)$	proportion d'exposés au facteur de risque
τ	année de l'intervention
a_τ	âge au moment de l'intervention

Valorisations Scientifiques

Publications

- Wanneveich M, Jacqmin-Gadda H, Dartigues J F, Joly P. Impact of intervention targeting risk factors on chronic disease burden. *Statistical Methods in Medical Research*, (2016).
- Wanneveich M, Jacqmin-Gadda H, Dartigues J F, Joly P. Projections of health indicators for chronic disease under Semi-Markov assumption. Under review, *Population Health Metrics*, (2016).

Communications orales

- Wanneveich M, Jacqmin-Gadda H, Helmer C, Joly P. Impact d'une intervention de Santé Publique ciblant un facteur de risque de la démence sur les projections de prévalence et les espérances de vie. 46^{èmes} *Journées de Statistique*, juin 2014, Rennes (France).
- Wanneveich M, Jacqmin-Gadda H, Helmer C, Joly P. Impact of preventive policy regarding dementia's risk factors, on prevalence projections and life expectancy. 27th *International Biometrics Conference*, juillet 2014, Florence (Italie).
- Wanneveich M, Jacqmin-Gadda H, Helmer C, Joly P. Impact of preventive policy regarding dementia's risk factors, on several health indicator projections. *GDR Statistique et Santé*, Octobre 2014, Toulouse (France).
- Wanneveich M, Jacqmin-Gadda H, Dartigues JF, Joly P. Impact of intervention targeting risk factors on chronic disease burden. *International Biometrics Society Channel*, Avril 2015, Nimègue (Pays-Bas).
- Wanneveich M, Jacqmin-Gadda H, Joly P. Projections d'indicateurs de santé pour maladies chroniques ; modèle multi-états et hypothèse semi-Markovienne. 6^{èmes} *Rencontres Jeunes Statisticiens*, Août 2015, Le Teich (France).
- Wanneveich M, Jacqmin-Gadda H, Dartigues JF, Joly P. Projections of health indicators for chronic disease : illness death model and semi-Markov assumption. 48^{èmes} *Journées de Statistique*, mai 2016, Montpellier (France).
- Wanneveich M, Jacqmin-Gadda H, Dartigues JF, Joly P. Projections of health indicators for chronic disease : illness death model and semi-Markov assumption. 31^{èmes} *International Workshop on Statistical Modelling*, juillet 2016, Rennes (France).
- Joly P, Wanneveich M, Dartigues JF, Jacqmin-Gadda H. Projections of health indicators for chronic disease under Semi-Markov assumption. 28th *International Biometrics Conference*, juillet 2016, Victoria (Canada).

Communications affichées

- Wanneveich M, Jacqmin-Gadda H, Dartigues JF, Joly P. Projections of health indicators for chronic diseases. Application to Dementia. n illness-death models. *Population Association of America Annual Meeting*, mars 2016, Washington DC(États-Unis).
- Joly P, Jacqmin-Gadda H, Dartigues JF, Wanneveich M. Impact of Public Health intervention targeting risk factors, on chronic disease burden. *Population Association of America Annual Meeting*, mars 2016, Washington DC(États-Unis).

Introduction

De nos jours, nous faisons le constat dans de nombreux pays, d'un allongement de l'espérance de vie générale de la population qui s'accompagne d'une augmentation de la prévalence (nombre de cas de personnes atteintes d'une maladie à un moment donné) de maladies liées à la vieillesse. Or, ceci devient un véritable problème de Santé Publique d'une part, car ces maladies sont généralement incurables et détériorent progressivement la santé des malades, les amenant à nécessiter des soins spécifiques ; d'autre part parce que le vieillissement de la population n'est pas prêt de s'arrêter notamment du fait des progrès de la médecine et de l'amélioration des conditions de vie. C'est pourquoi, le fardeau économique social et démographique de ces maladies devraient s'alourdir dans les années à venir.

Il devient donc primordial d'envisager l'évolution de ces maladies, pour permettre aux États d'anticiper et de trouver des solutions notamment pour la prise en charge de ces malades. Ce sujet, très d'actualité, intéresse à la fois les actuaires (spécialistes de l'application du calcul des probabilités et de la statistique aux questions d'assurances, de finances et de prévoyance sociale), les démographes, les bio-statisticiens, ainsi que les acteurs de la Santé Publique au sens large, tels que les épidémiologistes, les décideurs de la santé et les économistes de la santé.

Maladies chroniques liées au vieillissement

D'après l'Organisation Mondiale de la Santé (OMS), nous parlons de maladies chroniques dans le cas d'une affection de longue durée (ALD) qui évolue (généralement lentement) avec le temps. Ces maladies durent de nombreuses années, voire le plus souvent toute la vie de la personne atteinte, car la majorité d'entre elles sont incurables. De plus, elles entraînent souvent une invalidité et des complications graves. C'est pourquoi, les conséquences d'une maladie chronique sur la vie quotidienne des malades et de leur entourage sont lourdes à la fois d'un point de vue sociale, psychologique et économique.

Ils existent une multitude de maladies chroniques, nous pouvons citer en exemple :

- des maladies comme le diabète, l'asthme, les maladies cardio-vasculaires, ...
- des maladies transmissibles persistantes, comme le sida ou l'hépatite C,
- des maladies très handicapantes comme les scléroses en plaques, les maladies neuro-dégénératives,...
- des troubles mentaux tels que la dépression, la schizophrénie,...
- et d'autres maladies rares comme la mucoviscidose, la drépanocytose et les myopathies.

Le travail de cette thèse cible des maladies neuro-dégénératives ayant aussi un lien avec le vieillissement de la population et pour lesquelles aucun retour à l'état sain n'est envisageable au jour d'aujourd'hui. Dans cette thèse, nous nous focalisons donc principalement, sur les démences de type Alzheimer et la maladie de Parkinson. Ces maladies se développent tardivement, mais sont très fréquentes chez les personnes âgées, l'Alzheimer représente 2/3 des cas de démence et est la maladie neuro-dégénérative la plus développée suivie de la maladie de Parkinson. Aucune baisse de leurs prévalences n'est envisagée pour l'instant, au contraire, la situation démographique actuelle suggère, par exemple pour la démence, que le nombre de cas va augmenter proportionnellement au nombre de personnes âgées [Berr 2007]. Or, ces maladies nécessitent des soins médicaux spécifiques ayant un coût humain et économique conséquent.

Il est donc important de pouvoir estimer le nombre de cas attendus ainsi que la durée de vie moyenne des malades dans les années à venir afin de pouvoir planifier les ressources qui seront nécessaires.

Modèles et Projections

Les projections de population cherchent à déterminer l'évolution à venir de la population, en fournissant une estimation de données futures, à partir de la connaissance du passé et d'hypothèses pour le futur.

Ainsi, afin de prévoir quel sera le fardeau/poids futur de ces maladies dans nos sociétés, des modèles de projections qui prennent en compte du mieux possible l'évolution de la maladie, d'un point de vue épidémiologique, et l'évolution de la population doivent être développés.

La modélisation de la mortalité des populations a toujours fait l'objet de nombreuses publications, mais dernièrement, du fait de l'augmentation continue de l'espérance de vie, de nouveaux modèles ont vu le jour pour étudier l'évolution des taux de mortalité futurs, et des modèles de projection ont été proposés. Ces modèles sont, pour certains, basés sur l'évolution des taux de mortalité en fonction de l'âge, et pour d'autres plus élaborés, basés sur l'évolution des taux de mortalité en fonction de l'âge, mais aussi du temps calendaire.

Cependant, il ne s'agit pas ici de simplement étudier l'évolution générale de la population pour en donner des projections, mais de prendre en compte une maladie, son évolution et ses effets sur des indicateurs de santé, tout ceci dans le contexte démographique actuel et présumé pour les années à venir, afin de faire une évaluation du fardeau futur de cette maladie.

D'autres travaux se sont déjà penchés sur la problématique, ont proposé des modèles et ont donné des projections. Nous retrouvons principalement des travaux menant à des projections de la prévalence future de la démence, nous pouvons citer par exemple : Brookmeyer *et al.*, Hebert *et al.* et Joly *et al.*, [Brookmeyer 1998], [Hebert 2003], [Joly 2013].

Aujourd'hui, les modèles multi-états semblent très bien établis pour modéliser cette maladie [Joly 2002], et les méthodes de macro-simulation sont très souvent utilisées pour fournir des projections. Néanmoins, en s'attardant sur les différents modèles de projections proposés, nous constatons que bien qu'ils aient une base commune, les stratégies/hypothèses sont différentes. Ainsi, en fonction du modèle utilisé, nous n'aurions pas les mêmes projections pour une même population. Il est donc important d'établir un modèle adapté à la maladie dont nous voulons fournir des projections et également flexible dans ses hypothèses pour permettre de prendre en compte l'évolution et l'épidémiologie de la maladie ainsi que l'évolution de la population générale, afin de donner des projections du fardeau futur de la maladie.

Néanmoins, cela ne suffit pas de donner des projections en considérant que l'évolution de la maladie ou l'évolution générale de la population va suivre son cours actuel, sans connaître des changements importants entre-temps. En effet, la question se pose surtout sur l'évolution de la maladie du fait des progrès continus de la médecine. Il n'est pas dénué d'intérêt de penser que des changements du mode de vie, un traitement ou autres, entraînant une évolution de l'incidence de la maladie (nombre de nouveaux cas de malades par an, dans une population), pourraient avoir un impact sur le fardeau à venir de la maladie. Pour des maladies telles que l'Alzheimer, il est assez approprié d'envisager des interventions de Santé Publique ciblant des facteurs de risque de la maladie, ceci fait même déjà l'objet de quelques travaux. Cependant, avant de décider de mettre en place une intervention, il est important d'évaluer avec attention quel pourrait être son impact sur les projections futures du fardeau de la maladie, ce qui n'est pas chose facile de par les relations complexes qui existent entre les facteurs de risque, les maladies et la mortalité.

Actuellement, des approches ont été étudiées [Polidori 2012] et quelques études ont investigué l'impact d'une intervention sur les projections [Brookmeyer 2007], [Barnes 2011], [Jacqmin-Gadda 2013] mais, là encore, le type d'intervention, la façon de la considérer dans le modèle ainsi que les hypothèses fixées varient grandement en fonction des études.

C'est pourquoi, il est fort intéressant d'approfondir ce sujet qui semble encore en phase exploratoire et qui pourtant est d'un intérêt grandissant compte tenu de la situation actuelle et aurait des retentissements importants pour la Santé Publique. Il devient donc utile d'aboutir à une approche générale, un modèle souple dans ces hypothèses, avec un large "champ d'action" pour considérer d'éventuelles modifications à venir (comme des interventions), afin d'avoir les outils nécessaires pour répondre aux nombreuses questions sur cette problématique.

Objectifs de la thèse et structure du rapport

C'est dans ce contexte que sont nés les objectifs de cette thèse. En se basant sur les travaux entamés par l'équipe Biostatistique¹ de Bordeaux, nous avons fixé plusieurs objectifs.

Le principal objectif consistait à développer l'utilisation du modèle multi-états illness-death proposé par l'équipe pour faire des projections.

Pour cela, nous avons plusieurs sous-objectifs à atteindre :

- tout d'abord, assouplir/relâcher les hypothèses du modèle, principalement en permettant à la méthode de prendre en compte une évolution de l'incidence de la maladie au cours du temps, mais également de considérer une évolution de la mortalité des malades en fonction de la durée passée avec la maladie.
- puis, adapter la méthode à l'évaluation de l'impact d'une intervention beaucoup plus générale que celle qui avait été mise en place initialement.
- ensuite, faire des projections sur plusieurs indicateurs de santé pertinents pour permettre un meilleur diagnostic du fardeau futur de la maladie, ainsi que proposer des projections avec ou sans intervention.

Dans ce contexte, nous développerons les indicateurs de santé suivants :

- les espérances de vie dans différents états de santé : générale, sans la maladie, pour un non-malade et pour un malade,
- le nombre moyen d'années passées avec la maladie,
- la probabilité de développer la maladie,
- l'âge moyen de début de la maladie,
- et la prévalence de la maladie.
- enfin, appliquer cette approche à d'autres pathologies chroniques pour en vérifier l'adaptabilité.

Tout l'intérêt d'utiliser un tel modèle en agissant directement à la base des estimations, plutôt que de faire des projections à partir des taux actuels, est de pouvoir prendre en compte le plus possible de variabilité dans les paramètres susceptibles d'intervenir dans les projections et de pouvoir facilement les modifier.

Ce rapport est structuré en 7 chapitres.

Le chapitre **1** retrace les principaux modèles de projections qui sont utilisés, pour finir par les modèles multi-états et plus spécialement le modèle illness-death que nous utilisons pour faire des projections.

Dans le chapitre **2**, nous abordons quelques notions générales utiles à la compréhension du modèle illness-death, ainsi que les méthodes d'estimations et les quantités d'intérêts estimées avec le modèle, pour enfin présenter les différents indicateurs de santé que nous calculerons par la suite.

Le chapitre **3** présente la méthode, c'est-à-dire toutes les étapes qui nous permettent de faire des projections à partir de ce modèle. Nous présentons également

1. INSERM U 1219, Université de Bordeaux, ISPED

dans cette partie les modèles que nous utiliserons avec leurs hypothèses.

Le chapitre 4 concerne également la méthode, mais il s'agit ici de développer la façon dont nous prenons en compte l'impact d'une intervention dans la méthode de projections, ainsi que les conséquences que cela implique à la fois sur l'estimation des paramètres/fonctions, les projections et le calcul des indicateurs de santé.

Dans les chapitres 5 et 6, nous développons deux applications, l'une sur la démence de type Alzheimer et l'autre sur la maladie de Parkinson. L'application à la démence a été la base du développement du modèle, de l'approche présentée dans cette thèse et de la réalisation de la plupart des objectifs préfixés, alors que l'application à la maladie de Parkinson a davantage servi à vérifier la transposabilité de la méthode et à adapter le modèle à d'autres situations.

Pour finir, dans le chapitre 7, nous discutons du travail effectué, les avantages, les limites, les différents choix que nous avons fait et pourquoi... Puis nous parlons des perspectives, de prochains travaux qu'il serait fort intéressant de développer sur ce modèle de projections.

Enfin, les Annexes seront appelées tout au long de cette thèse pour apporter des développements supplémentaires ainsi que d'autres résultats.

État de l'art des principaux modèles de projections

Sommaire

1.1	Approche en démographie	12
1.1.1	Modèle de Lee Carter	12
1.1.2	Modèles âge-période-cohorte	15
1.2	Les modèles multi-états	18
1.2.1	Contexte	18
1.2.2	Principaux modèles	18
1.2.3	Définitions, notions	20
1.2.4	Le modèle illness-death	22

Il existe différents modèles permettant de donner des projections de population. En principe, ces modèles s'appuient sur des données passées afin de modéliser une tendance pour ensuite extrapoler et faire des projections pour les années futures.

Dans ce chapitre, nous présentons quelques modèles parmi les plus étudiés sans en faire une liste exhaustive ni détailler des extensions ou alternatives existantes. L'objectif est de connaître le cadre général et les différentes méthodes d'usages pour illustrer les divers points de vue sur ce sujet.

Ainsi, les deux premiers modèles relèvent plutôt d'une approche démographique, ils servent habituellement à donner des projections de mortalité. Ces deux modèles font partie des modèles stochastiques de mortalité rendant compte des variations des taux instantanés de mortalité autour de leurs valeurs tendanciennes, à chaque âge.

Tandis que le dernier est un modèle biostatistique de plus en plus utilisé en projections de population, car il permet d'introduire la notion de maladie et de varier les projections possibles. C'est d'ailleurs sur ce modèle qu'a porté le travail de cette thèse.

1.1 Approche en démographie

1.1.1 Modèle de Lee Carter

1.1.1.1 Contexte

En 1992, Ronald Lee et Lawrence Carter ont développé un modèle d'extrapolation de la fonction de mortalité qui dépend de l'âge et du temps (que nous pouvons représenter par une surface) dont l'estimation des paramètres s'effectue à partir des données historiques. Le modèle de Lee et Carter a été appliqué sur plusieurs populations dont les États-Unis d'Amérique [Lee 1992], le Chili [Lee 1994], le Canada, le Japon, la Belgique, ..., pour permettre par la suite d'extrapoler leurs taux de mortalité.

La méthode consiste à modéliser les taux de mortalité, puis faire des projections en considérant les effets « âge » et « période » observés sur les données. Cette méthode présente une composante propre à l'âge et une autre tendancielle. L'effet de l'âge est supposé constant au cours du temps, et l'effet de la période permet de prendre en compte l'évolution, notamment des comportements démographiques, au cours du temps. Ainsi, la distribution des taux de mortalité est supposée déterministe selon les âges et stochastique selon les années.

Néanmoins, cette approche présente des inconvénients, car elle n'incorpore pas l'avis d'expert sur le progrès de la médecine, sur l'évolution de la mortalité, l'apparition des nouvelles maladies voir même l'évolution du style de vie.

1.1.1.2 Modèle

Le modèle résulte de la combinaison de méthodes statistiques pour séries chronologiques et d'une approche paramétrique.

Dans ce modèle, l'estimateur du taux instantané de mortalité $\hat{\mu}$ est fonction de composantes qui dépendent de l'âge x et du temps t , auxquelles des contraintes sont imposées pour permettre l'identifiabilité. De plus, le modèle inclut une hypothèse d'homoscédasticité des erreurs, ε_{xt} , considérées comme bruit blanc gaussien ($N(0, \sigma^2)$).

L'écriture standard du modèle est la suivante [Lee 1992], [Brouhns 2002a] :

$$\ln(\hat{\mu}_x(t)) = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t} \quad (1.1)$$

sous les contraintes :

$$\sum_{x=x_{min}}^{x_{max}} \beta_x = 1$$

$$\sum_{t=t_{min}}^{t_{max}} \kappa_t = 0$$

Le taux instantané de mortalité est donc la somme d'une composante spécifique à l'âge α_x s'interprétant comme la valeur moyenne de $\ln(\hat{\mu}_x(t))$ au cours du temps

(pour chaque âge x) illustrant le comportement moyen des taux instantanés de mortalité relativement à l'âge x ; et d'une composante égale au produit de deux paramètres κ_t et β_x , le premier décrivant l'évolution générale de la mortalité au cours du temps et le second, propre à l'âge, décrivant la sensibilité du taux instantané de mortalité à l'âge x par rapport à l'évolution générale κ_t . Ainsi β mesure pour chaque âge x sa sensibilité au temps calendaire t , mais Lee et Carter supposent que cette sensibilité pour un âge ne varie pas en fonction de t .

Pour mieux comprendre :

- Pour tout âge x les variations relatives des taux de mortalité à des temps t différents ne dépendent pas de x . En d'autre terme, si la variation relative du taux de mortalité à 50 ans en 2000 par rapport à 1990 était de 80 %, ce coefficient reste le même pour tous les âges ;
- A un temps t fixé les variations relatives des taux de mortalité à des âges différents ne dépendent pas de t . En d'autre terme, si en 2000 la variation relative du taux de mortalité à 20 ans par rapport à 50 ans était de 50 %, ce coefficient s'appliquera à toutes dates ultérieures.

1.1.1.3 Estimation des paramètres

Lee et Carter ont proposé la méthode des moindres carrés ordinaires pour estimer les paramètres du modèle, il s'agit alors de résoudre :

$$(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t) = \underset{\alpha_x, \beta_x, \kappa_t}{\operatorname{argmin}} \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} (\ln(\hat{\mu}_x(t)) - \alpha_x - \beta_x \kappa_t)^2 \quad (1.2)$$

En d'autre terme, il faut trouver les valeurs de $\hat{\alpha}$, $\hat{\beta}$ et $\hat{\kappa}$ qui minimisent l'expression

$$\sum_{x,t} (\ln(\hat{\mu}_x(t)) - \alpha_x - \beta_x \kappa_t)^2$$

Dans cette expression, nous notons x_{\max} l'âge maximum, x_{\min} l'âge minimum, t_{\max} la date la plus récente et t_{\min} la plus ancienne, il y a alors $(2 * (x_{\max} - x_{\min} + 1) + t_{\max} - t_{\min} + 1)$ paramètres à estimer.

α_x est relativement simple à estimer, il suffit d'annuler la dérivée première de l'équation précédente, ainsi :

$$\hat{\alpha}_x = \frac{1}{t_{\max} - t_{\min} + 1} \sum_{t=t_{\min}}^{t_{\max}} \ln(\hat{\mu}_x(t)) \quad (1.3)$$

Puis, pour estimer β_x et κ_t et obtenir la solution unique et optimale, ils utilisent la méthode de la décomposition en valeurs singulières (Singular Value Decomposition (SVD)) de la matrice Z , de terme général :

$$Z_{xt} = \ln(\hat{\mu}_x(t)) - \hat{\alpha}_x$$

Z correspond à la matrice des taux centrés de dimension $(x_{max} + x_{min} + 1; t_{max} + t_{min} + 1)$ que nous pouvons approcher par l'écriture $Z = \hat{\beta}\hat{\kappa}$.

Remarque : si Z n'est pas une matrice rectangulaire, la méthode alternative est d'utiliser l'algorithme itératif de Goodman.

En utilisant la méthode SVD, nous déterminons u_1 (respectivement v_1) le vecteur propre normalisé de $Z^T Z$ (respectivement de $Z Z^T$) correspondant à la valeur propre la plus élevée λ_1 . Typiquement, la loi de mortalité $Z \approx \sqrt{\lambda_1} v_1 u_1^T$ explique environ 90 % de la variance de $\ln(\mu_x(t))$. Il suffit donc de prendre :

$$\hat{\beta} = \frac{v_1}{\sum_j v_{1j}} \quad (1.4)$$

et

$$\hat{\kappa} = \sqrt{\lambda_1} \sum_j v_{1j} u_1 \quad (1.5)$$

1.1.1.4 Projections

Faire des projections consiste à conjecturer l'évolution d'un phénomène dans une période ultérieure donnée. Pour construire des prévisions sur les taux de mortalité des années à venir avec ce modèle, il s'agit d'extrapoler la tendance temporelle de mortalité $\hat{\kappa}_t$ dans le futur *via* l'utilisation des techniques de prévisions des séries chronologiques.

Une fois les paramètres estimés, l'objectif est d'utiliser les résultats pour extrapoler les taux de mortalité pour $t > t_{max}$. Pour cela, la série des $\hat{\kappa}_t$ est analysée afin de lui ajuster un modèle de type ARIMA (AutoRegressive Integrated Moving Average).

Il s'agit d'une méthode de projection pour les séries temporelles longues, développée par Box et Jenkins en 1976 [Box 1976]. Cette méthode considère trois types de paramètres : p est le nombre de termes autorégressifs (AR), d est le nombre de différenciations effectuées pour rendre le modèle stationnaire et q est le nombre de moyennes mobiles (MA), c'est pourquoi elle se note ARIMA(p,d,q). Nous ne la détaillerons pas davantage.

Enfin, bien que ce modèle soit essentiellement utilisé pour des projections de taux de mortalité, récemment, des études proposent d'adapter la méthode pour donner des projections d'espérance de vie en santé, par exemple Majer *et al.* [Majer 2013].

1.1.1.5 Limites

Malgré le succès de cette méthodologie, plusieurs critiques ont été adressées à ce modèle. Notamment Gutterman et Vanderhoof [Gutterman 1999] ont critiqué l'approche purement extrapolative de cette méthode. En effet, seule l'histoire passée

peut être introduite dans le modèle. Le modèle ne peut prendre en compte les nouveaux progrès de la médecine, la découverte d'un nouveau traitement,..., ou toutes variations subites de la mortalité, ce qui biaise les projections de mortalité au-delà de la période d'étude. De plus, ce modèle repose sur un seul processus dynamique, c'est à dire que seule une composante du modèle est utilisée pour décrire l'évolution temporelle de la mortalité. Cela suppose donc une corrélation parfaite de l'évolution temporelle de la mortalité pour tous les âges. Une autre limite vient de l'utilisation de la méthode des moindres carrés ordinaires pour estimer les paramètres [Alho 2000] car celle-ci nécessite que les résidus vérifient l'hypothèse d'homoscédasticité, ce qui est rarement vérifiée.

Pour pallier ces limites, des extensions ou alternatives ont été développées. Nous pouvons citer le modèle log-Poisson, proposé par Brouhns et al. [Brouhns 2002b] et les extensions élaborées par Renshaw et Haberman [Renshaw 2003], [Renshaw 2008].

Cependant, le modèle de Lee-Carter reste le modèle le plus utilisé pour l'extrapolation de la fonction de mortalité, car sa mise en oeuvre est relativement simple.

1.1.2 Modèles âge-période-cohorte

1.1.2.1 Contexte

Les modèles âge-période-cohorte (APC) sont largement utilisés par les épidémiologistes pour analyser les tendances de l'incidence de certaines maladies [Clayton 1987], [Kupper 1985]). En effet, le taux de mortalité de l'âge a à l'instant t dépend de l'action conjointe de l'histoire passée de la génération $t - a$, de l'année de calendrier t et de l'âge a . L'étude simultanée de ces trois dimensions fait apparaître des effets période et génération : l'effet de période, lié au temps calendaire t , exprime l'influence rapide des événements qui déterminent donc des variations à court terme; l'effet de génération, lié à l'histoire passée, exprime les transformations lentes et profondes qui affectent les personnes nées une même année dans les mêmes contextes [Clayton 1987], [Decarli 1987], [Osmond 1985], [Verdecchia 1989]. Ce dernier, n'était pas considéré dans le modèle de Lee Carter, or il doit être pris en compte dans la projection de la mortalité.

La figure 1.1 résume par un diagramme de Lexis, les différentes représentations que nous obtenons en fonction de ces trois paramètres de temps.

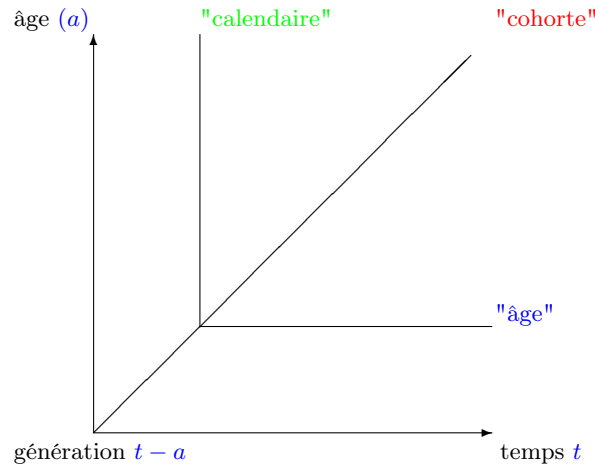


FIGURE 1.1 – Diagramme de Lexis

Sur le diagramme de Lexis, les axes horizontaux correspondent aux âges exacts entiers, les axes verticaux sont les années, et les obliques sont les générations. Pour repérer un événement sur ce diagramme, il faut en connaître les coordonnées en termes de date et d'âge.

Ici, si nous supposons que l'événement observé est la mortalité, nous pouvons donc la déterminer de trois manières différentes :

- soit la mortalité "calendaire", pour un temps t , quel que soit l'âge,
- soit la mortalité par "âge" a , quel que soit le temps calendaire,
- soit la mortalité d'une génération (ou "cohorte") $t - a$.

1.1.2.2 Modèle

De façon très générale, nous pouvons exprimer un risque de mortalité μ comme une fonction des facteurs Z de paramètres β : $\mu = f(Z, \beta)$. La formulation de base du modèle APC utilise la forme des modèles linéaires généralisés (GLM) :

$$l(\mu(a, t)) = a(a) + p(t) + c(t - a) \quad (1.6)$$

l est une fonction de la réponse ; a , p et c sont respectivement les fonctions de l'âge a , de la période t et de la cohorte $t - a$. Ces trois dernières n'ont pas de contraintes de linéarité et les variables peuvent être discrètes ou continues.

Les transformations logarithmique et logistique (les plus courantes en épidémiologie) de variables des modèles linéaires généralisés (GLM) améliorent généralement la qualité de l'ajustement des données en autorisant une certaine flexibilité des effets. La prise en compte des variables continues fait toujours l'objet de multiples publications et développements méthodologiques [Durrleman 1989], [Royston 2000],

[Greenland 1995]. Les méthodes de modélisations proposées sont variées : polynômes, polynômes fractionnaires, splines de régression ou de lissage...

1.1.2.3 Estimations et projections

Il y a donc plusieurs façons d'exprimer le modèle, ici, nous nous focaliserons sur l'expression et la méthode d'estimation proposées par Burgio et Frova [Burgio 2015] dont un des objectifs était de fournir ensuite des projections de mortalité. Leur modèle de projection est fondé sur une analyse séparée des trois composantes, âge, période et génération. Leur choix s'est porté sur une transformation logarithmique des taux de mortalité avec un ajustement par une fonction polynomiale des facteurs âge, période et génération.

Le modèle décrit par l'équation 1.6 s'écrit :

$$\ln(\mu(a, t)^*) = q + a(a) + p(t) + c(t - a)$$

avec

$$\ln(\mu(a, t)^*) = q + \sum_i b_i a^i + \sum_j c_j t^j + \sum_k d_k (t - a)^k$$

où $i = 1, \dots, h_1$; $j = 1, \dots, h_2$; $k = 1, \dots, h_3$

$\mu(a, t)^*$ désigne la valeur théorique du taux de mortalité à l'âge a , l'année t , et $q, b_1, \dots, b_{h_1}, c_1, \dots, c_{h_2}, d_1, \dots, d_{h_3}$ les paramètres estimés par la méthode des moindres carrés. Les degrés h_1, h_2, h_3 des polynômes sont déterminés par une procédure itérative incluant, à chaque itération, un paramètre différent dans le modèle (âge, période ou cohorte).

L'extrapolation des effets de période est la plus compliquée à envisager du fait de la nature imprévisible de facteurs conjoncturels. Dans le but d'exprimer au mieux l'effet période pour en donner des projections, Burgio et Frova ont proposé de le distinguer *via* deux composantes additives : un mouvement de fond et les écarts à la tendance. Pour la projection, seul le mouvement de fond sera extrapolé et les écarts à la tendance seront supposés nuls.

Ainsi, les trois fonctions a, p et c vont subir une même procédure et les nouvelles fonctions en résultant a^*, p^* et c^* seront liées par les équations suivantes :

$$\begin{cases} p^*(t) = p(t) - B \times (t - t_1) \\ a^*(a) = a(a) - B \times (a - a_1) \\ c^*(t - a) = c(t - a) - B \times ((t - a) - (t_1 - a_1)) \end{cases} \quad (1.7)$$

avec $B = \frac{p(t_n) - p(t_1)}{t_n - t_1}$, et $p^*(t_1) = p^*(t_n) = p(t_1)$, t_1 et t_n étant respectivement la première et la dernière année d'observation.

Lors de l'extrapolation, la seule différence dans les équations ci-dessus concerne $p^* = p^*(t_1) = p^*(t_n)$ du fait de l'hypothèse d'absence d'écart à la tendance. Cependant, compte tenu de la relation linéaire qui unit les trois facteurs (période = âge +

génération), le mouvement de fond de la composante de période reste implicitement compris dans les effets linéaires d’âge et de génération [Negri 1990].

1.1.2.4 Avantages et Limites

Le modèle APC permet d’analyser les séries temporelles de taux en faisant ressortir les effets de période et de génération. Ce modèle est parfaitement adapté à la description des informations contenues dans les données de mortalité par âge en conservant les caractéristiques passées, fonctions de chaque génération.

Dans la phase analytique, un poids est donné à chacune des trois composantes, l’âge, la période et la cohorte, pour mieux comprendre le phénomène étudié, et ainsi, permettre de faire des projections futures sur la forme des courbes de taux par âge et génération, et de formuler d’éventuelles hypothèses sur les effets de période. Les comportements de générations ont des variations beaucoup moins brutales que les à-coups d’une période.

Il reste cependant difficile d’estimer dans quelle proportion la variation d’un taux peut être attribuée à l’âge, à la période et à la génération puisque les trois variables sont interdépendantes. Plusieurs méthodes existent pour estimer les trois effets en tenant compte des problèmes associés à la multi-colinéarité [Tableau 1999], [Wilmoth 1990]. Il s’agit alors de choisir une méthode convenant au problème étudié.

1.2 Les modèles multi-états

1.2.1 Contexte

Les modèles multi-états ont connu un essor dans les années 1990 comme alternative aux modèles de survie classiques [Cox 1972], [Clayton 1978]. Ce type de modèle permet d’étudier des dynamiques de population en modélisant l’évolution d’un sujet dans différents états. C’est pourquoi, ils sont particulièrement adaptés à la modélisation de l’état de santé d’un sujet au cours du temps à travers les différents stades de la maladie. C’est le cas, notamment, des études longitudinales en santé où, très souvent, les données individuelles sont récupérées à certains temps de suivi, sous forme de mesures ou d’états.

La littérature est riche dans ce domaine, nous nous sommes principalement appuyés sur les travaux menés par Keiding [Keiding 1991], Andersen [Andersen 1993], Commenges [Commenges 1998], [Commenges 1999], et Hougaard [Hougaard 1999] pour décrire les notions et définitions générales que nous rencontrons dans ces modèles.

Une fois les bases posées, nous détaillons plus particulièrement le modèle multi-états qui sera central dans cette thèse.

1.2.2 Principaux modèles

Les modèles multi-états permettent une grande flexibilité de modélisation puisque le nombre d’états que l’ont fait intervenir ainsi que les échanges entre ces

états sont indéterminés. En effet, la construction du modèle, va dépendre de l'objet de l'étude, des problématiques, de la qualité et du détail des données à disposition. Cependant, certaines "catégories" de modèles multi-états sont déjà bien en place et très utilisées dans ce domaine.

Nous pouvons citer en exemple de modèles classiques :

- Les modèles de survie [Therneau 2000] : dans ce type de modèle, il n'y a que 2 états, l'état 0 qui est l'état initial du sujet et l'état k qui correspond dans ce cas à la survenue de l'évènement.

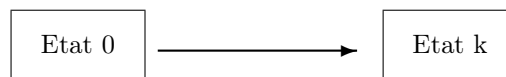


FIGURE 1.2 – Modèle de survie

- Les modèles progressifs [Hougaard 1999] : où les sujets passent de l'état 0 à l'état k en passant par les états intermédiaires i ($1 < i < k$).

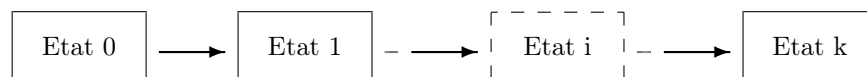


FIGURE 1.3 – Modèle progressif

Dans ces modèles, l'information sur l'état présent renseigne sur les états précédents. Nous parlons alors de processus pour représenter les différents états successivement occupés à chaque temps d'observation.

- Les modèles à risques compétitifs ou concurrents [Andersen 1993] : à l'état 0, les sujets peuvent aller dans un et seulement un des états i ($1 < i < k$), c'est pourquoi ces états sont dits en compétitions.

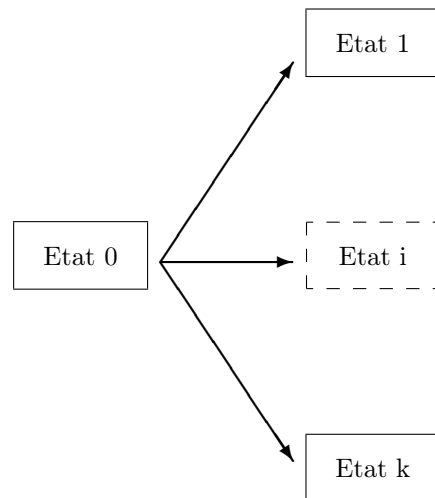


FIGURE 1.4 – Modèle à risques compétitifs

1.2.3 Définitions, notions

1.2.3.1 Les états et les intensités de transition

Les exemples que nous venons de voir nous permettent de comprendre qu'au sein d'un même modèle, nous pouvons distinguer différentes "natures" d'états. En effet, certains états sont transitoires, ce sont des états dans lesquels nous pouvons entrer et sortir ; d'autres sont absorbants, une fois entrée, nous ne pouvons plus en sortir (par exemple, ce sera le cas du décès) ; d'autres sont réversibles, car nous pouvons entrer et sortir plusieurs fois de ce même état. Ainsi, par définition un état absorbant ne peut pas être réversible.

Nous remarquons également que les chemins possibles entre les états sont représentés par des flèches. Celles-ci sont à sens unique dans les exemples présentés ce qui signifie (figure 1.2) que nous pouvons passer de l'état 0 vers l'état k, mais pas l'inverse. Dans les cas d'états réversibles, une flèche à double sens entre deux états (i,j) indique que le passage de i à j peut se faire plusieurs fois dans ce modèle.

Un changement d'état est appelé transition ou évènement et sa mesure est l'intensité de transition (cela correspond au risque instantané). L'étude de ces modèles consiste à analyser les intensités de transition notées α_{ij} d'un état i à un état j .

1.2.3.2 Les temps

Les intensités de transition peuvent dépendre de différentes bases de temps. Le choix d'une base de temps va dépendre de l'étude ou de l'application. En fonction des cas, nous privilégions généralement le temps qui est la principale source de variation pouvant mener à un changement d'état.

Dans les études épidémiologiques sur population, il est courant d'utiliser le temps calendaire comme base de temps du modèle. Nous verrons que ceci permet également de faire des projections d'indicateurs de santé. Au contraire, si l'étude a plutôt un intérêt individuel la base de temps préférentielle est l'âge. Il est également possible de choisir une durée, c'est-à-dire le temps écoulé depuis une date choisie : date d'inclusion, de suivi ou date de survenue d'un autre événement...etc.

Souvent, il est pertinent de faire dépendre les intensités de transition de plusieurs paramètres de temps. Dans la suite de ce rapport, nous notons t le temps calendaire, a l'âge et d une durée.

Donc l'écriture $\alpha_{ij}(a, t, d)$ signifie que l'intensité de transition de l'état i vers l'état j dépend de a , t et d .

1.2.3.3 Hypothèses

Le choix du temps est donc une étape importante dans la modélisation, il amène à la formulation de diverses hypothèses sur le type de modèle multi-états.

Tout d'abord, nous distinguons les modèles homogènes des modèles non-homogènes. Si les intensités de transition $\alpha_{ij}(a, t, d)$ dépendent de t , le modèle est dit non-homogène par rapport au temps t , à l'inverse le modèle est homogène par rapport au temps t si les $\alpha_{ij}(a, t, d)$ ne dépendent pas de t (ne varient pas en fonction de t). L'hypothèse d'homogénéité assume des intensités de transition constantes au cours du temps ce qui simplifie la modélisation et l'interprétation des résultats, mais cette hypothèse est souvent trop forte et restrictive pour beaucoup de maladie. C'est pourquoi, les modèles non-homogènes sont généralement plus adaptés.

Ensuite, nous distinguons les modèles sous hypothèse Markovienne des modèles sous hypothèse semi-Markovienne.

- Hypothèse Markovienne

Dans les modèles multi-états, l'hypothèse de Markov est couramment utilisée. Celle-ci suppose que l'information sur les états précédents est résumée par l'état présent (pour un temps t ou un âge a). Autrement dit, l'évolution dans l'état suivant dépendra uniquement de l'état au temps t .

Combiné aux notions précédentes, nous pouvons donc avoir des modèles Markoviens homogènes (les intensités de transition ne varient pas en fonction de t) et Markoviens non-homogènes (les intensités de transitions varient en fonction de t).

- Hypothèse semi-Markovienne

Dans le cas où les intensités de transition dépendent de la durée passée dans l'état (que nous appelons aussi temps de séjour), le modèle est semi-Markovien.

Là encore, il peut s'agir de modèles semi-Markoviens homogènes (si les intensités de transition dépendent de la durée seulement), ou non-homogènes (si elles dépendent aussi du temps t). En effet, il n'est pas rare que l'évolution d'un individu vers un autre état dépende de la durée depuis lequel il occupe l'état actuel, notamment quand cet état correspond à la survenue d'une maladie. Nous retrouvons ce type de modèles semi-Markoviens dans divers domaines tels que l'épidémiologie, par exemple Heutte et al [Heutte 2001] ont modélisé l'évolution d'un patient atteint du VIH, en fiabilité [Pérez-Ocón 2002], en finance [Janssen 1997], ...

Les modèles Markoviens non-homogènes et les modèles semi-Markoviens, peuvent être ajustés de façon à mieux modéliser l'évolution des individus et le choix du temps et donc du type de modèle dépendra de l'application et des objectifs.

1.2.4 Le modèle illness-death

Nous développons maintenant le modèle autour duquel tout le travail de cette thèse a découlé.

Un modèle illness-death est un cas particulier des modèles multi-états. Il comporte trois états, deux transitoires "non-malade" et "malade", et un absorbant "décédé" que nous noterons respectivement état 0, 1 et 2.

Dans la littérature, ces modèles sont usuellement associés à la notion de risques semi-compétitifs [Xu 2010]. Par analogie avec les modèles à risques compétitifs, depuis l'état 0 les sujets peuvent aller dans un seul des états en compétitions (1 ou 2). Cependant, dans le modèle illness-death, si le sujet passe de l'état 0 à l'état 1 il peut encore passer dans l'état 2, mais pas l'inverse, d'où la notion de "semi-compétitif". Nous parlons alors d'un évènement non-terminal pour l'état 1 (état transitoire) et d'un évènement terminal pour l'état 2 (état absorbant).

De cette façon, nous représentons (figure 1.5) les différents états possibles dans lesquels l'individu peut se trouver.

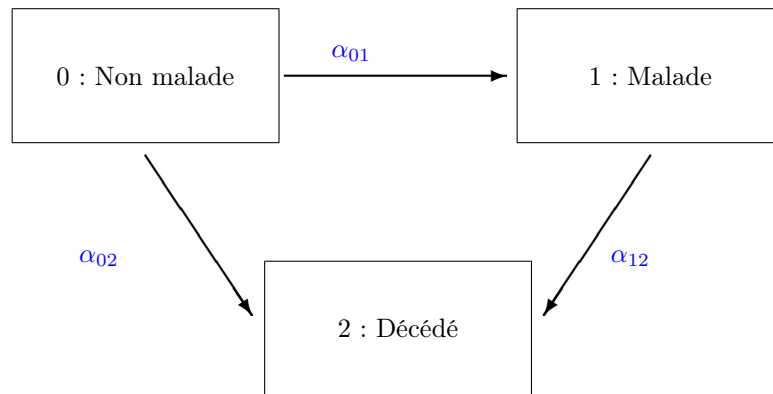


FIGURE 1.5 – Le modèle Illness-death

Ainsi, le sujet entrant dans le modèle est initialement non-malade et a deux trajectoires possibles. Il peut passer de l'état "non-malade" à l'état "décédé" soit directement, soit en passant par l'état intermédiaire "malade". Ici, nous figurons le modèle simple, mais il peut être complété en ajoutant une possibilité de guérison, dans ce cas, nous aurions l'état 1 réversible.

Comme nous l'avons défini précédemment, dans ces modèles, nous cherchons à estimer les intensités de transition entre les états qui vont permettre de mesurer le risque instantané de passer d'un état à l'autre. Dans le modèle "illness-death", elles sont au nombre de trois. Nous notons donc α_{01} l'intensité de transition entre l'état "non-malade" et l'état "malade" qui s'interprète comme l'incidence de la maladie ; α_{02} l'intensité de transition entre l'état "non-malade" et l'état "décédé" et α_{12} l'intensité de transition entre l'état "malade" et l'état "décédé" qui s'interprètent respectivement comme la mortalité des non-malades et la mortalité des malades. Il est alors possible de comparer les taux de mortalité des sujets non-malades et des sujets malades.

C'est avec ce modèle que nous faisons des projections dont l'approche méthodologique a été développée par Joly et al. [Joly 2013].

Les prochains chapitres sont développés dans le cadre du modèle illness-death. Les méthodes d'estimations sur ce modèle dans un contexte de troncature à gauche et censure à droite et par intervalles, sont détaillées dans le chapitre 2 ; et les méthodes de projections dans le chapitre 3.

Estimations avec un modèle illness-death et Indicateurs de santé

Sommaire

2.1	Notions générales sur l'analyse de données d'évènements	26
2.1.1	Censures et troncatures	26
2.1.2	Les paramètres du modèle	27
2.1.3	Méthodes d'estimations	27
2.2	Estimations	31
2.2.1	Probabilités de transition	31
2.2.2	Méthode du maximum de vraisemblance	33
2.3	Indicateurs de santé	34
2.3.1	Motivations épidémiologiques	34
2.3.2	Espérances de vie	36
2.3.3	Autres indicateurs	37
2.4	Intervalles de confiance pour les projections	37

Dans ce chapitre, nous développons quelques notions générales utiles à la compréhension du modèle illness-death que nous utilisons, ainsi que les principales méthodes d'estimations qui peuvent être utilisées pour estimer les paramètres du modèle.

Puis, nous abordons diverses quantités d'intérêts mesurables avec le modèle, notamment les probabilités de transition qui en facilitent l'interprétation.

Enfin, nous présentons le calcul d'indicateurs de santé d'un intérêt certains pour les applications qui viendront, ainsi que la méthode que nous utilisons pour calculer les intervalles de confiances.

2.1 Notions générales sur l'analyse de données d'évènements

2.1.1 Censures et troncatures

Lorsque nous utilisons des données qui ont été récoltées dans le temps, il faut pouvoir faire face à certains problèmes récurrents, car, très souvent, ces données sont en partie incomplètes. Nous voyons alors apparaître des phénomènes de troncatures et de censures.

- Censure à droite

La censure à droite est fréquemment rencontrée, elle correspond au cas où l'individu n'a pas subi l'évènement à sa date de dernière observation. Ce phénomène est toujours présent dans les données de cohorte, car l'échantillon n'est pas observé sur un temps infini. Dans ce cas de censure, nous distinguons principalement deux catégories d'individus :

- les perdus de vue : où la date de dernière observation est antérieure à la date de fin d'étude. Nous pouvons imaginer ce phénomène lors d'un déménagement du sujet par exemple.

L'hypothèse la plus simple suppose que la censure est indépendante de l'évènement, mais dans le cas contraire la censure est dite informative.

- les exclus vivants : où la date de dernière observation correspond à la date de fin d'étude.

- Censure par intervalle

La censure par intervalle correspond au cas où l'évènement s'est produit entre deux dates connues, mais le temps exact de transition d'un état à un autre est indéterminé. Ce cas se rencontre souvent quand les individus ne sont pas observés en temps continus. En présence de risque compétitif, il est aussi possible que, pour certains sujets, l'information sur la trajectoire suivie ne soit pas bien connue (par exemple, est-ce que le sujet était malade ou non avant de décéder)

- Troncature à gauche

Une observation est dite tronquée si elle est conditionnelle à un autre évènement. Plus particulièrement, nous parlons de troncature à gauche dans le cas où l'évènement d'intérêt n'est observable que si un autre évènement a eu lieu. Dans une cohorte, cet évènement est en général un critère d'inclusion dans l'étude.

Quelques exemples pour le modèle illness-death sont donnés *via* la figure A.1 en annexe A.1

2.1.2 Les paramètres du modèle

Rappelons que le modèle illness-death fait intervenir trois intensités de transition correspondant aux risques instantanés de passer d'un état à un autre. Ici, α_{01} , α_{02} et α_{12} sont interprétées respectivement comme l'incidence de la maladie, le risque de décès chez les non-malades et chez les malades à un moment donné. Ce sont donc les paramètres qu'il nous faut estimer, sachant que les transitions d'un état à l'autre peuvent être observées en temps continus ou en temps discrets.

Cependant, l'interprétation des intensités de transition n'est pas des plus naturelle et nous utilisons plus souvent les probabilités de transition (comme nous le verrons au paragraphe 2.2.1) faisant appel aux intensités de transition cumulées, notées :

$$A_{ij}(0, t) = \int_0^t \alpha_{ij}(u) du$$

2.1.3 Méthodes d'estimations

Nous pouvons distinguer les approches d'estimations paramétriques, non-paramétriques et semi-paramétriques.

L'utilisation d'une approche paramétrique nécessite de faire une hypothèse sur la forme des intensités de transition. Il s'agit donc de choisir une loi de distribution déterminant une classe de fonctions d'un nombre fini de paramètres relativement proches/semblables des intensités de transition. Cette approche a l'avantage de faciliter la phase d'estimation et de fournir des estimations en temps continus, mais la loi de distribution choisie pour décrire la forme des intensités de transition peut ne pas être la bonne.

L'approche non-paramétrique pallie ce problème puisqu'elle ne nécessite aucune hypothèse, en revanche les estimations ne sont pas des fonctions continues et ne permettent pas de traiter simplement les cas de censures par intervalles.

C'est pourquoi il est courant (surtout en survie [Therneau 2000]) de choisir un compromis en utilisant une approche semi-paramétrique. Ceci induit que les intensités de transition seront des fonctions en partie dépendantes de paramètres et en partie considérées comme non-paramétriques.

Nous pouvons citer différents exemples tirés de la littérature illustrant ces approches, Janssen et Limnios [Janssen 1999] présentent un panel de méthodes paramétriques et non-paramétriques.

2.1.3.1 Principales lois paramétriques

Dans le cas des modèles paramétriques, nous détaillons quelques-unes des principales distributions que nous avons utilisées, que nous retrouvons fréquemment

dans le domaine de la santé et notamment en survie.

– Loi de Weibull

La loi de Weibull est une loi de probabilité continue à deux paramètres : γ le paramètre d'échelle de la distribution, et κ le paramètre de forme (avec $\gamma > 0$ et $\kappa > 0$). Sa fonction de densité f au temps t en fonction des paramètres κ et γ est la suivante :

$$f(t; \kappa, \gamma) = \frac{\kappa}{\gamma} \left(\frac{t}{\gamma}\right)^{\kappa-1} e^{-\left(\frac{t}{\gamma}\right)^\kappa}$$

Il est d'usage de l'écrire simplement $f(t)$.

Ainsi par la relation qui lie la fonction de densité, la fonction de répartition F et la fonction de survie S au temps t , à savoir $f(t) = -\frac{dS(t)}{dt} = \frac{dF(t)}{dt}$ (que nous écrivons plus simplement $f(t) = -S'(t) = F'(t)$), nous obtenons :

$$S(t) = e^{-\left(\frac{t}{\gamma}\right)^\kappa}$$

La fonction de risque λ au temps t qui représente la probabilité de subir l'évènement dans un court intervalle de temps après t , conditionnellement au fait d'avoir survécu jusqu'à t s'écrit :

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\kappa}{\gamma} \left(\frac{t}{\gamma}\right)^{\kappa-1}$$

Et donc, la fonction de risque cumulé Λ n'est autre que :

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

Dans le cas de la loi de Weibull, la fonction de risque est monotone (croissante ou décroissante) au cours du temps. Nous pouvons remarquer que lorsque nous fixons $\kappa = 1$, nous retrouvons une loi exponentielle de paramètre $(1/\gamma)$.

– Loi de Gompertz Makeham

Cette fois, il s'agit d'une loi de probabilité continue à trois paramètres. Le modèle de Gompertz-Makeham date de 1860 et fait partie des modèles de référence pour l'ajustement de taux de mortalité en fonction de l'âge. Le taux de mortalité est modélisé comme la somme de termes indépendants de l'âge (termes de Makeham) et de termes dépendants de l'âge (fonction de Gompertz). La fonction de densité de cette loi au temps t s'écrit :

$$f(t) = (\alpha + \beta e^{\gamma t}) e^{-\alpha t - \frac{\beta}{\gamma} e^{\gamma t} - 1}$$

avec β la mortalité de base, γ l'influence de l'âge et $\alpha (> 0)$ pour tenir compte de la mortalité accidentelle.

On en déduit sa fonction de survie S au temps t :

$$S(t) = e^{-\alpha t - \frac{\beta}{\gamma}(e^{\gamma t} - 1)}$$

et sa fonction de risque λ au temps t :

$$\lambda(t) = \alpha + \beta e^{\gamma t}$$

2.1.3.2 Principales approches non-paramétriques

Les approches non-paramétriques reposent sur une approximation des intensités par des fonctions constantes par morceaux, il s'agit d'obtenir les valeurs de la fonction sur chaque intervalle. Cependant, il existe aussi des méthodes donnant des estimations continues.

Les principaux estimateurs non-paramétriques (d'un modèle de survie) sont l'estimateur de Kaplan-Meier de la fonction de survie et l'estimateur de Nelson-Aalen du risque cumulé.

– Kaplan-Meier

L'estimateur de Kaplan-Meier suppose que survivre après un temps t , c'est être en vie juste avant t et ne pas mourir au temps t .

Il permet d'estimer la fonction de survie au temps t :

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

n_i étant le nombre de personnes à risque juste avant le temps t_i et d_i le nombre d'évènements au temps t_i .

Cette méthode prend en compte les données censurées à droite. Nous pouvons également obtenir un estimateur de Kaplan-Meier dans le cas de données tronquées, mais pas pour les données censurées par intervalles, car les temps de décès ne sont pas connus.

– Nelson-Aalen

Introduit en 1978, cet estimateur généralise celui de Nelson (1972).

L'estimateur de la fonction de risque cumulé s'écrit :

$$\hat{\Lambda}(t) = \sum_{t_i < t} \frac{d_i}{n_i}$$

n_i étant le nombre de personnes à risque juste avant le temps t_i et d_i le nombre d'évènements au temps t_i .

L'interprétation du risque cumulé est moins naturelle que la fonction de survie, mais sa dérivée donne une estimation de la fonction de risque. Cet estimateur peut être appliqué dans le cas de données censurées ou incomplètes, mais pas lors de données censurées par intervalles, pour les mêmes raisons évoquées pour l'estimateur de Kaplan-Meier.

– Autres estimateurs

D'autres estimateurs, que nous ne détaillerons pas ici, existent, nous pouvons citer notamment l'estimateur de la probabilité de faire une transition pour les modèles multi-états Markoviens non-homogènes, proposé par Aalen and Johansen [Aalen 1978], ainsi que les alternatives récentes développées par Titman [Titman 2015]. Notons aussi que Frydman et Szarek [Frydman 1999], [Frydman 2009], [Frydman 2010] ont proposé des estimateurs non-paramétriques des intensités de transition cumulées, pour des modèles illness-death avec données censurées par intervalles. Ou encore, l'estimateur des intensités de transition proposé par Ouhbi et Limnios [Ouhbi 1999] dans le cadre d'un modèle semi-Markovien avec un cas de censure, et sa généralisation apportée par Saint-Pierre [Saint-Pierre 2005] pour l'adapter à plusieurs cas de censures ainsi qu'aux modèles avec des états absorbants.

2.1.3.3 Principales approches semi-paramétriques

Une autre approche pour estimer des fonctions continues sans faire d'hypothèses paramétriques est d'utiliser des fonctions splines [Rosenberg 1995]. Comme l'approche paramétrique, ce type d'approche est basé sur la maximisation de la vraisemblance (qui sera décrite dans la section 2.2.2) et permet aisément de considérer les cas de données censurées par intervalles.

Par exemple Joly et Commenges [Joly 1999] utilisent une approche semi-paramétrique avec maximisation de la vraisemblance pénalisée, pour un modèle à trois états ayant des données censurées et tronquées.

L'utilisation de la maximisation par vraisemblance pénalisée permet de favoriser certains estimateurs par rapport à d'autre dans le but d'obtenir une meilleure adéquation. Cette technique est particulièrement utilisée afin d'obtenir des estimateurs non-paramétriques lisses de fonctions de risque ou d'intensités de transition. Pour cela, il s'agit d'ajouter une pénalité au logarithme de la vraisemblance et de choisir le terme de la pénalisation. Un exemple de vraisemblance pénalisée est le suivant :

$$pl(\lambda) = l(\lambda) - \kappa \int_0^\infty \lambda'^2(u) du$$

Avec l la log-vraisemblance, κ le paramètre de lissage (il y a autant de paramètres de lissage que de transitions entre états possibles) et λ la fonction de risque. Cet estimateur est calculé *via* une base de fonctions splines, qui permet d'obtenir une approximation de λ . La log-vraisemblance est donc pénalisée par un terme qui est d'autant plus grand que la fonction de risque est peu lisse.

Dans une autre étude, Joly *et al.* ([Joly 2002]) utilisent une base de M-splines cubiques, qui sont des fonctions polynomiales par morceaux combinées linéairement pour approcher une fonction sur un intervalle.

Alors la fonction de risque s'écrit :

$$\lambda(t) = \theta^T M(t)$$

où $M(t) = (M_1(t), \dots, M_m(t))^T$ est le vecteur des valeurs en t de m fonctions splines et θ le vecteur de coefficients des splines. Son estimateur $\hat{\theta}$ pour un κ fixé est obtenu en maximisant $pl(\lambda)$.

L'approche par vraisemblance pénalisée est vue comme une approche intermédiaire entre les approches paramétriques et non-paramétriques. L'estimateur y est défini non-paramétriquement comme le maximum de la vraisemblance pénalisée et il est approché par une base de fonctions splines.

Ces fonctions permettent de considérer des données, dans un contexte de censures par intervalles et de troncatures, comme une fonction paramétrique, mais sans hypothèses sur la distribution si ce n'est qu'elle doit être deux fois dérivables.

Une des limites de l'utilisation de la méthode d'estimation par vraisemblance pénalisée se trouve dans le choix du nombre de fonctions splines à utiliser ainsi que le choix des paramètres de lissage. En effet, le nombre de fonctions splines ne doit être ni trop grand ni trop petit pour minimiser l'erreur d'approximation et éviter les problèmes numériques, quant au choix du paramètre de lissage, il se fait en utilisant une méthode de validation croisée approchée.

De manière générale, nous utilisons plutôt des approches paramétriques ou semi-paramétriques qui permettent d'inclure facilement des variables et d'obtenir des estimations continues des intensités de transition. Celles-ci vont donc faire appel à la méthode de la maximisation de la vraisemblance.

2.2 Estimations

2.2.1 Probabilités de transition

Dans un modèle multi-états les fonctions d'intérêts sont les intensités de transition. Elles permettent l'estimation de beaucoup d'autres quantités ayant une interprétation différente, parfois plus simple, comme les probabilités de transition d'un état à un autre, ou divers indicateurs de santé (pour une interprétation d'ordre populationnelle).

Les probabilités de transition de l'état i à l'état j notées P_{ij} représentent la probabilité de passer d'un état i à un état j entre deux-temps. Elles peuvent être exprimées directement en fonction des intensités de transition $\alpha_{ij}(\cdot)$ (voir par exemple les travaux de Putter *et al.* [Putter 2007]).

Pour être en accord avec les différents travaux menés dans cette thèse, les probabilités suivantes sont définies en fonction des âges et pour le modèle illness-death de la figure (1.5).

La probabilité notée $P_{ij}(a, a^*)$ se définit comme la probabilité pour un sujet dans l'état i à l'âge a , d'être dans l'état j à l'âge a^* (avec $a^* > a$).

Nous notons :

$$P_{00}(a, a^*) = e^{-A_{01}(a, a^*) - A_{02}(a, a^*)}$$

la probabilité pour un sujet vu vivant non-malade à l'âge a , de le rester jusqu'à l'âge a^* .

$$P_{01}(a, a^*) = \int_a^{a^*} e^{-A_{01}(a, u) - A_{02}(a, u)} \alpha_{01}(u) e^{-A_{12}(u, a^*)} du$$

la probabilité pour un sujet vu vivant non-malade à l'âge a , de le devenir entre les âges a et a^* et de rester vivant.

$$P_{11}(a, a^*) = e^{-A_{12}(a, a^*)}$$

la probabilité pour un sujet vu vivant malade à l'âge a , de le rester jusqu'à l'âge a^* .

$$P_{02}^1(a, a^*) = \int_a^{a^*} e^{-A_{01}(a, u) - A_{02}(a, u)} \alpha_{01}(u) e^{-A_{12}(u, a^*)} \alpha_{12}(a^*) du$$

la probabilité pour un sujet vu vivant non-malade à l'âge a , de décéder à l'âge a^* , après avoir développé la maladie entre ces deux âges.

Nous notons également P_v la somme de deux probabilités :

$$P_v(a, a^*) = P_{00}(a, a^*) + P_{01}(a, a^*)$$

Cela correspond à la probabilité pour un sujet vivant et non-malade à l'âge a , d'être en vie à l'âge a^* (malade ou non-malade).

Remarque : dans le modèle illness-death, si nous nous plaçons dans le cas d'un modèle Markovien non-homogène, alors les intensités de transition dépendent du temps et il en est de même pour le risque de décéder d'un individu malade. Dans le cas d'un modèle semi-Markovien, l'intensité de transition α_{12} est modifiée. En effet, l'incidence de la maladie α_{01} et le risque de décès des non-malades α_{02} ne diffèrent pas du cadre Markovien (les sujets étant initialement non-malades, le temps passé dans l'état 0 et le temps courant sont semblables), alors que le risque de décès d'un malade α_{12} dépend du temps depuis lequel il est malade.

La plupart des quantités présentées s'écrivent de la même manière quelle que soit l'hypothèse (Markovienne/semi-Markovienne), cependant, il faut faire attention de bien différencier α_{12} selon l'hypothèse. En effet, ici, nous présentons cette intensité comme dépendante de l'âge, mais il va de soi que dans le cadre d'un modèle semi-Markovien, elle doit dépendre de la durée passée avec la maladie.

Nous distinguons donc deux cas qui nous seront utiles pour la suite :

- dans le cas d’un modèle semi-Markovien homogène, nous écrivons l’intensité de transition de l’état malade à l’état décédé $\alpha_{12}(d)$ et son intensité de transition cumulée sera $A_{12}(0, d)$ (où d est la durée passée avec la maladie).
- dans le cas d’un modèle semi-Markovien non-homogène, nous ferons dépendre la mortalité des malades de l’âge et de la durée passée avec la maladie. Nous développerons un modèle additif où la mortalité des malades s’écrit comme la somme d’une fonction de l’âge et d’une fonction de la durée d passée avec la maladie. Ce modèle sera décrit dans la section 3.3, ici nous le simplifions par :

$$\alpha_{12}(a, d) = \alpha_{02}(a) + \alpha^d(d) \quad (2.1)$$

et son intensité de transition

$$A_{12}(a, a^*, d) = A_{02}(a, a^*) + A^d(0, a^* - a)$$

Remarque : la notation $A_{12} = (a, a^*, d)$ plutôt que $A_{12} = (a, a^*)$ sert à éviter de faire une confusion en rappelant que la durée intervient également dans l’expression de la mortalité des malades.

2.2.2 Méthode du maximum de vraisemblance

La vraisemblance est un outil d’inférence statistique, c’est une fonction de probabilités conditionnelles qui décrit les valeurs y_i d’une loi statistique (i les individus, et les observations y_i étant i.i.d. (indépendantes et identiquement distribuées)) en fonction du vecteur de paramètres θ supposé connu. Elle s’exprime à partir de la fonction de densité $f(Y; \theta)$, ou, plus souvent dans notre cas, à partir des contributions individuelles de chaque individu i , notée L_i .

$$L = \prod_{i=1}^n L_i$$

Les différentes contributions pouvant intervenir dans un modèle illness-death sont détaillées dans la thèse de C. Touraine [Touraine 2013a] et sont résumées dans l’annexe A.1.

Cependant, le cas d’un modèle semi-Markovien non-homogène en âge n’a pas été traité. Celui-ci étant un des développements de cette thèse, nous présentons, ici, les contributions individuelles dans le cas où α_{12} s’exprime sous la forme d’un modèle additif tel celui de l’équation 2.1.

Nous avons défini 4 trajectoires possibles pour les individus (figure A.1, annexe A.1). On note \tilde{T} l’âge, qui correspond à l’âge de dernière nouvelles C , si le sujet était vivant à la dernière visite, ou bien à l’âge de décès T ; et δ_2 est l’indicateur de décès (1 si le sujet est décédé, 0 sinon). Pour ce modèle semi-Markovien non-homogène, nous avons donc les contributions suivantes :

- Si le sujet i n'était pas malade à l'âge L auquel il a été vu pour la dernière fois :

$$L_i = e^{-A_{01}(a,L)-A_{02}(a,L)} \left\{ e^{-A_{01}(L,\tilde{T})-A_{02}(L,\tilde{T})} \alpha_{02}(\tilde{T})^{\delta_2} + \int_L^{\tilde{T}} e^{-A_{01}(L,u)-A_{02}(L,u)} \alpha_{01}(u) e^{-(A_{02}(u,\tilde{T})+A^d(0,\tilde{T}-u))} (\alpha_{02}(\tilde{T}) + \alpha^d(\tilde{T}-u))^{\delta_2} du \right\} \quad (2.2)$$

- Si le sujet i a été diagnostiqué malade à l'âge R (et vu non-malade pour la dernière fois en L) :

$$L_i = e^{-A_{01}(a,L)-A_{02}(a,L)} \left\{ \int_L^R e^{-A_{01}(L,u)-A_{02}(L,u)} \alpha_{01}(u) e^{-(A_{02}(u,\tilde{T})+A^d(0,\tilde{T}-u))} (\alpha_{02}(\tilde{T}) + \alpha^d(\tilde{T}-u))^{\delta_2} du \right\} \quad (2.3)$$

sachant que $\tilde{T} > R$.

Ces deux contributions correspondent aux 4 trajectoires en fonction de la valeur de δ_2 .

Pour les approches paramétriques et semi-paramétriques, la méthode du maximum de vraisemblance permet d'obtenir des estimateurs des paramètres de modèles statistiques avec de bonnes propriétés asymptotiques, et également d'effectuer des tests sur ces paramètres.

L'estimateur du maximum de vraisemblance (MLE pour "maximum likelihood estimator" en anglais) $\hat{\theta}$ est la valeur de θ qui maximise la vraisemblance $L(\theta; Y)$. En général, nous maximisons la log-vraisemblance $\log L(\theta; Y)$ qui est la somme des log-vraisemblances individuelles pour chaque i .

$$\log L(\theta; Y) = \sum_{i=1}^n \log L_i(Y; \theta)$$

2.3 Indicateurs de santé

Pour faciliter la lecture, toutes les quantités définies dans cette partie ne sont dépendantes que de l'âge au temps t (t étant volontairement omis dans les notations). De plus, nous considérons également l'année de naissance des individus $t - a$.

2.3.1 Motivations épidémiologiques

Les indicateurs de santé déterminent l'état de santé de l'individu. Ils peuvent être de natures personnels, sociaux économiques et politiques ou encore environnementaux.

Ils en existent beaucoup, pour notre travail, nous avons plus spécialement besoin de développer des indicateurs de morbidité pour décrire la fréquence des maladies (comme la prévalence, l'incidence...), des indicateurs de mortalité pour décrire la fréquence des décès, et des indicateurs plus généraux tels que l'espérance de vie..., pour décrire l'état de la population.

Différentes sources d'informations peuvent servir pour obtenir des indicateurs de santé :

- les sources de données démographiques telles que l'INSEE (Institut national de la statistique et des études économiques) et l'INED (Institut national d'études démographiques) en France,
- les statistiques de morbidité avec le système des maladies à déclaration obligatoire, les réseaux sentinelles, les systèmes de registre, l'assurance maladie et également les centres hospitaliers,
- et certaines études épidémiologiques de type cohorte où un certain nombre de personnes sont suivies dans le temps, ainsi que l'évolution de leur état de santé. Nous pouvons distinguer les cohortes de malades, où tous les individus sont porteurs de la même maladie, et les cohortes en population générale, où les individus sont choisis indépendamment de leur état de santé. Les cohortes font partie des meilleurs dispositifs pour étudier le rôle de facteurs environnementaux dans l'apparition de certaines maladies, pour ensuite essayer de mieux les contrôler.

En épidémiologie l'un des principaux objectifs est d'étudier la distribution et les déterminants des problèmes de santé dans la population et les applications à la prévention des problèmes de santé. Ainsi, les indicateurs de santé sont des outils de décision pour l'épidémiologie.

En fonction des objectifs, ces indicateurs serviront à décrire l'état de santé d'une population, puis à analyser les déterminants des problèmes de santé en vue de proposer des interventions efficaces et enfin à évaluer l'impact des interventions.

Dans la partie suivante, nous détaillons le calcul de plusieurs indicateurs pertinents pour une année t , que nous avons choisi de développer après en avoir discuté avec plusieurs experts des maladies chroniques liées au vieillissement. Pour ce faire, nous utilisons les probabilités de transition dans l'écriture de ces indicateurs. De plus, notez que nous donnons les définitions générales des indicateurs de santé en intégrant jusqu'à l'infini, mais en réalité, nous nous arrêtons vers 105 ans (car, nous avons trop peu d'informations pour pouvoir estimer les paramètres du modèle après cet âge, et de plus, la probabilité d'être encore vivant est faible.

2.3.2 Espérances de vie

Étudier les espérances de vie dans ce modèle se révèle particulièrement intéressant, car il est possible de les distinguer en fonction de l'état de santé de départ et des évolutions possibles. Nous pouvons donc nous intéresser à l'espérance de vie sans la maladie que nous distinguons de l'espérance de vie d'un non-malade : en effet pour cette dernière, nous estimons l'espérance de vie d'un individu non-malade à un âge donné, quelle que soit l'évolution de l'état de santé qui l'attend ; mais également à l'espérance de vie d'un malade et à l'espérance de vie globale (que l'individu soit malade ou non).

Sur la base des travaux menés par Touraine *et al.* [Touraine 2013b] nous adaptons et détaillons les espérances de vie en fonction de l'état de santé du sujet à l'âge a .

- *L'espérance de vie sans la maladie* à l'âge a qui est le nombre d'années qu'un sujet non-malade à l'âge a peut s'attendre à vivre sans la maladie. Nous la notons :

$$LE_{00}(a) = \int_a^\infty P_{00}(a, u) du$$

- *L'espérance de vie d'un malade* à l'âge a qui est le nombre d'années qu'il reste à vivre à un sujet déjà malade à l'âge a . Nous la notons :

$$LE_{11}(a) = \int_a^\infty P_{11}(a, u) du$$

- *L'espérance de vie d'un non-malade* à l'âge a qui est le nombre d'années qu'il reste à vivre à un sujet non-malade à l'âge a , quelle que soit l'évolution de son état (malade/non-malade). Nous la notons :

$$LE_0(a) = \int_a^\infty (P_{00}(a, u) + P_{01}(a, u)) du$$

- *L'espérance de vie générale* à l'âge a qui est la moyenne pondérée de l'espérance de vie d'un malade LE_{11} et d'un non-malade LE_0 . à l'âge a . Nous la notons :

$$LE_{..}(a) = \pi_0(a) \int_a^\infty (P_{00}(a, u) + P_{01}(a, u)) du + (1 - \pi_0(a)) \int_a^\infty P_{11}(a, u) du$$

avec $\pi_0(a) = \frac{P_{00}(0, a)}{P_v(0, a)}$ la proportion de non-malades à l'âge a , parmi les sujets vivants.

- *Le nombre moyen d'années passées avec la maladie* à l'âge a qui est le nombre moyen d'années qu'un sujet non-malade à l'âge a peut s'attendre à vivre avec la maladie. Cela revient à soustraire deux espérances de vie. Nous le notons :

$$T_{11}(a) = LE_0(a) - LE_{00}(a)$$

2.3.3 Autres indicateurs

- La probabilité de développer la maladie correspond au risque de développer la maladie avant de décéder pour un sujet vivant et non-malade à l'âge a . Nous la notons :

$$F_{01}(a) = \int_a^{\infty} P_{00}(a, u) \alpha_{01}(u) du$$

- L'âge moyen de début de la maladie pour quelqu'un qui développe la maladie, se note :

$$AV_{01} = \frac{\int_0^{\infty} u P_{00}(0, u) \alpha_{01}(u) du}{F_{01}(0)}$$

- La prévalence de la maladie l'année t qui est le nombre total de personnes malades de tous âges l'année t (ici par exemple, nous fixons à 100 l'âge maximum que peut atteindre la population). Nous la notons :

$$Prev(t) = \sum_{k=0}^{100} \nu(0, t - 100 + k) P_{01}(0, 100 - k)$$

Remarque : dans l'écriture de la prévalence, nous donnons un âge maximum plutôt que ∞ , car, généralement, nous avons très peu d'informations sur les maladies après cet âge. Or, si nous n'avons pas d'informations, nous ne pouvons pas estimer les paramètres du modèle. Cette limite d'âge est donc généralement imposée par les données propres à la maladie, car, souvent, les limites d'âge des données de population (nationales) sont supérieures.

2.4 Intervalles de confiance pour les projections

Pour chaque quantité décrite, il est possible de calculer des intervalles ou bandes de confiance. L'approche par simulation de van den Hout *et al.* [van den Hout 2008], [van den Hout 2010] est utilisée. Les intervalles de confiance (IC) sont construits en utilisant la variabilité des paramètres estimés du modèle.

La démarche est la suivante :

Nous notons $\hat{\phi}$ le vecteur des paramètres estimés et $\hat{V}_{\hat{\phi}}$ la matrice de variance-covariance estimée. Puis, nous générons n vecteurs de paramètres $\phi^{(1)}, \dots, \phi^{(n)}$ suivant une loi Normal d'espérance $\hat{\phi}$ et de matrice de variance covariance $\hat{V}_{\hat{\phi}}$.

Pour chaque $\phi^{(k)}$, les intensités de transition, probabilités de transition et indicateurs de santé sont estimés au temps t .

Enfin, nous récupérons le 2,5^{ème} et le 97,5^{ème} percentile empirique de ces simulations pour déterminer les intervalles de confiance à 95%. Cette procédure est répétée pour chaque temps t pour obtenir des intervalles de confiance pour tous les indicateurs dépendants du temps.

Méthodes de projections avec un modèle illness death

Sommaire

3.1 Principales sources de données	40
3.1.1 Données de cohorte	40
3.1.2 Données Nationales démographiques	40
3.1.3 Autres sources de données possibles	40
3.2 Modèle Illness death et hypothèses Makoviennes	41
3.2.1 Modèles	41
3.2.2 Méthodologie pour l'estimation des paramètres	42
3.2.3 Équations différentielles	43
3.3 Modèle Illness death et hypothèses semi-Makoviennes	45
3.3.1 Modèles	45
3.3.2 Méthodologie pour l'estimation des paramètres	46
3.3.3 Équations différentielles	47
3.4 Projections	48

Dans ce chapitre, nous développons toutes les étapes de l'approche proposée, de l'estimation à partir des données et en fonction des hypothèses, jusqu'à la méthode de projection.

Tout d'abord, nous présentons les différentes sources de données qui sont utilisées. Puis, nous présentons les modèles illness-death que nous avons choisi de développer (Markoviens ou semi-Markoviens), en détaillant leurs hypothèses et les étapes de l'estimation. Enfin, nous expliquons comment faire des projections à partir de ces estimations.

3.1 Principales sources de données

3.1.1 Données de cohorte

Les données de cohorte permettent de suivre dans le temps un échantillon de personnes, sélectionnées selon des critères, afin d'étudier l'évolution de leur état de santé et la survenue d'évènements. Au cours du suivi, les mêmes informations sont recueillies sur chacun des individus, nous avons donc des données individuelles et longitudinales. L'intérêt de ce type de données et qu'elles permettent notamment de mesurer l'incidence (nombre de nouveau cas) d'une pathologie ou d'un événement, et d'évaluer et suivre dans le temps les relations expositions-risques...etc.

Ainsi, les données de cohorte peuvent être utilisées pour déterminer les paramètres du modèle illness-death à savoir l'incidence de la maladie, la mortalité des malades et des non-malades. Plusieurs méthodes d'estimations sont présentées dans le chapitre 2, il s'agit de choisir celle qui répond le mieux à nos objectifs en fonction de la qualité des données et des hypothèses du modèle.

Cependant, notons que lorsqu'il s'agit d'une cohorte "fermée", c'est-à-dire que se sont les mêmes sujets suivis tout au long de l'étude sans inclusion nouvelle, il est rarement possible de prendre en compte une évolution au cours du temps dans l'estimation des paramètres du modèle.

3.1.2 Données Nationales démographiques

Notre travail s'élevant à l'échelle de la population, nous avons besoin d'accéder à certaines informations sur la population. Pour cela, il existe des instituts nationaux de statistiques, démographie et/ou d'économie dont les objectifs principaux sont de produire, analyser et publier des statistiques nationales officielles (voir même élargies à d'autres pays). La récolte et l'analyse des données font l'objet d'un long travail de recensement et d'étude de la population, leurs résultats et bases de données sont généralement en libre accès sur leur site. De plus, en utilisant des techniques et méthodes évoluées, ils produisent des projections de certains indicateurs de santé (en restant dans le cadre général, sans considérer de maladie). Pour cela, ils s'appuient sur des hypothèses d'évolution par sexe et âge et présentent plusieurs scénarios qui peuvent également tenir compte d'autres critères comme le mode de vie ou les progrès médicaux. L'avis des experts est, en général, considéré pour établir plusieurs théories.

Certains paramètres, tels que la taille de la population à risque et la mortalité générale dans la population, seront issus de ces données nationales

3.1.3 Autres sources de données possibles

Les deux sources de données que nous avons décrites sont celles le plus couramment utilisées, mais il en existe d'autres. Par exemple, les paramètres du modèle peuvent être estimés via des données d'assurance maladie avec lesquelles il est possible d'obtenir les taux d'incidence α_{01} de la maladie ou encore le nombre de cas

prévalent pour une année. La littérature peut également fournir ce type d'informations ainsi que les taux de mortalité des malades α_{12} .

De plus, nous nous arrêtons ici à une échelle nationale pour faire nos projections, mais il va de soi que si nous avons des données de cohorte européenne (internationale ou autre) relative à une maladie, il n'est pas difficile d'accéder à des données démographiques du même ordre et donc de faire des projections à cette échelle.

En réalité, ce sont les données permettant d'estimer les paramètres du modèle pour une maladie donnée qui sont généralement les plus difficiles à trouver.

3.2 Modèle Illness death et hypothèses Makoviennes

3.2.1 Modèles

En santé, les modèles Markoviens homogènes sont très utilisés et de nombreux travaux ont été menés sur l'ajustement de ces modèles [Alioum 2001]. Mais cette hypothèse d'homogénéité, qui suppose que les intensités de transition sont constantes au cours du temps, est assez restrictive et n'est pas toujours adaptée. Ceci sera le cas pour notre application, c'est pourquoi le modèle Markovien développé (figure 3.1) est non-homogène.

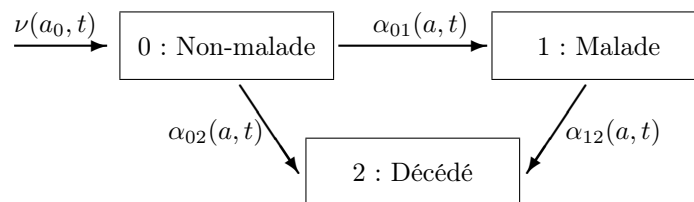


FIGURE 3.1 – Le modèle illness-death Markovien.

Dans un contexte où l'âge, la durée passée avec la maladie et le temps calendaire sont étroitement liés, diverses hypothèses sur les intensités de transition peuvent être considérées. Il convient donc de déterminer quelles hypothèses sont les plus adaptées.

Dans ce modèle, les intensités de transition sont interprétées comme des taux, or les intensités de transition d'un état non-pathologique à des états pathologiques sont généralement des taux d'incidence par âge des maladies. C'est pourquoi les trois intensités de transition dépendent de deux temps, t le temps calendaire et a l'âge (ainsi $t - a$ définit une année de naissance).

À l'entrée du modèle, nous définissons $\nu(a_0, t)$, la taille de la population à risque de développer la maladie pour un âge fixé a_0 au temps t . Ainsi, nous ferons l'hypothèse que l'incidence de la maladie est nulle avant cet âge a_0 . Dans ce modèle, a_0 est choisi en fonction de l'âge à partir duquel les sujets peuvent développer la maladie. Il est donc dépendant de la maladie d'intérêt et $a_0 = 0$ est possible si les sujets sont à risque dès la naissance.

Enfin, nous utilisons également la fonction $\alpha_2(a, t)$ pour définir le taux de mortalité générale dans la population au temps t à l'âge a .

3.2.2 Méthodologie pour l'estimation des paramètres

Habituellement, les données de cohorte permettent d'estimer les paramètres "propres" aux modèles illness-death, c'est-à-dire les intensités de transition, car se sont sur les données de cohorte que nous observons les changements d'état au cours du temps.

Alors que les données nationales démographiques, elles, vont nous permettre d'estimer la taille de la population ν et le taux de mortalité dans la population générale α_2 .

- À partir des données démographiques

La population à risque de développer la maladie ainsi que le taux de mortalité générale sont des fonctions propres à la population générale qui sont donc obtenues via les données nationales (section 3.1.2). Nous récupérons les valeurs passées et futures de α_2 et ν , récoltées en temps discrets. Puis, nous utilisons la loi de Gompertz-Makeham [Gompertz 1825] pour lisser les données de α_2 et obtenir des fonctions continues sur l'âge pour chaque année de naissance. Ceci est utile, car tous les paramètres du modèle seront estimés en temps continus, et comme nous allons le voir, certains vont être réécrits de façon à faire intervenir α_2 la mortalité générale dans la population.

- À partir des données de cohorte

Commenges [Commenges 2002] propose une modélisation basée sur la vraisemblance pénalisée, développée par Joly *et al.* [Joly 2002] avec une estimation semi-paramétrique des intensités de transition par vraisemblance pénalisée et approximation par fonctions splines. Cette méthode traite le risque semi-compétitif entre décéder et devenir malade ainsi que la censure par intervalle de l'âge de début de la maladie.

Les données de cohortes permettent d'estimer les paramètres comme des fonctions continues dépendantes de l'âge, mais souvent, elles ne captent pas l'effet du temps calendaire. Or, dans l'optique de faire des projections, il est indispensable de prendre en compte cet effet. Notamment en ce qui concerne la mortalité des non-malades, car il est très probable qu'elle ne soit pas homogène dans le temps calendaire du fait de sa similitude avec la mortalité générale dans la population, qui a évolué et continue d'évoluer dans le temps.

C'est pourquoi, nous proposons de ne pas utiliser la mortalité des non-malades estimée via les cohortes et le modèle illness-death, mais de l'estimer en fonction de la mortalité générale α_2 issue des données démographiques.

Deux cas sont distinguables, le plus souvent, il est pertinent de distinguer la mortalité des non-malades de la mortalité générale dans la population :

$\alpha_{02}(a, t) \neq \alpha_2(a, t)$, surtout pour les maladies avec une forte prévalence où une différence significative est observée. Cette démarche nécessite de résoudre une équation différentielle pour estimer $\alpha_{02}(a, t)$, que nous détaillons dans la section 3.2.3.

Sinon, il peut être supposé que $\alpha_{02}(a, t) = \alpha_2(a, t)$ ce qui simplifie les équations.

Quant à la mortalité des malades α_{12} , dans un premier temps, nous l'estimons comme fonction de l'âge directement sur les données de cohorte, sans prendre en compte une évolution de cette mortalité en fonction du temps calendaire.

$$\alpha_{12}(a, t) = \alpha_{12}(a) \quad (3.1)$$

Dans ce cas, un sujet malade âgé de 85 ans, n'a pas la même mortalité qu'un sujet malade âgé de 70 ans.

Ce premier modèle a très vite fait place à un modèle plus adapté dans le cadre de ces maladies, car la mortalité des malades varie, très vraisemblablement, aussi en fonction du temps calendaire. L'hypothèse est donc de considérer α_{12} comme proportionnelle à la mortalité des non-malades multipliée par un risque relatif de décès des malades par rapport aux non-malades :

$$\alpha_{12}(a, t) = \alpha_{02}(a, t) \times g(a, t) \quad (3.2)$$

avec $g(a, t)$ le risque relatif de décès des malades par rapport aux non-malades. Celui-ci peut être estimé directement sur les données de cohorte (et dans ce cas, il sera seulement fonction de l'âge) ou issu de la littérature.

Remarque : par la suite, nous distinguerons ces deux modèles Markoviens en les nommant ainsi :

- **M1** si la mortalité des malades s'exprime sous la forme de l'équation 3.1,
- **M2** si la mortalité des malades s'exprime sous la forme de l'équation 3.2.

Enfin, concernant l'incidence de la maladie, si aucune hypothèse d'évolution au cours du temps n'est avérée (ce qui souvent est le cas) alors nous l'estimons simplement en fonction de l'âge sur les données de cohorte.

3.2.3 Équations différentielles

Dans les équations qui vont suivre, la dépendance au temps calendaire t est omise pour alléger les notations. De plus, notons que la résolution de l'équation différentielle pour déterminer α_{02} est effectuée pour chaque année de naissance ($t - a$ avec $a > a_0$) dont nous avons besoin pour les calculs de projections.

Le taux de mortalité générale dans la population peut s'écrire comme la moyenne pondérée du risque de décéder chez les malades et les non-malades :

$$\alpha_2(a) = \frac{P_{00}(a_0, a)\alpha_{02}(a) + P_{01}(a_0, a)\alpha_{12}(a)}{P_v(a_0, a)} \quad (3.3)$$

De ce fait, en isolant α_{02} , nous avons :

$$\alpha_{02}(a) = \frac{\alpha_2(a)P_v(a_0, a) - P_{01}(a_0, a)\alpha_{12}(a)}{P_{00}(a_0, a)}$$

puis, en remplaçant P_{01} par $(P_v - P_{00})$ et en développant P_{00} nous obtenons :

$$\alpha_{02}(a) = \frac{\alpha_2(a)P_v(a_0, a) - P_v(a_0, a)\alpha_{12}(a) + [e^{-A_{01}(a_0, a) - A_{02}(a_0, a)}]\alpha_{12}(a)}{e^{-A_{01}(a_0, a) - A_{02}(a_0, a)}}$$

et donc,

$$\alpha_{02}(a) = e^{A_{01}(a_0, a) + A_{02}(a_0, a)} P_v(a_0, a) [\alpha_2(a) - \alpha_{12}(a)] + \alpha_{12}(a) \quad (3.4)$$

À cette étape, nous devons choisir l'hypothèse que nous faisons sur la mortalité des malades. Nous allons donc traiter les deux cas :

– **Modèle M1 :**

La mortalité des malades ne dépend que de l'âge, dans ce cas α_{12} est une quantité connue que nous estimons sur les données de cohorte. Nous nous ramenons donc à une équation différentielle de la forme :

$$y'(s) = a(s)e^{y(s)} + c(s)$$

où $y(s) = A_{02}(a_0, a)$, $y'(s) = \alpha_{02}(a)$, $c(s) = \alpha_{12}(a)$, et

$$a(s) = e^{A_{01}(a_0, a)} P_v(a_0, a) [\alpha_2(a) - \alpha_{12}(a)]$$

La résolution de cette équation différentielle linéaire du premier ordre à coefficients non constants, nous donne la solution suivante :

$$\alpha_{02}(a) = \alpha_{12}(a) + \frac{e^{A_{01}(a_0, a)} P_v(a_0, a) [\alpha_2(a) - \alpha_{12}(a)] e^{A_{12}(a_0, a)}}{1 - \int_{a_0}^a e^{A_{01}(a_0, u)} P_v(a_0, u) [\alpha_2(u) - \alpha_{12}(u)] e^{A_{12}(a_0, u)} du}$$

– **Modèle M2 :**

La mortalité des malades dépend de l'âge et du temps calendaire, dans ce cas α_{12} est une quantité inconnue que nous remplaçons par son expression (définie par l'équation 3.2) dans l'équation 3.4

$$\alpha_{02}(a) = e^{A_{01}(a_0, a) + A_{02}(a_0, a)} P_v(a_0, a) [\alpha_2(a) - \alpha_{02}(a)g(a)] + \alpha_{02}(a)g(a)$$

nous nous ramenons à l'équation différentielle :

$$\alpha_{02}(a) = \frac{P_v(a_0, a)\alpha_2(a)}{e^{-A_{01}(a_0, a) - A_{02}(a_0, a)} [1 - g(a)] + g(a)P_v(a_0, a)}$$

qui se trouve être de la forme

$$y'(s) = \frac{a(s)}{e^{-y(s)} + c(s)}$$

où $y(s) = A_{02}(a_0, a)$, $y'(s) = \alpha_{02}(a)$,

$$c(s) = \frac{P_v(a_0, a)g(a)}{e^{-A_{01}(a_0, a)}[1 - g(a)]}$$

$$a(s) = \frac{P_v(a_0, a)\alpha_2(a)}{e^{-A_{01}(a_0, a)}[1 - g(a)]}$$

La résolution de cette équation différentielle n'a pas de solution analytique, c'est pourquoi nous utilisons une méthode numérique de Runge-Kutta d'ordre 4 (détaillée en annexe A.2.1).

3.3 Modèle Illness death et hypothèses semi-Makoviennes

3.3.1 Modèles

Dans le cas d'un modèle semi-Markovien homogène la transition de l'état "Malade" à l'état "décédé" ne dépend que de la durée passée avec la maladie. C'est l'hypothèse la plus simple que nous pouvons faire, mais de la même manière que pour le modèle Markovien, il est souvent trop restrictif de considérer une homogénéité au cours du temps des intensités de transition.

En effet, si le taux de mortalité des malades ne dépend que de la durée passée dans la maladie en négligeant l'effet de l'âge et du temps calendaire, alors par exemple :

- Deux sujets malades depuis une même durée auraient des taux de mortalité semblables, même si l'un à 20 ans de moins.
- Deux sujets malades depuis une même durée auraient des taux de mortalité semblable, même si l'un vivait en 1930 et l'autre en 2010.

C'est pourquoi, le modèle semi-Markovien développé (figure 3.2) est non-homogène, et la mortalité des malades dépend de la durée passée avec la maladie ainsi que de l'âge au temps t :

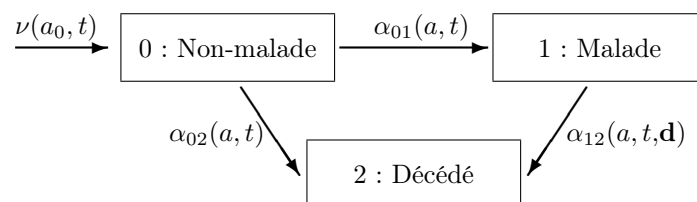


FIGURE 3.2 – Le modèle illness-death semi-Markovien.

La description des paramètres et les hypothèses que nous avons formulées pour le modèle Markovien (en section 3.2.1) sont les mêmes pour le modèle semi-Markovien.

La seule différence se trouve dans la définition de l'intensité de transition de l'état malade vers l'état décédé, qui dépend de l'âge a au temps t et de la durée passée avec la maladie d .

3.3.2 Méthodologie pour l'estimation des paramètres

Les données démographiques fournissent les valeurs passées et futures de α_2 et ν . Les données de cohorte, elles, permettent d'estimer les intensités de transition. Nous avons adaptés la méthode d'estimation semi-paramétrique avec vraisemblance pénalisée et approximation par fonctions splines à ce modèle.

Pour les mêmes raisons que celles énoncées pour le modèle Markovien, nous préférons estimer la mortalité des non-malades α_{02} en fonction de la mortalité générale α_2 , ce qui nécessite encore une fois de résoudre une équation différentielle pour estimer $\alpha_{02}(a, t)$, que nous détaillons dans la section 3.3.3.

Remarque : il est toujours possible de faire l'hypothèse que $\alpha_{02}(a, t) \neq \alpha_2(a, t)$ ou bien que $\alpha_{02}(a, t) = \alpha_2(a, t)$ en fonction de l'épidémiologie de la maladie.

De même, concernant l'incidence de la maladie, si aucune hypothèse d'évolution au cours du temps n'est avérée alors nous l'estimons simplement en fonction de l'âge sur les données de cohorte.

Pour la mortalité des malades, nous avons tout d'abord développé le modèle simple où elle en dépend que de la durée passée avec la maladie : $\alpha_{12}(d)$. Puis, conscient que cette hypothèse est trop restrictive, nous avons développé et proposé un modèle additif pour prendre en compte à la fois l'âge et le temps calendaire (et indirectement une évolution au cours du temps à travers la mortalité des non-malades). Pour cela nous l'écrivons comme la mortalité des non-malades à laquelle nous ajoutons un sur-risque de décès pour les malades :

$$\alpha_{12}(a, d) = \alpha_{02}(a, t) + \alpha^d(d) \quad (3.5)$$

Avec α^d le sur-risque de décès des malades par rapport aux non-malades à estimer directement sur les données de cohorte, ou à déterminer via la littérature, et α_{02} issu de la résolution de l'équation différentielle.

Avant de proposer ce modèle, nous en avons testé beaucoup d'autres : des modèles multiplicatifs, ou d'autres additifs, ... Puis, notre choix s'est porté sur ce modèle additif qui nous semblait plus intuitif, plus simple, plus facilement interprétable et qui permettait de trouver une solution analytique à la résolution de l'équation différentielle (ce choix sera discuté dans la section discussion (7.2.2)).

Remarque : par la suite, nous distinguerons ces deux modèles semi-Markoviens en les nommant ainsi

- **S1** si la mortalité des malades s'exprime seulement en fonction de la durée passée avec la maladie $\alpha_{12}(d)$,
- **S2** si la mortalité des malades s'exprime sous la forme de l'équation 3.5

3.3.3 Équations différentielles

Ici encore, nous omettons la dépendance au temps calendaire t dans les équations qui vont suivre, pour alléger les notations, et la résolution de l'équation différentielle, pour déterminer α_{02} , est effectuée pour chaque année de naissance ($t - a$ avec $a > a_0$).

L'évolution au cours du temps de la mortalité des non-malades α_{02} peut être prise en compte en impliquant les informations des données démographiques.

Comme pour la première résolution d'équation différentielle, nous réécrivons α_2 , mais cette fois si l'expression du risque de décès chez les malades est différente puisqu'il dépend également de la durée d :

$$\alpha_2(a) = \frac{P_{00}(a_0, a)\alpha_{02}(a) + P_{02}^1(a_0, a)}{P_v(a_0, a)}$$

Ce qui permet d'écrire la mortalité des non-malades ainsi :

$$\alpha_{02}(a) = \frac{\alpha_2(a)P_v(a_0, a) - P_{02}^1(a_0, a)}{P_{00}(a_0, a)} \quad (3.6)$$

Cette formulation est différente de celle que nous avons utilisée dans le cas des modèles Markoviens (équation 3.3), ceci vient du fait que la mortalité des déments fait intervenir la durée passée avec la maladie.

Ici, nous développons la résolution de l'équation différentielle seulement dans le cas où la mortalité des malades dépend de l'âge au temps t et de la durée passée avec la maladie (modèle **S2**). Alors α_{12} est une quantité inconnue que nous exprimons par l'équation 3.5.

Dans ce cas, l'intensité de transition cumulée s'écrit :

$$A_{12}(a_0, a, d) = A_{02}(a_0, a) + A^d(0, a - a_0)$$

et donc, P_{02}^1 s'écrit :

$$P_{02}^1(a_0, a) = \int_{a_0}^a e^{-A_{01}(a_0, u) - A_{02}(a_0, u)} \alpha_{01}(u) e^{-[A_{02}(u, a) + A^d(0, a - u)]} [\alpha_{02}(a) + \alpha^d(a - u)] du$$

A présent, il suffit de remplacer P_{00} , P_{02}^1 et P_v par leur écriture mathématique dans l'équation 3.6, de simplifier, puis d'isoler α_{02} , pour obtenir l'équation suivante :

$$\alpha_{02}(a) = \alpha_2(a) - e^{-A_{02}(a_0, a)} \frac{\int_{a_0}^a e^{-A_{01}(a_0, u)} \alpha_{01}(u) e^{-A^d(0, a - u)} \alpha^d(a - u) du}{P_v(a_0, a)}$$

Ainsi, pour estimer α_{02} , il faut résoudre l'équation différentielle :

$$y'(s) = c(s) - e^{-y(s)}a(s)$$

avec $y'(s) = \alpha_{02}(a)$, $y(s) = A_{02}(a_0, a)$, $c(s) = \alpha_2(a)$ et

$$a(s) = \frac{\int_{a_0}^a e^{-A_{01}(a_0, u)} \alpha_{01}(u) e^{-A^d(0, a-u)} \alpha^d(a-u) du}{P_v(a_0, a)}$$

Cette équation différentielle a une solution analytique que nous résolvons à l'aide d'un changement de variable et de la méthode de la variation de la constante (voir l'annexe A.2.2). La solution finale est donc exprimée qu'avec des quantités connues :

$$\begin{aligned} \alpha_{02}(a) &= \alpha_2(a) \\ &- \frac{e^{-A_2(a_0, a)} \int_{a_0}^a e^{-A_{01}(a_0, u)} \alpha_{01}(u) e^{-A^d(0, a-u)} \alpha^d(a-u) du}{P_v(a_0, a)} \\ &- \frac{1 - \int_{a_0}^a \frac{\int_{a_0}^u e^{-A_{01}(a_0, v)} \alpha_{01}(v) e^{-A^d(0, u-v)} \alpha^d(u-v) dv}{P_v(a_0, u)} e^{-A_2(a_0, u)} du}{P_v(a_0, a)} \end{aligned}$$

3.4 Projections

Une fois que le choix du modèle à appliquer est fixé et que les paramètres ou fonctions sont estimés, ceux-ci vont servir à faire des projections de divers indicateurs de santé. C'est également dans cette phase-là qu'il est essentiel d'utiliser les projections nationales, qui concernent la mortalité générale dans la population et la taille de la population à risque.

La méthode pour faire des projections est la suivante :

Tout d'abord, il faut fixer une année t pour laquelle nous voulons donner des projections. En fonction de t et sachant que les sujets sont à risque à partir de l'âge a_0 nous pouvons en déduire les années de naissance qui vont intervenir dans le calcul des indicateurs de santé au temps t . En effet, pour tous les âges (au temps t) supérieurs ou égaux à a_0 nous calculons l'année de naissance des sujets concernés en faisant $t - age$.

Puis, les estimations sont faites pour tous les âges supérieurs à a_0 et pour chaque année de naissance concernée par l'année de projection t .

Remarque 1 : il est évident que si une fonction ne dépend que de l'âge, nous n'avons pas besoin de faire des estimations pour chaque année de naissance puisqu'elles seront inchangées.

Enfin, ces estimations sont celles que nous utilisons dans nos formules pour calculer les indicateurs de santé présentés dans la section 2.3.

Remarque 2 : si nous souhaitons connaître la projection d'un indicateur de santé pour un âge donné (par exemple, l'espérance de vie à 75 ans) alors seulement une année de naissance est nécessaire.

Pour récapituler, voici de manière plus concise les différentes étapes de la méthode de projection :

- 1. Choisir le modèle, Markov/semi-Markov, et les hypothèses sur les fonctions d'intensités de transition.
- 2. Fixer t l'année de projection et en déduire les années de naissance $t - a$ (avec $a_0 \leq a \leq 100$).
- 3. En fonction du modèle choisi dans étape (1.), pour tous âges a et pour chaque année de naissance déterminée dans l'étape (2.) :
 - 3.a Estimer, *via* les données de cohorte et/ou de la littérature, l'incidence de la maladie, α_{01} , et soit le risque relatif, g , soit le sur-risque, α^d .
 - 3.b Utiliser les estimations de l'étape 3.a et les données de populations pour résoudre l'équation différentielle et en déduire les taux de mortalité des malades, α_{12} , et des non-malades, α_{02} .
- 4. Utiliser les estimations faites dans l'étape (3.) dans les formules permettant de calculer les indicateurs de santé.

Concernant les intervalles de confiance pour les projections, ils sont calculés comme décrit dans la section 2.4, en prenant en compte une incertitude pour les paramètres estimés à même les données (généralement données de cohortes). C'est-à-dire, une incertitude sur l'incidence α_{01} , puis, selon les modèles, une incertitude sur $\alpha_{12}(a)$ (modèle **M1**), $g(a)$ (modèle **M2**), $\alpha_{12}(d)$ (modèle **S1**) ou $\alpha^d(d)$ (modèle **S2**).

Il faut noter que nous ne prenons pas en compte de variation/d'incertitude dans les projections des données nationales démographiques même s'il nous est possible de le faire. En effet, l'INSEE propose différents scénarios pour les projections de ces fonctions, par exemple, en distinguant le scénario bas, du scénario central et du scénario haut. Ainsi, nous pourrions faire des projections sous ces différents scénarios ce qui nous donnerait une variation supplémentaire dans nos projections, cependant, pour ce travail, nous avons choisi de rester sous le scénario central de l'INSEE.

Étude de l'impact d'interventions

Sommaire

4.1	Introduction	52
4.2	Prise en compte de l'effet d'un facteur de risque	53
4.3	Schéma d'intervention	53
4.3.1	Description	53
4.3.2	Paramètres pour l'intervention	54
4.3.3	Scénarios d'intervention	55
4.3.4	Conséquences	56
4.4	Adaptation du modèle.	56
4.4.1	Estimations et projections	56
4.4.2	Indicateurs de santé	58

4.1 Introduction

Les maladies chroniques sur lesquelles nous travaillons ne connaissent pas encore de traitements capables de les guérir. Une fois que la maladie est avérée, il n'y a pas de retour possible à un état "non-malade". Il est donc difficile d'imaginer comment agir directement sur la maladie même. En revanche, beaucoup d'études sont menées pour connaître les facteurs de risque de ces maladies et étudier leur impact sur l'apparition ou l'évolution de celles-ci.

La stratégie serait donc d'envisager d'agir sur ces facteurs de risque pour indirectement agir sur la maladie.

Nous pouvons donner en exemple quelques études qui ont cherché de différentes manières à évaluer l'effet d'une intervention. Brookmeyer *et al.* [Brookmeyer 2007] ont étudié l'impact d'une intervention retardant le début de la démence d'un ou deux ans et/ou l'évolution de la maladie en modifiant les intensités de transition. Pour modéliser l'impact de l'intervention, ils multiplient les intensités de transition par plusieurs risques-relatifs et en déduisent le retard moyen du début de la maladie et de sa progression, mais sans considérer les changements sur le décès.

Barnes et Yaffe [Barnes 2011] quant à eux, ont utilisé les risques attribuables (pourcentage de cas attribuable à un facteur donné) pour calculer le nombre de cas qui pourraient être prévenus en réduisant la prévalence de facteurs de risque dans la population.

Cependant, ces études ne considèrent pas le fait que modifier un facteur de risque d'une maladie impacte aussi la mortalité des "non-malades", puisque la plupart des facteurs de risque de ces maladies chroniques sont également des facteurs de risque de décès. Donc, si nous modifions les effets du facteur de risque, il faut modifier ces effets sur la maladie et sur le décès.

Les premiers travaux menés par l'équipe pour mesurer l'impact d'interventions utilisent un modèle Markovien non-homogène et prennent en compte l'impact d'une intervention en considérant ses effets à la fois sur le décès et la survenue de la maladie [Joly 2013], [Jacqmin-Gadda 2013]. Cependant, leur intervention cible seulement les individus âgés de 65 ans en 2010 et n'étudie l'impact que sur les projections de prévalence de la maladie.

Notre objectif était donc d'améliorer la méthode initiale pour pouvoir considérer des interventions plus générales et donner des projections de différents indicateurs de santé.

Dans les sections suivantes, nous détaillons la méthode pour prendre en compte l'impact d'une intervention dans le modèle de projection. Puis, nous définissons les interventions que nous pouvons faire. Enfin, nous précisons les adaptations qui ont été apportées à notre méthode (décrite dans le chapitre 3) pour l'estimation, les projections ainsi que pour le calcul des indicateurs de santé, dans le cadre d'une

intervention.

4.2 Prise en compte de l'effet d'un facteur de risque

Nous tenons compte d'une modification d'un facteur de risque due, par exemple, à une campagne de prévention, en introduisant une variable explicative binaire z qui prend la valeur 1 si l'individu est exposé au facteur de risque et 0 sinon. Ce facteur de risque peut avoir un effet différent sur chaque transition.

Pour modéliser l'effet de ce facteur, nous utilisons un modèle à intensités de transition proportionnelles :

$$\alpha_{ij}(a|z) = \alpha_{ij}^0(a)\theta_{ij}^z \quad (4.1)$$

α_{ij}^0 étant l'intensité de transition de base entre l'état i et l'état j ($i = 0, 1, j = 1, 2, i \neq j$), et θ_{ij} le risque relatif associé au facteur de risque z pour la transition i, j . La prévalence du facteur de risque dans la population à un âge donné (soit a_0) est supposée connue avant l'intervention et est notée $p(a_0)$.

Les valeurs initiales des θ_{ij} et de la prévalence du facteur de risque sont généralement déterminées à l'aide de la littérature, elles sont donc supposées connues avant l'intervention.

En conséquence, les intensités de transition cumulées s'écrivent également avec un conditionnement sur z :

$$A_{ij}(a_0, a|z) = \int_{a_0}^a \alpha_{ij}(u|z) du$$

Lors d'une intervention, nous considérons que la prévalence et/ou les risques relatifs du facteur de risque peuvent être modifiés. Ces modifications entraînent un changement du risque de devenir malade et/ou de décéder des individus exposés à ce facteur. Nous supposons donc que ce facteur de risque est un facteur causal de la maladie [Commenges 2015].

4.3 Schéma d'intervention

4.3.1 Description

Mettre en place une intervention ciblant un facteur de risque de maladie chronique est actuellement une des interventions préventives les plus réalistes.

En effet, il n'y pas de traitements qui guérissent les maladies sur lesquels notre intérêt se porte et les causes sont peu connues, alors qu'il existe des traitements efficaces ainsi que des stratégies de prévention pour beaucoup de facteurs de risque connus de ces maladies.

Nous envisageons donc une campagne d'intervention mise en place à partir d'un temps τ choisi, sur tous les individus ou bien seulement sur une tranche d'âge précise. Cette campagne d'intervention ciblerait un facteur de risque de la maladie, ou des facteurs de risques combinés, et aurait donc un impact sur la maladie, mais également sur le risque de décès des individus exposés au(x) facteur(s) de risque, qu'ils soient malades ou non-malades.

4.3.2 Paramètres pour l'intervention

Nous faisons l'hypothèse que l'intervention est mise en place au temps τ et ses effets sont maintenus dans le temps. Elle peut modifier la prévalence p du facteur de risque et/ou les risques relatifs θ_{ij} associés au facteur de risque. Ces deux paramètres sont supposés connus avant l'intervention.

De plus, nous supposons que les sujets non-exposés au facteur de risque à a_0 restent non-exposés pour le reste de leur vie. En d'autre terme, nous considérons soit que le facteur de risque ne peut pas se développer après a_0 ans, soit que l'acquisition du facteur de risque après cet âge ne modifie ni le risque de décès ni le risque de développer la maladie. Ainsi, les exposés sont définis comme les sujets ayant le facteur de risque à l'âge a_0 et de ce fait, la prévalence connue du facteur de risque est définie pour cet âge, d'où $p(a_0)$.

Si l'intervention cible un groupe d'âge, nous supposons que les individus trop jeunes au moment de l'intervention seront impactés par ses effets lorsqu'ils entreront dans la catégorie d'âge fixée.

Une telle intervention nécessite de distinguer les individus non-malades des individus malades au temps τ , car pour les sujets exposés au facteur de risque et malades au temps τ seul le risque de décès sera impacté, alors que pour les sujets exposés et non-malades, le risque de développer la maladie est lui aussi modifié. Pour les sujets non-exposés au facteur de risque, les risques de décéder ou de développer la maladie seront inchangés par l'intervention.

De nouveaux paramètres sont introduits dans le calcul des probabilités.

La prévalence $p(a)$ du facteur de risque à l'âge a est scindée en $p_0(a)$ et $p_1(a)$ qui sont respectivement la prévalence du facteur de risque chez les non-malades et chez les malades à l'âge a .

Lorsque nous ajoutons le signe $*$ aux paramètres définissant la prévalence et les risques relatif associés au facteur de risque, nous signifions qu'il s'agit des valeurs de ces paramètres après l'intervention (elles peuvent être inchangées).

De ce fait, p^* est la modification de la prévalence du facteur de risque et θ_{ij}^* la modification des risques relatifs associés au facteur de risque, après l'intervention au temps τ .

Enfin :

- $\omega(a)$ représente la proportion de sujets d'âge a , exposés au facteur de risque, qui restent exposés après l'intervention,

– $1 - \omega(a)$, le complémentaire, est la proportion de sujets exposés au facteur de risque, qui perdent leur exposition après l'intervention.
 Alors, $\omega(a) = \frac{p^*(a)}{p(a^-)}$, sachant que $p(a^-)$ est la prévalence du facteur de risque à l'âge a juste avant l'intervention. La figure 4.1 décrit ce phénomène.

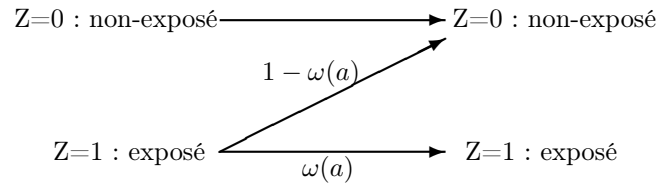


FIGURE 4.1 – Exposition avant l'intervention (sur la gauche) versus après l'intervention (sur la droite).

Nous pouvons remarquer que le groupe des non-exposés après l'intervention se compose de sujets qui n'ont jamais été exposé au facteur de risque et de sujets qui étaient exposés avant l'intervention, mais ne le sont plus après l'intervention. En effet, nous pouvons noter que le nombre de cas non exposés au facteur de risque après l'intervention $1 - p^*(a)$ s'écrit comme une somme : $1 - p^*(a) = 1 - \omega(a)p(a^-) = [1 - p(a^-)] + (1 - \omega(a))p(a^-)$ (soit la somme des non-exposés juste avant l'intervention et des exposés qui deviennent non-exposés après l'intervention).

Remarque : Si nous voulions étudier l'effet qu'aurait eu une intervention dans le passé (par exemple en 1980) sur les indicateurs de santé aujourd'hui, nous serions dans un cadre contrefactuel [Pearl 2009].

4.3.3 Scénarios d'intervention

L'intervention que nous mettons en place est assez flexible vis à vis de la population visée et de son impact sur les différents paramètres du modèle.

Il s'agit ensuite d'envisager des scénarios d'intervention. Il est intéressant d'avoir différents types de scénarios : des scénarios plausibles proposés par des médecins spécialistes de la maladie, de son évolution et des facteurs de risque, ou des scénarios qui ressortent de la littérature ; mais aussi des scénarios extrêmes, afin d'envisager les variations maximum de nos projections d'indicateurs de santé et de faire des analyses de sensibilité. Comme scénarios extrêmes, nous proposons que l'intervention diminue de moitié la prévalence du facteur de risque dans la population voir l'élimine complètement. Ou bien que l'intervention diminue de moitié ou élimine l'effet du facteur de risque sur le risque de développer la maladie θ_{01} .

De manière générale, nous construisons des scénarios extrêmes en parallèle des scénarios plausibles, dans le but de mieux comprendre les résultats des scénarios

plausibles, en illustrant la façon dont les différents paramètres interviennent.

4.3.4 Conséquences

Dans un premier temps, nous devons considérer le fait que la prévalence du facteur de risque, dans la population, peut être modifiée par l'intervention. Si cela est le cas, la proportion d'exposés et de non-exposés va changer entre les moments avant et après l'intervention. De plus, l'impact de l'intervention est différent selon le statut d'exposition au facteur de risque, d'où la nécessité de distinguer ces deux populations (exposés/non-exposés).

Ensuite, chaque intensité de transition peut être modifiée par le facteur de risque, *via* les risques relatifs θ_{ij} . Et comme nous l'avons décrit précédemment, les facteurs de risque qui nous intéressent sont très souvent à la fois facteurs de risque de la maladie et du décès, d'où la nécessité de distinguer au moment de l'intervention les sujets malades (dont le risque de décès θ_{12} peut être modifié) des sujets non-malades (dont le risque de décès θ_{02} et/ou le risque de développer la maladie θ_{01} peuvent être modifiés).

4.4 Adaptation du modèle.

L'utilisation des données démographiques pour estimer la mortalité générale dans la population, α_2 , et la taille de la population, ν , est inchangée.

En revanche, les paramètres (α_{ij}) du modèle que nous estimons (en ajustant le modèle illness-death semi-paramétrique par vraisemblance pénalisée et approximation par fonctions splines, puis en résolvant l'équation différentielle afin de considérer une évolution au cours du temps pour chaque mortalité) ne seront pas utilisés tels quels lors de l'intervention, mais serviront à estimer les intensités de transitions de base α_{ij}^0 . Dans ce cadre, ces paramètres α_{ij} sont alors appelés intensités de transition marginales.

4.4.1 Estimations et projections

Prendre en compte l'impact d'une intervention dans le modèle nécessite de modifier les étapes de l'estimation des paramètres et également la phase de projection. C'est pourquoi nous distinguons en générale le contexte "avant l'intervention" qui équivaut à un contexte "sans intervention" et le contexte "après l'intervention".

La démarche à suivre en fonction du contexte est détaillée dans les deux parties qui suivent.

4.4.1.1 Sans intervention

Étape (a) Estimations des paramètres du modèle.

Ceci est effectué sur les données de cohorte en ajustant le modèle illness-death semi-paramétrique avec vraisemblance pénalisée et approximation par

fonctions splines [Joly 2002] en utilisant le package R SmoothHazard. Cette méthode gère le risque semi-compétitif et la censure par intervalle et permet d'obtenir des fonctions continues sur l'âge (choix que nous avons fait).

Puis les données démographiques fournissent les paramètres de la population α_2 et $\nu(a_0, t)$ sur lesquelles nous faisons une approximation avec le modèle de Gompertz-Makeham pour obtenir également des fonctions continues sur l'âge.

Étape (b) Estimations de la mortalité des non-malades et des malades.

La mortalité des non-malades est estimée par résolution d'une équation différentielle analytique ou numérique (suivant le modèle illness-death utilisé, section 3.2.2 ou 3.3.2) impliquant des quantités connues ainsi que la mortalité générale dans la population α_2 obtenues à l'étape (a).

Puis la mortalité des malades est estimée en fonction du choix du modèle : **M1** ou **M2** pour les modèles Markoviens, **S1** ou **S2** pour les modèles semi-Markoviens.

Étape (c) Projections de divers indicateurs de santé.

Connaissant toutes les quantités d'intérêt après les étapes (a) et (b) les indicateurs de santé pertinents pour l'étude sont estimés pour n'importe quel temps t de projection, ce qui permet d'avoir un résumé du fardeau futur de la maladie (*via* les formules présentées dans la section 2.3).

4.4.1.2 Avec intervention

Les étapes (a) et (b) sont similaires, alors que l'étape (c) est remplacée par les étapes (c_1) et (c_2). L'étape (c_1) permet de considérer l'intervention et l'étape (c_2) évalue l'impact de l'intervention sur les projections.

Étape (c_1) Estimations des intensités de transition de base.

Estimer les intensités de transition de base α_{ij}^0 requiert de connaître les intensités de transition marginales α_{ij} (sans distinguer les exposés des non-exposés), les risques relatifs θ_{ij} ainsi que la prévalence du facteur de risque avant l'intervention. Les intensités de transition marginales α_{ij} sont données par les étapes (a) et (b), et les paramètres spécifiques du facteur de risque (θ_{ij} et p) peuvent être fournis par d'autres études, des experts, la littérature, ou imaginés pour les scénarios extrêmes par exemple.

Puis, il s'agit de résoudre un système de trois équations à trois fonctions inconnues, reliant les intensités α_{ij} aux intensités de transition spécifiques fonctions de l'exposition au facteur de risque, en faisant intervenir les risques relatifs et la prévalence associés à ce facteur de risque, et en se plaçant dans le contexte sans intervention.

Le système est le suivant :

$$\left\{ \begin{array}{l}
\alpha_{01}(a)P_{00}(a_0, a) = \\
\quad = p(a_0)\theta_{01}\alpha_{01}^0(a)e^{-\theta_{01}A_{01}^0(a_0, a) - \theta_{02}A_{02}^0(a_0, a)} \\
\quad \quad + (1 - p(a_0))\alpha_{01}^0(a)e^{-A_{01}^0(a_0, a) - A_{02}^0(a_0, a)} \\
\alpha_{02}(a)P_{00}(a_0, a) = \\
\quad = p(a_0)\theta_{02}\alpha_{02}^0(a)e^{-\theta_{01}A_{01}^0(a_0, a) - \theta_{02}A_{02}^0(a_0, a)} \\
\quad \quad + (1 - p(a_0))\alpha_{02}^0(a)e^{-A_{01}^0(a_0, a) - A_{02}^0(a_0, a)} \\
\alpha_{12}(a)P_{01}(a_0, a) = \\
\quad = p(a_0)\theta_{12}\alpha_{12}^0(a) \int_{a_0}^a e^{-\theta_{01}A_{01}^0(a_0, u) - \theta_{02}A_{02}^0(a_0, u)} \theta_{01}\alpha_{01}(u)e^{-\theta_{12}A_{12}^0(u, a)} du \\
\quad \quad + (1 - p(a_0))\alpha_{01}^0(a) \int_{a_0}^a e^{-A_{01}^0(a_0, u) - A_{02}^0(a_0, u)} \alpha_{01}(u)e^{-A_{12}^0(u, a)} du
\end{array} \right. \quad (4.2)$$

A_{ij}^0 étant les intensités de transition de base cumulées.

Ce système se résout en utilisant la méthode des moindres carrés et une approximation par fonction splines cubiques comme décrit dans l'article de Joly *et al.* [Joly 2013]

Enfin, les intensités de transition prenant en compte l'effet d'un facteur de risque $\alpha_{ij}(a|z)$ sont calculées *via* le modèle à intensités de transition proportionnelles décrit en section 4.2 :

$$\left\{ \begin{array}{l}
\alpha_{ij}(a|z = 1) = \alpha_{ij}^0\theta_{ij} \\
\alpha_{ij}(a|z = 0) = \alpha_{ij}^0
\end{array} \right. \quad (4.3)$$

Étape (c₂) Projections de divers indicateurs de santé après l'intervention.

Nous avons développé les formules utilisées dans l'étape (c) dans la section qui suit (4.4.2), en prenant en compte l'impact de l'intervention sur les sujets exposés au facteur de risque. Ainsi, nous donnons les projections d'indicateurs de santé pour la maladie au temps t , et nous pouvons évaluer l'impact d'une intervention mise en place à un temps τ ($\tau < t$) sur ces projections.

Remarque : Les étapes (b) à (c₂) sont programmées en Fortran.

4.4.2 Indicateurs de santé

Les intensités de transition et les intensités de transition cumulées s'écrivent maintenant conditionnellement au statut d'exposition au facteur de risque z . De ce fait, les probabilités de transition ainsi que le calcul des indicateurs de santé seront distingués chez les exposés et chez les non-exposés au facteur de risque. Puis, pour obtenir les estimations en population générale, il s'agit de faire une moyenne

pondérée par la proportion d'exposés et de non-exposés au facteur de risque à chaque âge a au temps t , tout en considérant les différents états malades/non-malades quand cela est nécessaire.

Pour simplifier les équations, nous écrivons encore une fois toutes les quantités seulement en fonction de l'âge a , mais elles sont également fonction du temps calendaire t et de l'année de naissance $t - a$.

Enfin, rappelons que pour pouvoir déterminer les prévalences du facteur de risque chez les non-malades et chez les malades à l'âge a (respectivement $p_0(a)$ et $p_1(a)$), nous devons connaître la prévalence du facteur de risque à l'âge a_0 , qui n'est autre que $p_0(a_0)$ puisque par définition tous les individus sont non-malades à l'âge a_0 (nous pouvons également faire dépendre cette prévalence de l'année de naissance $t - a$ sans problème).

4.4.2.1 Sans intervention

Dans un premier temps, nous présentons les indicateurs de santé en fonction du statut d'exposition, mais en restant dans le contexte sans intervention.

– *L'espérance de vie sans la maladie* LE_{00} à l'âge a s'écrit dans ce contexte :

$$LE_{00}(a) = p_0(a)LE_{00}(a|z = 1) + (1 - p_0(a))LE_{00}(a|z = 0)$$

Avec $p_0(a)$ la proportion de sujets exposés au facteur de risque d'âge a parmi les non-malades.

$$p_0(a) = \frac{p_0(a_0)P_{00}(a_0, a|z = 1)}{p_0(a_0)P_{00}(a_0, a|z = 1) + (1 - p_0(a_0))P_{00}(a_0, a|z = 0)} \quad (4.4)$$

Sachant également que $LE_{00}(a|z) = \int_a^\infty P_{00}(a, u|z)du$

et $P_{00}(a, u|z) = e^{-A_{01}(a, u|z) - A_{02}(a, u|z)}$

– *L'espérance de vie d'un non-malade* LE_0 à l'âge a s'écrit dans ce contexte :

$$LE_0(a) = p_0(a)LE_0(a|z = 1) + (1 - p_0(a))LE_0(a|z = 0)$$

Sachant que $LE_0(a|z) = \int_a^\infty (P_{00}(a, u|z) + P_{01}(a, u|z))du$

et $P_{01}(a, u|z) = \int_a^u e^{-A_{01}(a, v|z) - A_{02}(a, v|z)} \alpha_{01}(v|z) e^{-A_{12}(v, u|z)} dv$

– *L'espérance de vie d'un malade* LE_{11} à l'âge a s'écrit dans ce contexte :

$$LE_{11}(a) = p_1(a)LE_{11}(a|z = 1) + (1 - p_1(a))LE_{11}(a|z = 0)$$

Avec $p_1(a)$ la proportion de sujets exposés au facteur de risque d'âge a parmi les malades.

$$p_1(a) = \frac{p_0(a_0)P_{01}(a_0, a|z = 1)}{p_0(a_0)P_{01}(a_0, a|z = 1) + (1 - p_0(a_0))P_{01}(a_0, a|z = 0)} \quad (4.5)$$

Sachant également que $LE_{11}(a|z) = \int_a^\infty P_{11}(a, u|z) du$
 et $P_{11}(a, u|z) = e^{-A_{12}(a, u|z)}$

– *L'espérance de vie générale* à l'âge a s'écrit dans ce contexte :

$$LE_{..}(a) = \pi_0(a)LE_0(a) + (1 - \pi_0(a))LE_{11}(a)$$

où $\pi_0(a)$ est la proportion de non-malades d'âge a (au temps t) parmi les sujets vivants : $\pi_0(a) = \frac{P_{00}(a_0, a)}{P_v(a_0, a)}$

– *Le nombre moyen d'années passées avec la maladie* à l'âge a s'écrit dans ce contexte :

$$T_{11}(a) = p_0(a)T_{11}(a|z = 1) + (1 - p_0(a))T_{11}(a|z = 0)$$

Sachant que $T_{11}(a|z) = LE_0(a|z) - LE_{00}(a|z)$

– *La probabilité de développer la maladie* à l'âge a s'écrit dans ce contexte :

$$F_{01}(a) = p_0(a)F_{01}(a|z = 1) + (1 - p_0(a))F_{01}(a|z = 0)$$

Sachant que $F_{01}(a|z) = \int_a^\infty P_{00}(a, u|z)\alpha_{01}(u|z) du$

– *L'âge moyen de début de la maladie* s'écrit dans ce contexte :

$$AV_{01}(a_0) = p_0(a)AV_{01}(a_0|z = 1) + (1 - p_0(a))AV_{01}(a_0|z = 0)$$

Sachant que $AV_{01}(a_0|z) = \frac{\int_{a_0}^\infty uP_{00}(a_0, u|z)\alpha_{01}(u|z) du}{F_{01}(a_0|z)}$

– *La prévalence de la maladie* entre les âges a_0 et 100 ans s'écrit dans ce contexte :

$$Prev(a_0) = \sum_{z=0}^1 \sum_{k=0}^{100} \nu(a_0, t - 100 + k|z) P_{01}(a_0, 100 - k|z)$$

avec :

$$\begin{cases} \nu(a_0, t - 100 + k|z = 1) = p_0(a_0)\nu(a_0, t - 100 + k) \\ \nu(a_0, t - 100 + k|z = 0) = (1 - p_0(a_0))\nu(a_0, t - 100 + k) \end{cases} \quad (4.6)$$

4.4.2.2 Après intervention

En utilisant le modèle à intensités de transition proportionnelles et en considérant les nouvelles valeurs des paramètres associés au facteur de risque, p^* et θ_{ij}^* , nous pouvons fournir les projections des indicateurs de santé après l'intervention. Dans cette partie, nous détaillons les calculs des indicateurs de santé à l'âge a au temps t après l'intervention.

Les espérances de vie

Chaque espérance de vie s'évalue comme la moyenne de l'espérance de vie chez les exposés et les non-exposés, pondérée par la prévalence du facteur de risque à l'âge et au temps considérés.

La tranche d'âge ciblée par l'intervention peut être choisie. Par commodité et dans le but de proposer une intervention possible à mettre en place, nous supposons ici que l'intervention cible tous les sujets de plus de a_0 ans, âge à partir duquel les sujets sont à risque de développer la maladie pour une maladie donnée.

C'est pourquoi, il nous faut différencier les sujets de plus de a_0 ans des moins de a_0 ans au moment de l'intervention.

- Sujets de plus de a_0 ans au moment de l'intervention.

- LE_{00} à l'âge a :

$$LE_{00}(a|\tau) = p_0^*(a)LE_{00}(a|\tau, z = 1) + (1 - p_0^*(a))LE_{00}(a|\tau, z = 0)$$

$p_0^*(a)$ étant la proportion après l'intervention de sujets exposés au facteur de risque d'âge a parmi les non-malades.

$$p_0^*(a) = \frac{\omega p_0(a_0)P_{00}(a_0, a_\tau|z = 1)P_{00}(a_\tau, a|z = 1)}{P_{00}(a_0, a|\tau)} \quad (4.7)$$

avec a_τ l'âge au moment de l'intervention.

Sachant que $P_{00}(a_0, a|\tau)$ est la probabilité d'être vivant non-malade entre les âges a_0 et a après l'intervention :

$$\begin{aligned} P_{00}(a_0, a|\tau) &= \omega p_0(a_0)P_{00}(a_0, a_\tau|z = 1)P_{00}(a_\tau, a|z = 1) \\ &+ (1 - \omega)p_0(a_0)P_{00}(a_0, a_\tau|z = 1)P_{00}(a_\tau, a|z = 0) \\ &+ (1 - p_0(a_0))P_{00}(a_0, a|z = 0) \end{aligned}$$

Le premier terme correspond à la probabilité d'être exposé avant et après l'intervention et de rester en vie et non-malade entre les âges a_0 et a .

Le second terme correspond à la probabilité d'être exposé avant l'intervention et non-exposé après l'intervention et de rester en vie et non-malade entre les âges a_0 et a .

Le dernier terme correspond à la probabilité d'être non-exposé, vivant et non-malade entre les âges a_0 et a .

- LE_{11} à l'âge a :

$$LE_{11}(a|\tau) = p_1^*(a)LE_{11}(a|\tau, z = 1) + (1 - p_1^*(a))LE_{11}(a|\tau, z = 0)$$

$p_1^*(a)$ étant la proportion après l'intervention de sujets exposés au facteur de risque d'âge a parmi les malades. C'est la somme de $p_1^{*0}(a)$ la proportion de sujets non-malades à τ qui restent exposés et sont vivants malades à l'âge a , et de $p_1^{*1}(a)$ la

proportion de sujets malade à τ qui restent exposés et sont vivants malades à l'âge a :

$$p_1^*(a) = p_1^{*0}(a) + p_1^{*1}(a) \quad (4.8)$$

$$\text{avec } p_1^{*0}(a) = \frac{\omega p_0(a_0) P_{00}(a_0, a_\tau | z = 1) P_{01}(a_\tau, a | z = 1)}{P_{01}(a_0, a | \tau)},$$

$$\text{et } p_1^{*1}(a) = \frac{\omega p_0(a_0) P_{01}(a_0, a_\tau | z = 1) P_{11}(a_\tau, a | z = 1)}{P_{01}(a_0, a | \tau)}.$$

Sachant que $P_{01}(a_0, a | \tau)$ est la probabilité pour les non-malades et vivants à l'âge a_0 d'être malades et vivants à l'âge a après l'intervention.

$$\begin{aligned} P_{01}(a_0, a | \tau) &= \omega p_0(a_0) P_{00}(a_0, a_\tau | z = 1) P_{01}(a_\tau, a | z = 1) \\ &+ \omega p_0(a_0) P_{01}(a_0, a_\tau | z = 1) P_{11}(a_\tau, a | z = 1) \\ &+ (1 - \omega) p_0(a_0) P_{00}(a_0, a_\tau | z = 1) P_{01}(a_\tau, a | z = 0) \\ &+ (1 - \omega) p_0(a_0) P_{01}(a_0, a_\tau | z = 1) P_{11}(a_\tau, a | z = 0) \\ &+ (1 - p_0(a_0)) P_{01}(a_0, a | z = 0) \end{aligned}$$

Le premier terme correspond à la probabilité d'être exposé avant et après l'intervention, de devenir malade après l'intervention et d'être vivant à l'âge a .

Le second terme est la même chose sauf que le sujet est déjà malade au moment de l'intervention (il devient malade entre les âges a_0 et a_τ).

Le troisième terme correspond à la probabilité d'être exposé avant l'intervention et non-exposé après l'intervention, de devenir malade après l'intervention et d'être vivant à l'âge a .

Le quatrième terme est similaire au troisième sauf que le sujet est malade au moment de l'intervention.

Le dernier terme correspond à la probabilité d'être non-exposé, vivant et malade entre les âges a_0 et a .

- LE_0 , à l'âge a :

$$LE_0(a | \tau) = p_0^*(a) LE_0(a | \tau, z = 1) + (1 - p_0^*(a)) LE_0(a | \tau, z = 0)$$

- $LE_{..}$, à l'âge a :

$$LE_{..}(a | \tau) = \pi_0^*(a) LE_0(a | \tau, z) + (1 - \pi_0^*(a)) LE_{11}(a | \tau, z)$$

avec $\pi_0^*(a)$ la proportion de non-malades à l'âge a parmi les sujets vivants :

$$\pi_0^*(a) = \frac{P_{00}(a_0, a | \tau)}{P_{00}(a_0, a | \tau) + P_{01}(a_0, a | \tau)}$$

- **Sujets d'âge a_0 ou plus jeunes au moment de l'intervention.**

Par hypothèse, ces sujets sont non-malades au moment de l'intervention ($a_\tau \leq a_0$). Ils percevront les effets de l'intervention quand ils atteindront l'âge a_0 . Par conséquent, les probabilités P_{00} et P_{01} utilisées dans les formules précédentes sont simplifiées.

En remplaçant a_τ par a_0 , qui est l'âge à partir duquel les sujets sont soumis aux effets de l'intervention, nous obtenons :

$$P_{00}(a_0, a|\tau) = \omega p_0(a_0)P_{00}(a_0, a|z = 1) + (1 - \omega p_0(a_0))P_{00}(a_0, a|z = 0)$$

$$P_{01}(a_0, a|\tau) = \omega p_0(a_0)P_{01}(a_0, a|z = 1) + (1 - \omega p_0(a_0))P_{01}(a_0, a|z = 0)$$

Les autres indicateurs

Pour les autres indicateurs de santé, nous suivons la même démarche.

– T_{11} à l'âge a :

$$T_{11}(a) = p_0^*(a)T_{11}(a|\tau, z = 1) + (1 - p_0(a))T_{11}(a|\tau, z = 0)$$

– F_{01} à l'âge a :

$$F_{01}(a) = p_0^*(a)F_{01}(a|\tau, z = 1) + (1 - p_0^*(a))F_{01}(a|\tau, z = 0)$$

– AV_{01} :

$$AV_{01}(a_0) = p_0^*(a)AV_{01}(a_0|\tau, z = 1) + (1 - p_0^*(a))AV_{01}(a_0|\tau, z = 0)$$

– $Prev$ entre les âge a_0 et 100 :

$$Prev(a_0) = \sum_{z=0}^1 \sum_{k=0}^{100} \nu(a_0, t - 100 + k|z)P_{01}(a_0, 100 - k|\tau, z)$$

Rappelons que toutes les quantités définies dans cette partie sont dépendantes de l'âge au temps t (nous omettons le conditionnement sur le temps calendaire dans les notations), l'année de naissance des individus est aussi considérée.

Ainsi, tous ces calculs sont effectués séparément pour chaque année de naissance et il est facile de faire dépendre la prévalence ($p_0(a_0)$) et les risques relatifs associés au facteur de risque de l'année de naissance.

Application à la démence de type Alzheimer

Sommaire

5.1	Contexte et données	66
5.1.1	La démence : définition et état des lieux	66
5.1.2	Cohorte PAQUID	68
5.1.3	Données démographiques de l'INSEE	70
5.2	Modèles et hypothèses	71
5.2.1	Modèle Markovien	71
5.2.2	Modèle semi-Markovien	75
5.3	Intervention et scénarios	76
5.3.1	Facteurs de risque	76
5.3.2	Hypothèses	77
5.3.3	Scénarios d'intervention	77
5.4	Résultats	79
5.4.1	Estimations de l'incidence de la démence	79
5.4.2	Mortalité générale dans la population	80
5.4.3	Sous les modèles Markovien	81
5.4.4	Sous le modèle semi-Markovien	91
5.5	Choix du modèle	95
5.6	Conclusion sur l'application à la démence	97
5.6.1	Intervention	98
5.6.2	L'Hypothèse sur l'incidence de la démence	99
5.6.3	Les hypothèses sur la mortalité des déments et des non-déments	99

Notre modèle de projections peut être appliqué aux maladies chroniques pour lesquelles un retour à l'état non-malade n'est pas envisageable une fois que le sujet a développé la maladie. Il est nécessaire de disposer d'un certain nombre d'informations pour pouvoir appliquer le modèle, notamment les informations qui nous permettront d'estimer les intensités de transitions.

L'application à la démence type Alzheimer a été la base des principaux développements méthodologiques de cette thèse.

5.1 Contexte et données

5.1.1 La démence : définition et état des lieux

5.1.1.1 Définition et diagnostic

La démence ou syndrome démentiel se caractérise par une dégradation irréversible des fonctions cognitives telles que la mémoire ou le raisonnement, avec un retentissement sur l'aptitude à réaliser les activités quotidiennes menant à une perte d'autonomie et à un décès prématuré. Touchant principalement les personnes âgées, la distinction entre les personnes atteintes de démence et le vieillissement cognitif normal, n'est pas toujours évidente.

Dans le cadre de la recherche clinique, de l'épidémiologie et des essais thérapeutiques, les critères de la démence les plus utilisés sont ceux de l'American Psychiatric Association (APA) décrits dans le Diagnostic and Statistical Manual 4th edition (DSM IV) et ceux de l'Organisation Mondiale de la Santé (OMS) inscrits dans la Classification Internationale des Maladies 10^{ème} version (ICM-10). Ces deux manuels précisent que le diagnostic d'une démence nécessite la combinaison d'un trouble de la mémoire et de l'atteinte d'au moins une autre fonction cognitive, d'un degré qualifié de suffisant pour perturber les activités quotidiennes du sujet.

Cependant, le diagnostic est souvent tardif, et parfois même il ne sera jamais fait. La difficulté du diagnostic vient de la démence en elle-même, les causes étant trop peu connues, et les symptômes étant nombreux et difficilement évaluables.

Par ailleurs, les études s'intéressant à l'impact du mode de vie sur le risque de démence se développent et mettent l'accent sur les activités physiques, intellectuelles et sociales et sur le rôle de la nutrition [Berr 2007]. Ainsi, ces dernières années, il ressort des données disponibles divers facteurs de risque potentiels de la démence, notamment les facteurs de risque des maladies vasculaires, tels le diabète, l'hypertension, l'obésité, la consommation de tabac et l'inactivité physique d'après l'OMS.

Une meilleure connaissance des facteurs de risque de la démence serait un atout de taille pour la compréhension de cette maladie, sa survenue et son diagnostic, mais également pour envisager des possibilités de traitements.

Les conséquences de la démence sont donc principalement cognitives et physiques pour les sujets touchés, mais l'impact sur l'entourage du malade est important. En effet, la prise en charge du malade est lourde, c'est pourquoi la démence reste aujourd'hui un des principaux motifs d'institutionnalisation des personnes âgées.

Jusqu'à présent, aucun traitement ne permet de guérir la démence, mais les avancées médicales permettent un diagnostic plus précoce de la maladie, une meilleure prise en charge des patients, et donc allongent l'espérance de vie des sujets atteints de démence.

5.1.1.2 État des lieux de la démence en France

La situation démographique actuelle laisse supposer que le nombre de déments risque d'augmenter de façon proportionnelle au nombre de sujets âgés. Afin d'évaluer le nombre futur de déments, l'indicateur de morbidité couramment utilisé est la prévalence de la maladie (qui correspond ici au nombre de personnes démentes dans une population à un temps donné). Le nombre de déments en France avoisine en 2010 les 1 million de cas [Jacqmin-Gadda 2013], sans changement majeur et en considérant une incidence par âge constante au cours du temps, les projections pour 2030 avoisineraient 1,75 million [Joly 2013], [Wanneveich 2016].

D'autre part, le taux de prévalence de la maladie augmente avec l'âge et diffère selon le sexe. Des études menées sur les données de cohortes (PAQUID en 1999 et EURODEM en 2000) montrent que la prévalence est plus élevée chez les femmes après 80 ans [Ramaroson 2003], [Dartigues 1991], comme l'illustre la figure 5.1.

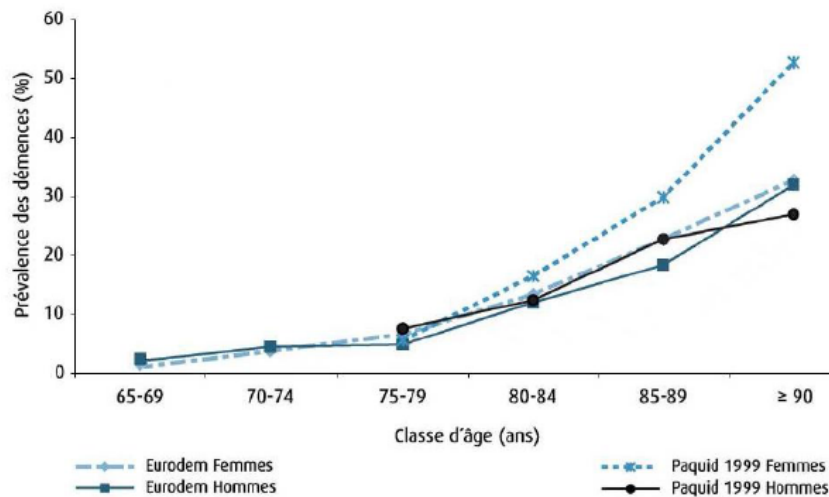


FIGURE 5.1 – Evolution de la prévalence de la démence selon l'âge, en fonction du sexe. Données PAQUID 1999 et EURODEM 2000.

Par ailleurs, un certain nombre de publications ont attesté du rôle de facteurs de risque modifiables sur le risque d'apparition d'une démence [Polidori 2012]. Ces facteurs sont donc des cibles naturelles d'interventions qui viseraient à réduire le risque de démence. En effet, l'hypertension artérielle est un facteur de risque bien connu de mortalité cardio-vasculaire et de plus en plus d'études l'associent également à un risque accru de démence [Tolppanen 2012], [Duron 2008]. Cependant, davantage d'études sont attendues dans ce domaine pour comprendre si une meilleure prise en charge de l'hypertension artérielle, mais aussi du diabète, voire des dyslipidémies (concentration anormale de lipide dans le sang) pourrait diminuer le risque de démence.

Actuellement, le temps moyen de survie est estimé à 5 ans [Berr 2007] pour les sujets déments dont l'âge moyen au début de la démence avoisine 82 ans. D'une

façon générale, les femmes atteintes par une démence ont une survie plus longue que les hommes déments.

Quelques études récentes commencent à entrevoir la possibilité d'une baisse de l'incidence de la démence au cours du temps [Schrijvers 2012], [Rocca 2011], mais les résultats sont encore incertains.

En définitive, les constats laissent entrevoir une augmentation du nombre de sujets à risque et donc du nombre futur de déments. En effet, le taux de prévalence de la démence augmente avec l'âge, et les limites du vieillissement général de la population sont constamment repoussées. Le plus souvent, les patients ont besoin de soins médicaux spécialisés, ce qui engendre un coût humain et économique significatif. C'est pourquoi, le poids socio-économique de la démence risque d'augmenter de façon importante.

Il s'avère donc essentiel de pouvoir prédire le "fardeau" futur de la démence aux travers de divers indicateurs de santé, à la fois pour prévoir les ressources nécessaires aux soins et à la prise en charge des futurs malades, mais également pour réfléchir à des interventions qui permettraient de diminuer la prévalence future des déments. Pour décider de ce type d'interventions, il sera nécessaire de cibler des facteurs de risque avérés de démence, et d'être capable d'évaluer l'impact qu'une intervention, sur un ou plusieurs de ces facteurs de risque, pourrait avoir sur l'évolution de la prévalence, mais aussi sur d'autres indicateurs de santé.

5.1.2 Cohorte PAQUID

La cohorte française PAQUID (Personnes Agées QUID) est une étude épidémiologique prospective menée en Aquitaine qui fut initiée en 1988 dans le but d'étudier le vieillissement cérébral de la population et plus particulièrement la démence. L'ensemble des travaux réalisés, depuis plus de 20 ans, a fourni des données essentielles sur l'épidémiologie et l'étiologie (étude des causes et des facteurs de risque) de la démence et de l'incapacité du sujet âgé en France. À l'inclusion dans l'étude les individus doivent avoir plus de 65 ans, être non-déments et non-institutionnalisés. Ils ont été sélectionnés par tirage au sort sur les listes électorales de 37 communes de Gironde et de 38 communes de Dordogne. Le taux initial de participation était de 68%.

Les participants ont été régulièrement suivis depuis l'entrée dans l'étude, avec une recherche active des cas de démence et un suivi de la mortalité. Les visites sont effectuées au domicile des participants et sont menées par une enquêtrice psychologue spécialement formée. De nombreuses variables ont été recueillies : les caractéristiques sociodémographiques, les habitudes de vie et activités pratiquées, l'entourage social, les aides professionnelles ou informelles et enfin des données concernant l'état de santé des sujets (avec l'évaluation notamment des consommations médicamenteuses régulières, de l'autonomie, de la symptomatologie dépressive, et des performances cognitives). Une recherche active des cas de démence était effectuée à chaque visite, appliquant les critères de diagnostic du

DSM-III-R (précédente version du DSM IV). Les sujets pour lesquels une démence était suspectée étaient ensuite examinés par un neurologue afin de confirmer le diagnostic. Les sujets ont été revus à plusieurs reprises à intervalle de 1 à 3 ans, soit 1 an après l'inclusion (seulement pour les sujets de Gironde), puis 3, 5, 8, 10, 13, 15, 17, 20 et 22 ans après la visite initiale. Une procédure d'actualisation du statut vital de chaque sujet participant à la cohorte était menée à chacun des suivis.

Le principal problème rencontré est dû à la censure par intervalle, car le diagnostic de la démence ne se fait qu'au moment des visites. Il est donc possible que des sujets deviennent déments puis décèdent entre deux visites.

L'échantillon se constituait de 3777 sujets de 65 ans ou plus, vivant dans le sud-ouest de la France. La figure 5.2 résume l'évolution des états de santé des sujets de la cohorte à la dernière mise à jour, qui correspond à 22 ans de suivis.

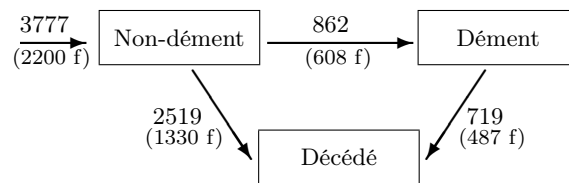


FIGURE 5.2 – Cohorte PAQUID après 22 ans de suivis (f=femmes).

Ainsi, parmi les 3777 sujets initialement inclus dans l'étude 2200 étaient des femmes, 862 cas incidents de démence ont été détectés (dont 608 chez les femmes) et 3238 sujets sont décédés dont 719 ayant été diagnostiqués déments. Les sujets encore en vie aujourd'hui sont toujours suivis dans le cadre de la cohorte.

Cet échantillon à l'inclusion est représentatif de la population française en terme d'âge et de sexe [Dartigues 1991] de l'époque.

À l'entrée dans l'étude, en 1988, l'âge moyen était de 75,46 ans (74,76 pour les hommes et 75,96 pour les femmes), pour une étendue comprise entre 65 et 101 ans. Par la suite, la concordance entre les taux spécifiques de décès par âge et sexe dans tout l'échantillon PAQUID a été comparée aux taux de mortalité générale de l'INSEE pour l'année 2000 (figure 5.3).

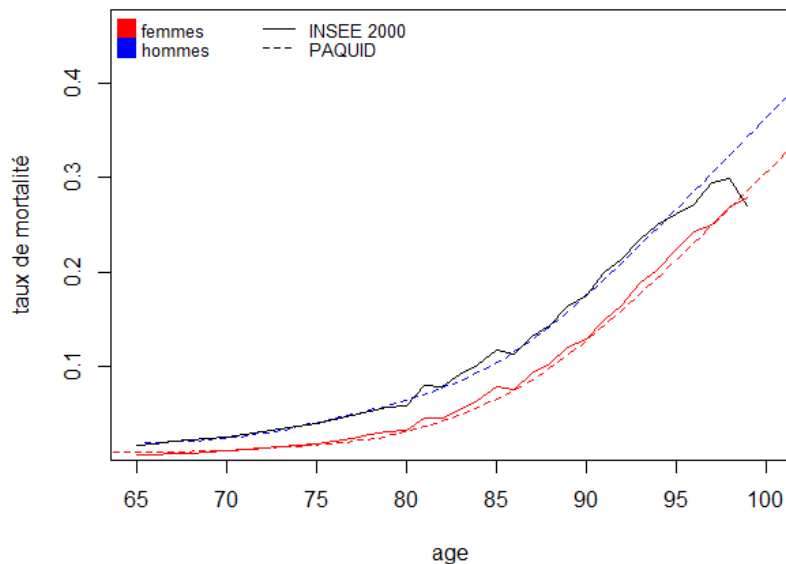


FIGURE 5.3 – Comparaison de la mortalité générale dans la population (INSEE en 2000) et de la mortalité globale dans la cohorte PAQUID, pour les hommes et les femmes.

La concordance entre les taux de mortalité de PAQUID et de l'INSEE atteste la représentativité de cette cohorte.

5.1.3 Données démographiques de l'INSEE

Nous utilisons les données démographiques de la France que nous pouvons récupérer *via* l'INSEE en libre accès sur leur site¹.

L'INSEE est une administration publique créée en 1946, qui joue un rôle majeur dans le service statistique public Français. Un des objectifs de l'INSEE est de produire des données de conjoncture, c'est-à-dire résumant la situation actuelle, comme des taux de croissance ou divers indices.

Les données de l'INSEE utilisées ici concernent les effectifs hommes et femmes avec les quotients de mortalité pour chaque âge et année de naissance avant 2007 (données de recensements), ainsi que les projections de population entre 2007 et 2060 (sous le scénario central) [Blanpain 2010]. Concernant les projections, nous disposons : du nombre annuel de décès par âge de 2007 à 2060, ainsi que du quotient de mortalité par âge de 2007 à 2060, et de la population vivante de 2007 à 2060 (pour les hommes, les femmes, et la population générale).

Ainsi, ces données permettent d'estimer le taux de mortalité générale α_2 et la population vivante ν par âge, sexe et année de naissance entre 1950 et 2060.

1. <http://www.insee.fr>

Les données de α_2 sont ensuite lissées en utilisant le modèle de Gompertz-Makeham [Gompertz 1825] sur l'âge pour avoir des taux de mortalités continus en fonction de l'âge, par sexe et par année de naissance.

5.2 Modèles et hypothèses

5.2.1 Modèle Markovien

Pour cette application, nous avons utilisé deux modèles Markoviens qui diffèrent de par leur façon de considérer la mortalité des déments α_{12} .

Cependant, certaines des hypothèses que nous faisons sont communes aux deux modèles. Nous allons d'abord développer ces hypothèses puis nous énoncerons les différences entre les deux modèles Markoviens.

Les hypothèses communes aux deux modèles Markoviens que nous mettons en place concernent l'incidence de la démence et la mortalité des non-déments.

Tout d'abord, concernant l'incidence de la démence, il est bien connu que la démence est rare avant 65 ans, et les taux d'incidence avant cet âge ne sont pas estimables de façon fiable. C'est pourquoi, sans perte de généralité, nous pouvons considérer que l'incidence est nulle avant 65 ans et donc fixer $a_0 = 65$ dans l'application.

Nous avons adapté la méthode pour qu'elle puisse prendre en compte un changement de l'incidence au cours du temps, mais peu d'hypothèses satisfaisantes existent actuellement pour la démence. Ainsi, le plus souvent dans les projections, nous la considérerons homogène au cours du temps, elle ne dépend donc que de l'âge :

$$\forall a > a_0, \alpha_{01}(a, t) = \alpha_{01}(a)$$

L'estimation de $\alpha_{01}(a)$ se fait sur les données de la cohorte avec la méthode de la vraisemblance pénalisée et l'approximation par des fonctions splines.

Enfin, concernant la mortalité des non-malades, il est préférable, dans le cas de la démence, de la distinguer de la mortalité générale dans la population, comme l'attestent les figures 5.4 et 5.5.

$$\alpha_{02}(a, t) \neq \alpha_2(a, t)$$

De ce fait, il faudra résoudre une équation différentielle.

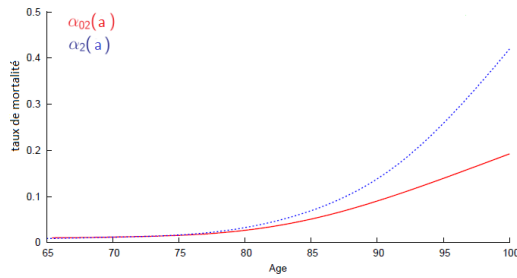


FIGURE 5.4 – Mortalité des non-déments α_{02} (PAQUID) et mortalité générale α_2 (INSEE 2000, lissée) par âge chez les femmes.

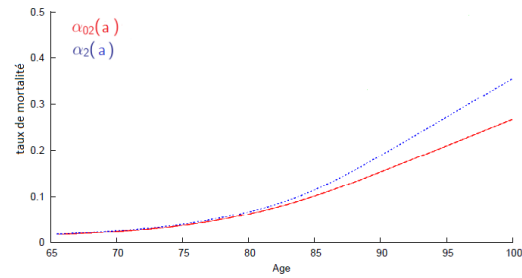


FIGURE 5.5 – Mortalité des non-déments α_{02} (PAQUID) et mortalité générale par âge α_2 (INSEE 2000, lissée) chez les hommes.

Sur ces figures, nous observons que la mortalité des non-déments estimée sur la cohorte est proche de la mortalité générale de la population française pour les âges jeunes. Puis, la différence entre ces deux estimations apparaît vers 80 ans pour les femmes et 85 ans pour les hommes et est d'autant plus importante que le sujet est âgé. Ceci est dû au taux de prévalence de la démence qui augmente en fonction de l'âge et à la sur-mortalité des déments.

Remarque : comme nous l'avons vu dans la figure 5.3 la mortalité globale dans la cohorte PAQUID équivaut à la mortalité générale dans la population (INSEE 2000), c'est pourquoi la constatation précédente est valable également si nous comparons la mortalité des non-déments à la mortalité globale de la cohorte.

La figure 5.6 représente le modèle multi-états Markovien à partir duquel nous allons distinguer deux cas en fonction de l'hypothèse considérée pour la mortalité des déments.

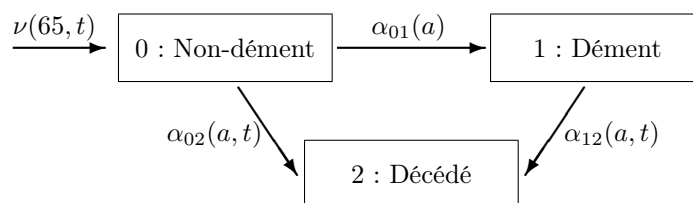


FIGURE 5.6 – Modèle illness-death Markovien.

Modèle M1 : mortalité des déments homogène selon t

Pour ce modèle, nous supposons que la mortalité des déments ne dépend que de l'âge des sujets. Nous faisons donc une hypothèse d'homogénéité pour cette transition :

$$\alpha_{12}(a, t) = \alpha_{12}(a)$$

De ce fait, la mortalité des déments α_{12} est estimée directement sur les données de cohorte par vraisemblance pénalisée et approximation par fonctions splines.

L'avantage de ce modèle est que l'équation différentielle à résoudre pour déterminer la mortalité des non-déments α_{02} se fait analytiquement, par contre il fait l'hypothèse forte que la mortalité des déments ne varie pas avec le temps. Ainsi, un dément de 70 ans en 1990 aurait le même risque de décéder qu'un dément de 70 ans en 2015, ce qui est fort improbable.

- Hypothèses alternatives testées sous le modèle M1

Nous avons testé le cas où la mortalité des non-déments n'est pas distinguée de la mortalité générale dans la population, pour apporter un élément de comparaison supplémentaire (en restant sous l'hypothèse $\alpha_{01}(a)$ concernant l'incidence de la démence) :

$$\alpha_{02}(a, t) = \alpha_2(a, t)$$

Il n'y a donc pas d'équation différentielle à résoudre sous cette hypothèse.

Ces hypothèses sont donc fortement contestables et ce ne sont pas celles que nous avons privilégiées. C'est pourquoi, avec ce modèle, nous ne présentons que des résultats de projections de prévalence afin de pouvoir faire des comparaisons et discuter les hypothèses avec les autres modèles.

Modèle M2 : mortalité des déments dépendante de t

Dans ce deuxième modèle, nous faisons l'hypothèse de non-homogénéité de la mortalité des déments. Ainsi, $\alpha_{12}(a, t)$ dépend à la fois de l'âge et du temps calendaire. Pour ce faire, nous utilisons l'équation 3.2 qui fait intervenir α_{02} et $g(a, t)$:

$$\alpha_{12}(a, t) = \alpha_{02}(a, t) \times g(a, t)$$

Cette formule sous-entend que l'évolution de la mortalité au cours du temps est la même parmi les déments et les non-déments, ce qui semble être le plus réaliste, et g représente la différence entre ces deux mortalités.

En ce qui concerne g , le risque relatif de décès des déments par rapport aux non-déments, nous l'estimons à partir des données de la cohorte (sur une base de fonctions splines) comme une fonction de l'âge, et nous assumons que ce risque est homogène au cours du temps : $g(a, t) = g(a)$.

Cette fois-ci, la résolution de l'équation différentielle pour déterminer la mortalité des non-déments α_{02} est nécessairement numérique, et permet ensuite d'estimer α_{12} (voir la section 3.2.3).

Ce modèle Markovien s'avère être le plus satisfaisant actuellement pour étudier l'évolution de la démence, c'est pour cela qu'il est celui avec lequel nous fournissons

le plus de projections. Nous donnons des résultats sur tous les indicateurs de santé que nous avons présentés, ainsi que des résultats avec ou sans intervention.

De plus, nous avons souhaité tester sous ce modèle **M2** d'autres hypothèses pour l'incidence de la démence afin d'enrichir la compréhension des phénomènes, la comparaison des modèles et des hypothèses, ainsi que la réflexion sur l'évolution possible du nombre futur de déments, notamment si l'incidence évolue.

- Hypothèses alternatives testées sous le modèle M2

La prévention de certains facteurs de risque de la démence, mais aussi l'évolution de la population en générale (par exemple, le niveau d'éducation ou d'avantage de sollicitation intellectuelle) pourrait faire varier l'incidence de la démence au cours du temps. De récentes études, menées par Schrijvers *et al.* [Schrijvers 2012], et Rocca *et al.* [Rocca 2011], constatent effectivement une tendance à la baisse de l'incidence de la démence. Une baisse de l'incidence pourrait être la conséquence d'un niveau d'études plus élevé, ou bien être attribuée à une meilleure prise en charge des facteurs de risque vasculaires au niveau de la population...

Bien que les résultats ne soient pas toujours significatifs et que l'interprétation soit encore incertaine, envisager une évolution de l'incidence en fonction de l'âge et du temps n'est pas dénuée d'intérêt.

C'est pourquoi, nous avons testé deux manières de prendre en compte une évolution de l'incidence dans le temps.

- La première hypothèse suppose que l'incidence évolue proportionnellement de génération en génération et résulte donc d'un "effet cohorte" (comme la conséquence de l'évolution du niveau d'éducation).

L'intensité de transition α_{01} est donc différente en fonction de l'année de naissance $t - a$. La méthode consiste à choisir une année de naissance de référence $(t - a)_0$ qui sera estimée sur PAQUID et d'en déduire les autres années de naissance $t - a$ en fonction, avec l'équation :

$$\alpha_{01}(a, t) = \alpha_{01}(a|(t - a)_0)f((t - a)) \quad (5.1)$$

- La seconde hypothèse suppose que l'incidence évolue proportionnellement d'année en année quelle que soit la génération, et résulte donc d'un "effet calendaire" (comme la conséquence de l'évolution ou de la prise en charge d'un facteur de risque de la démence). Donc, pour deux années consécutives, la variation (baisse ou hausse) appliquée à l'incidence est la même quel que soit l'âge des sujets.

Encore une fois, la méthode consiste à choisir une année calendaire de référence t_0 qui sera estimée sur PAQUID et d'en déduire les autres années t en fonction, avec l'équation :

$$\alpha_{01}(a, t) = \alpha_{01}(a|t_0)f(t) \quad (5.2)$$

Dans les deux cas, la fonction f modélise l'évolution de l'incidence. Celle-ci sera déterminée d'après la littérature ou en fonction d'hypothèses d'experts ou encore après un travail sur les données à disposition.

Quant au choix de $(t - a)_0$ et t_0 , il dépendra des données dont nous disposons.

Ainsi, les données de la cohorte PAQUID, nous permettent d'estimer $\alpha_{01}(a|t_0)$ ou $\alpha_{01}(a|(t - a)_0)$ et d'en déduire $\alpha_{01}(a, t)$ en fonction de l'hypothèse testée.

Remarque 1 : la prise en compte d'une évolution de l'incidence dans le temps, nous amène à poser des hypothèses quant aux autres fonctions du modèle (notamment α_2). Si l'incidence de la démence a varié, par exemple baissé, au cours des 20 dernières années, nous supposons que l'impact de cette variation sur la mortalité est pris en compte dans les taux de mortalité fournis par l'INSEE. C'est pourquoi nous ne modifierons pas la mortalité générale de la population donnée par l'INSEE.

Remarque 2 : considérer une évolution de l'incidence dans le temps de par un effet cohorte ne change en rien l'estimation de la mortalité des non-déments α_{02} (qui implique α_{01} , section 3.2) ; si l'évolution est calendaire il faudra recalculer l'incidence pour chaque année t .

Même remarque en ce qui concerne la méthode de projections présentée section 3.4.

5.2.2 Modèle semi-Markovien

Une étude a déjà démontré que l'utilisation d'un modèle semi-Markovien homogène **S1** n'était pas un modèle souhaitable pour représenter les données de la cohorte PAQUID [Commenges 2007]. C'est pourquoi, pour cette application, nous ne développons que le modèle semi-Markovien non-homogène **S2** qui considère que la mortalité des déments α_{12} est fonction de l'âge au temps t et de la durée passée avec la maladie.

Certaines hypothèses sont identiques à celles faites pour les modèles Markoviens. En effet, nous fixons toujours $a_0 = 65$ et la mortalité des non-malades est distinguée de la mortalité générale dans la population :

$$\alpha_{02}(a, t) \neq \alpha_2(a, t)$$

La résolution de l'équation différentielle pour déterminer cette mortalité est analytique (section 3.3.3). En revanche, nous restons sous l'hypothèse d'une incidence homogène dans le temps, soit $\alpha_{01}(a)$ estimée directement *via* les données de la cohorte.

La figure 5.7 représente le modèle multi-états semi-Markovien non-homogène où la mortalité des déments dépend de l'âge a , de la durée passée avec la démence d et implicitement du temps calendaire t .

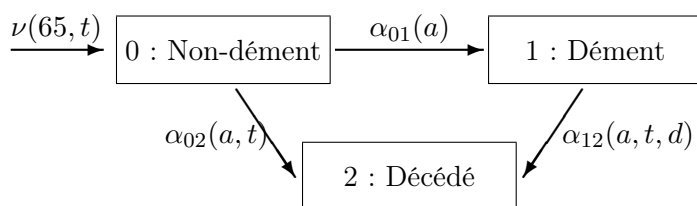


FIGURE 5.7 – Modèle illness-death semi-Markovien non-homogène.

Pour déterminer $\alpha_{12}(a, t, d)$ nous utilisons l'équation 3.5 qui fait intervenir $\alpha_{02}(a, t)$ (d'où la prise en compte du temps calendaire) et $\alpha^d(d)$.

$$\alpha_{12}(a, t, d) = \alpha_{02}(a, t) + \alpha^d(d)$$

Concernant le sur-risque de décès des déments par rapport aux non-déments, $\alpha^d(d)$, nous l'estimons directement en fonction de la durée passée avec la démence, sur les données de la cohorte à travers le modèle (par vraisemblance pénalisée et approximation par fonctions splines). Ensuite, le modèle nécessite de résoudre analytiquement une équation différentielle pour déterminer α_{02} puis, en conséquence, estimer α_{12} (voir la section 3.3.3).

Nous présentons ce modèle de projections sur un grand nombre d'indicateurs de santé que nous avons défini.

5.3 Intervention et scénarios

5.3.1 Facteurs de risque

Un exemple typique de facteur de risque est l'hypertension artérielle dite HBP (pour "High Blood Pressure"). L'hypertension influence la mortalité cardiovasculaire et ce facteur de risque a été trouvé associé à une augmentation du risque de démence [Tolppanen 2012], [Duron 2008]. Bien qu'encore controversés [Power 2011], ces résultats suggèrent qu'une intervention visant à réduire la prévalence de l'hypertension dans la population pourrait avoir un impact tant sur la mortalité que sur le risque de développer une démence.

Un meilleur contrôle de l'hypertension pourrait d'une part, réduire le risque de démence, mais d'autre part, augmenter l'espérance de vie, le nombre de personnes âgées et donc de sujets déments.

En conséquence, l'évaluation de l'impact d'une telle campagne sur la prévalence de la démence doit forcément prendre en compte la modification des taux de décès.

Un autre facteur de risque bien connu de la démence est l'apolipoprotéine E (ApoE4), qui est le principal facteur de risque génétique connu de l'Alzheimer et

qui est aussi un facteur de risque de la mortalité générale [Farrer 1997], [Genin 2011].

5.3.2 Hypothèses

Pour cette application, nous fixons $\tau = 2015$ (l'année de l'intervention). Nous supposons que l'intervention cible tous les individus de plus de 65 ans et que son effet est maintenu dans le temps. De ce fait, les individus de moins de 65 ans au moment de l'intervention percevront ces effets seulement lorsqu'ils atteindront 65 ans.

L'intervention a un effet soit sur la prévalence du facteur de risque, soit sur les risques relatifs des transitions entre chaque états pour les exposés par rapport aux non-exposés, ou les deux.

Sans autre hypothèse réaliste pour cette application, le ratio de la prévalence du facteur de risque avant et après l'intervention, noté $\omega(a) = \frac{p^*(a)}{p(a^-)}$, est supposé constant avec l'âge a , donc : $\omega(a) = \omega$.

Enfin, selon notre hypothèse décrite en section 4.3.2, les sujets non-exposés au facteur de risque à 65 restent non-exposés pour le reste de leur vie. Cette hypothèse qui peut sembler forte est en réalité très appropriée dans le cas de la démence, car la plupart des facteurs de risque de la démence sont dits à "mid-life", c'est à dire qu'ils sont facteurs de risque s'ils sont présents pendant la semi-vie du sujet (estimée entre 45 et 64 ans [Tolppanen 2012]).

5.3.3 Scénarios d'intervention

Nous avons repris certains des scénarios proposés par Jacqmin-Gadda *et al.* [Jacqmin-Gadda 2013] où l'intervention affectait seulement les sujets de 65 ans exactement au moment de l'intervention.

Ceci nous permettra de faire quelques comparaisons.

Les Scénarios 1 et 2 concernent une intervention qui cible l'HBP. Pour ces scénarios, nous supposons que le taux de prévalence de l'HBP à 65 ans est $p_0(65) = 0,40$ (40% de la population vivante âgée de 65 ans est exposée à l'HBP avant l'intervention).

Pour le scénario 1, nous supposons que dans le contexte général (*i.e* sans intervention) l'effet de l'hypertension artérielle est plus fort sur le décès ($\theta_{12} = \theta_{02} = 2$) que sur le risque de développer la démence ($\theta_{01} = 1,5$).

Nous distinguons ensuite les sous-scénarios (a,b) en fonction de l'impact de l'intervention. Dans le sous-scénario 1.a l'intervention réduit de moitié la prévalence de l'HBP ($\omega = 0,5$); alors que dans le sous-scénario 1.b, la prévalence de l'HBP est inchangée ($\omega = 1$) et (d'après la littérature sur l'efficacité de traitement anti-hypertenseur) nous considérons qu'un traitement de ce type pourrait réduire le risque relatif de la démence de 15% ($\theta_{01}^* = 1,27$) [Peters 2008], et le risque relatif

de décès de 13% ($\theta_{12}^* = \theta_{02}^* = 1,74$) parmi les sujets hypertendus [Staessen 2000].

Pour le scénario 2, avant l'intervention, l'effet de l'HBP est plus fort sur le risque de développer la démence ($\theta_{01} = 2$) que sur le risque de décéder ($\theta_{12} = \theta_{02} = 1,5$). Pour ce scénario, nous évaluons seulement l'impact d'une intervention qui réduit la prévalence de l'HBP : $\omega = 0,5$ pour le sous-scénario 2.a ; $\omega = 0$ pour le sous-scénario 2.b. Nous faisons donc l'hypothèse que la prise en charge va faire "disparaître" le facteur de risque chez 50% des sujets (scénario 2.a) voir 100% (scénario 2.b).

Le Scénario 3 concerne l'ApoE4 avec un hypothétique traitement diminuant le risque de démence parmi les sujets exposés. La prévalence de l'ApoE4 à 65 ans est $p_0(65) = 0,20$ [Winnock 2002]. Dans ce scénario, avant l'intervention, le risque relatif de démence des exposés est deux fois plus important que leurs risques relatifs de décès qu'ils soient malades ou non-malades ($\theta_{01} = 3$, et $\theta_{02} = \theta_{12} = 1,5$). L'intervention réduit le risque relatif de démence des exposés de moitié ($\theta_{01}^* = 1,5$).

5.4 Résultats

5.4.1 Estimations de l'incidence de la démence

Quel que soit le modèle choisi l'incidence de la démence est estimée en fonction de l'âge sur les données de la cohorte PAQUID. La figure 5.8 représente l'estimation de cette incidence, $\alpha_{01}(a)$, chez les hommes et chez les femmes de la cohorte, en distinguant le modèle Markovien du modèle semi-Markovien.

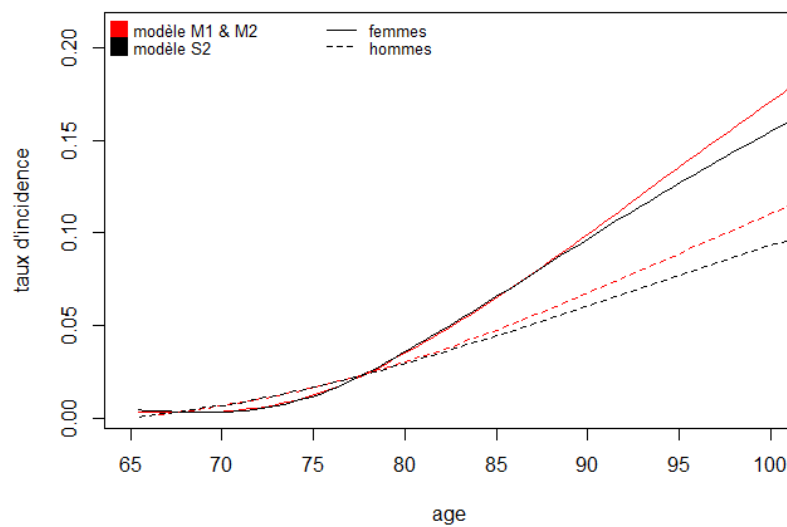


FIGURE 5.8 – Estimation de l'incidence de la démence par âge sur les données de la cohorte PAQUID en fonction du type de modèle appliqué, chez les hommes et chez les femmes.

Tout d'abord, nous observons que quel que soit le modèle l'incidence de la démence croit avec l'âge et est comparable chez les hommes et chez les femmes jusqu'à 80 ans, âge à partir duquel elle augmente beaucoup plus fortement chez les femmes.

Maintenant, si nous comparons les modèles Markoviens au modèle semi-Markovien, bien que la même méthode soit utilisée pour estimer cette incidence, nous observons des estimations plus faibles sous le modèle semi-Markovien (**S2**) pour les âges avancés, à la fois chez les femmes et chez les hommes (notez que pour le modèle **S1**, non présenté ici, les incidences estimées par âge sont encore plus faibles).

Les estimations commencent à vraiment différer à partir de 90 ans chez les femmes et 85 ans chez les hommes, ceci est en grande partie dû à la modélisation différente de la mortalité des sujets déments. En effet, nous n'avons pas les mêmes paramètres estimés en fonction de l'âge sur PAQUID, ce qui conduit à des résultats différents

pour l'estimation de l'incidence de ces modèles.

Enfin, nous constatons logiquement que l'incidence de la démence estimée se superposent dans les modèles Markoviens **M1** et **M2**. Ceci vient du fait que nous estimons exactement les mêmes paramètres, α_{01} , α_{02} et α_{12} , en fonction de l'âge sur PAQUID (mais la mortalité des déments $\alpha_{12}(a)$ estimée telle quelle ne sera pas utilisée dans le modèle **M2**).

5.4.2 Mortalité générale dans la population

La figure 5.9 décrit le taux de mortalité générale des femmes en France en fonction de leur âge et pour un large panel d'années. Les données sont issues de l'INSEE et nous les avons lissées à l'aide de la loi de Gompertz pour avoir des fonctions continues en fonction de l'âge. Les données passées (antérieures à 2007) viennent des recensements de population effectués par l'INSEE, tandis que les autres sont des projections de population faites par l'INSEE selon diverses hypothèses (nous pouvons avoir accès à des données allant jusqu'à 2060).

La représentation graphique de cette mortalité est indépendante du choix du modèle, ce sont ces mêmes taux de décès qui sont utilisés pour tous les modèles (en distinguant hommes et femmes).

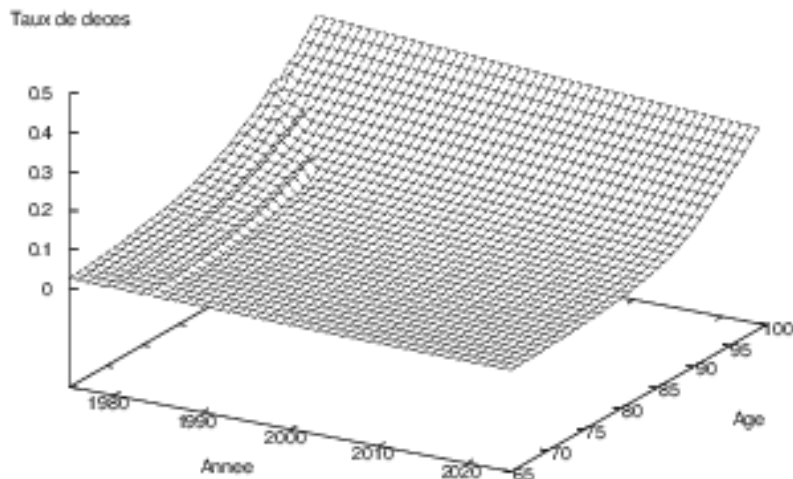


FIGURE 5.9 – Représentation 3D du taux de mortalité générale des femmes en France (INSEE, lissée), en fonction du temps et de l'âge.

Comme nous nous y attendions, le taux de mortalité générale augmente avec l'âge. Nous observons également que la mortalité générale diminue en fonction du temps calendaire, mais ceci est plus visible pour les âges avancés. Les mêmes observations ont été faites pour les hommes en France, mais les taux de mortalité sont systématiquement plus élevés à tous âges et pour toutes années.

5.4.3 Sous les modèles Markovien

Dans un premiers temps, nous développons les principaux résultats de l'application avec les modèles Markoviens.

5.4.3.1 Estimations

Sur les données de la cohorte, nous pouvons estimer les paramètres de mortalité du modèle en fonction de l'âge exclusivement. Nous avons déjà présenté la mortalité globale dans la cohorte PAQUID estimée en fonction de l'âge *via* la figure 5.3 ; ainsi que la mortalité des non-déments estimée sur la cohorte PAQUID en fonction de l'âge (figures 5.4 et 5.5). Nous retrouvons ces deux mortalités en annexe A.3.1 avec en plus la mortalité des déments estimée sur la cohorte PAQUID en fonction de l'âge.

Par ailleurs, la méthode de projection ne fait pas intervenir la mortalité des non-déments, $\alpha_{02}(a)$ estimée sur PAQUID, mais la mortalité générale de l'INSEE lissée par âge, $\alpha_2(a, t)$, puisque nous réestimons $\alpha_{02}(a, t)$ en fonction de α_2 pour capter également l'évolution au cours du temps. De même, la mortalité des déments, $\alpha_{12}(a)$, estimée sur PAQUID n'est utilisée telle quelle que dans le cas du modèle **M1**, car dans le modèle **M2** nous la ré-estimons pour prendre en compte une possible évolution au cours du temps.

Cependant, la mortalité des déments et des non-déments estimées directement en fonction de l'âge sur la cohorte, permettent d'estimer le risque relatif de décès des déments par rapport aux non-déments en fonction de l'âge (soit $g(a)$, dont nous avons besoin pour le modèle **M2**). Pour cela, nous faisons un rapport des taux en fonction de l'âge. La figure 5.10 représente l'estimation non-paramétrique de ce risque $g(a)$ sur la cohorte PAQUID.

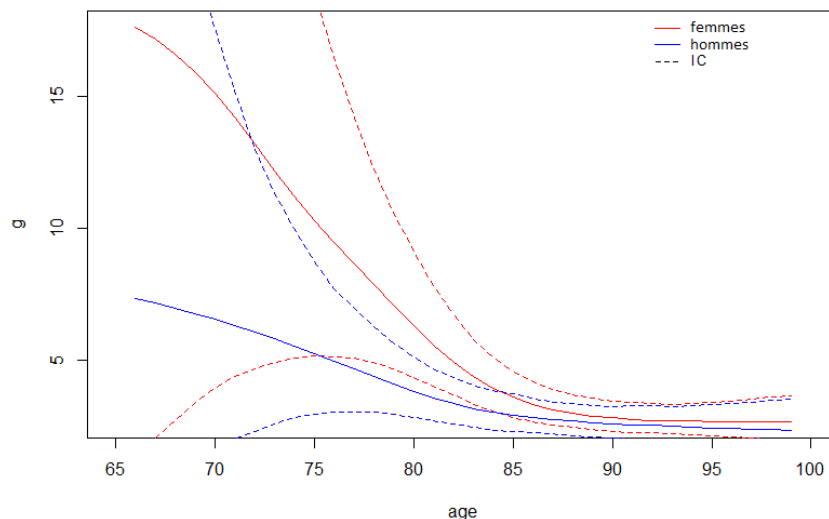


FIGURE 5.10 – Estimation du risque relatif de décès des déments par rapport aux non-déments en fonction de l'âge et du sexe sur la cohorte PAQUID, IC à 95% (modèle **M2**).

Ce paramètre, utilisé dans le modèle **M2**, permet de prendre en compte une évolution de la mortalité des déments au cours du temps. Les intervalles de confiance sont très larges pour les plus jeunes sujets à cause de la faible incidence de la démence et du peu de décès à ces âges. Mais, ces risques relatifs élevés (pour ces tranches d'âge) n'ont pas de réels impacts sur les projections des indicateurs de santé. Ceci est vérifié par analyse de sensibilité, en imposant des risques relatifs beaucoup plus faibles pour les classes jeunes et en comparant les projections et les intervalles de confiance obtenus.

5.4.3.2 Projections de prévalence en fonction des hypothèses sur l'incidence (modèle M2).

Sous l'hypothèse d'une incidence qui évolue au cours du temps de par un "effet cohorte", nous avons fixé la cohorte de référence à l'année de naissance 1913, car à l'entrée dans l'étude PAQUID, en 1988, les sujets avaient entre 65 et 101 ans, avec une moyenne d'âge de 75 ans. C'est pourquoi nous fixons la cohorte de référence à l'année 1913 (1988-75). Ainsi, nous considérons que l'intensité de transition α_{01} estimée par PAQUID est celle qui correspond aux individus nés en 1913.

Pour choisir la forme de la fonction f , nous nous sommes appuyés sur les résultats de deux études [Schrijvers 2012], [Rocca 2011] en testant des baisses simples : tout d'abord, un "effet cohorte" fort qui induit une baisse de l'incidence de la démence de 3% par année de naissance ; puis, un effet moindre d'une baisse de 1%.

Sous l'hypothèse d'une incidence qui évolue au cours du temps de par un "effet calendaire", nous avons fixé l'année de référence à l'année 2000, car les dernières mises à jour de la cohorte datent de 2012. La durée d'étude est donc comprise entre 1988 et 2012 d'où notre choix. Ainsi, nous supposons que l'intensité de transition α_{01} estimée par PAQUID est celle de l'année 2000.

Pour permettre la comparaison entre l'évolution de l'incidence sous un "effet cohorte" et sous un "effet calendaire", nous avons choisi la même forme de la fonction f induisant une baisse de l'incidence de 1% par an. Le tableau 5.1 fournit les projections de la prévalence de la démence (avec IC à 95%) en 2030 en France, en fonction des hypothèses faites sur l'évolution de l'incidence dans le temps.

TABLE 5.1 – Estimations de la prévalence de la démence (IC à 95%) en France, en fonction des hypothèses sur l'incidence (modèle **M2**).

Année	Femme	Homme	Total
Incidence homogène dans le temps : $\alpha_{01}(a, t) = \alpha_{01}(a)$			
2015	821 000	368 000	1 189 000
	[757 000;958 000]	[320 000;456 000]	[1 077 000;1 414 000]
2030	1 146 000	604 000	1 750 000
	[1 054 000;1 348 000]	[525 000;743 000]	[1 579 000;2 091 000]
"effet cohorte" baisse de 3% : $\alpha_{01}(a, t) = \alpha_{01}(a 1913).0,97^{(t-a)-1913}$			
2030	478 000	218 000	696 000
	[440 000;553 000]	[189 000;262 000]	[629 000;815 000]
"effet cohorte" baisse de 1% : $\alpha_{01}(a, t) = \alpha_{01}(a 1913).0,99^{(t-a)-1913}$			
2030	866 000	431 000	1 297 000
	[798 000;1 009 000]	[376 000;527 000]	[1 174 000;1 536 000]
"effet calendaire" baisse de 1% : $\alpha_{01}(a, t) = \alpha_{01}(a 2000).0,99^{t-2000}$			
2030	857 000	445 000	1 302 000
	[788 000;1 007 000]	[385 000;547 000]	[1 173 000;1 554 000]

Sous l'hypothèse d'une baisse de l'incidence de 3% par année de naissance les estimations de la prévalence sont très inférieures à celles estimées pour une incidence homogène dans le temps (baisse de 60% des projections en 2030 par rapport aux projections sous une incidence homogène dans le temps). Le nombre de cas serait moins important qu'en 2015.

Cette hypothèse nous semble irréaliste, nous avons donc testé l'hypothèse d'une baisse de 1% par année de naissance et le nombre de déments estimé est plus important, 1 297 000 cas en 2030.

Lorsque nous considérons que la baisse est de 1% par an pour un "effet calendaire", le nombre de déments est estimé à 1 301 000 cas.

Les projections de la prévalence de la démence totale en 2030 sont très proches quelle que soit l'hypothèse ("effet cohorte" ou "calendaire") faite sur l'incidence. De plus, sous ces deux hypothèses, l'augmentation du nombre de déments sur 20 ans est beaucoup plus modeste que lorsque nous considérons une incidence homogène dans le temps.

Cependant, nous remarquons une différence entre les hommes et les femmes puisque, pour une baisse de l'incidence de 1% par an, les projections de la prévalence de la démence sont plus importantes chez les femmes sous l'hypothèse d'un "effet cohorte", alors que chez les hommes les projections sont plus importantes sous l'hypothèse d'un "effet calendaire".

L'impact d'une hypothèse est donc différent en fonction du sexe.

5.4.3.3 Projections d'indicateurs de santé

Les projections résumées dans les tableaux 5.2, et 5.3 sont faites sous l'hypothèse d'une incidence homogène dans le temps et dans un cadre sans intervention. Cependant, le premier tableau (5.2) fournit les projections de prévalence sous le modèle **M1** où la mortalité des déments est supposée homogène dans le temps (avec les IC à 95%), en distinguant deux cas concernant la mortalité des non-déments. Le second tableau (5.3) fournit les projections des indicateurs de santé décrits dans la section 2.3 pour les hommes et les femmes (avec les IC à 95%) sous le modèle **M2** qui prend compte une mortalité des déments qui évolue au cours du temps.

TABLE 5.2 – Projections de la prévalence de la démence entre 65 et 100 ans, en France (IC à 95%) sous le modèle **M1**.

	Femmes	Hommes	Total
Hypothèse : $\alpha_{02}(a, t) \neq \alpha_2(a, t)$			
En 2015	743 000	326 000	1 069 000
	[688 000 ; 839 000]	[283 000 ; 396 000]	[971 000 ; 1 235 000]
En 2030	926 000	470 000	1 396 000
	[856 000 ; 1 054 000]	[408 000 ; 569 000]	[1 264 000 ; 1 623 000]
Hypothèse : $\alpha_{02}(a, t) = \alpha_2(a, t)$			
En 2015	628 000	154 000	782 000
	[566 000 ; 709 000]	[143 000 ; 210 000]	[709 000 ; 919 000]
En 2030	787 000	416 000	1 203 000
	[717 000 ; 911 000]	[358 000 ; 511 000]	[1 075 000 ; 1 422 000]

Nous pouvons observer que les estimations en 2015 et les projections en 2030 de la prévalence de la démence sont différentes en fonction de la façon de considérer la mortalité des non-déments. En effet, si la mortalité des non-déments est égale à la mortalité générale dans la population, les estimations et projections sont plus faibles que lorsque ces deux mortalités sont distinguées.

Ce constat appuie notre hypothèse et affirme que ne pas considérer cette différence induit une sous-estimation de la prévalence de la démence et donc une sous-estimation des projections, à cause d'une sur-estimation de la mortalité des non-déments.

En ce qui concerne les valeurs estimées en elles-mêmes, nous développons plutôt les estimations et projections de prévalence du tableau 5.3 car elles correspondent aux valeurs estimées sous l'hypothèse que nous privilégions pour le modèle Markovien.

TABLE 5.3 – Projections de la prévalence de la démence (en milliers de cas) entre 65 et 100 ans, de l'âge moyen d'entrée en démence (AV_{01}), de la probabilité de développer la maladie à 75 ans (F_{01}), des diverses espérances de vie à 75 ans (LE_{00} , $LE_{0.}$, LE_{11} , $LE_{..}$) et du nombre moyen d'années passées en démence (T_{11}), en France (IC à 95%), (modèle **M2**).

	Prev	AV_{01}	F_{01}	LE_{00}	$LE_{0.}$	LE_{11}	$LE_{..}$	T_{11}
	(N)	à 75 ans						
Femmes								
En 2015	821	84,6	0,658	11,97	15,79	8,57	15,52	3,82
	[757;958]	[83,4;85,3]	[0,621;0,696]	[11,72;12,26]	[15,71;16,00]	[7,39;10,07]	[15,50;15,58]	[3,45;4,28]
En 2030	1,146	85,3	0,703	12,44	17,04	10,41	16,77	4,60
	[1 054;1 348]	[84,2;85,9]	[0,666;0,740]	[12,19;12,72]	[16,97;17,25]	[9,17;11,73]	[16,74;16,83]	[4,25;5,06]
Hommes								
En 2015	368	81,7	0,431	10,73	12,67	6,71	12,35	1,93
	[320;456]	[80,1;82,5]	[0,383;0,481]	[10,49;10,99]	[12,59;12,82]	[5,60;7,85]	[12,32;12,39]	[1,60;2,33]
En 2030	604	82,8	0,478	11,48	13,90	8,07	13,55	2,41
	[525;743]	[81,3;83,6]	[0,425;0,531]	[11,15;11,77]	[13,82;14,02]	[6,96;9,07]	[13,51;13,59]	[2,05;2,87]

Comparé au modèle **M1** (sous l'hypothèse d'une mortalité des non-déments distincte de la mortalité générale), les estimations et les projections sont ici plus importantes pour les hommes et les femmes. Ainsi, le fait de prendre en compte une évolution de la mortalité des déments, α_{12} , au cours du temps n'est pas négligeable.

Une hausse de 39,6 % du nombre de femmes démentes entre 65 et 100 ans est attendue entre 2015 et 2030 due à une baisse naturelle de la mortalité générale. En même temps, l'âge moyen de début de la démence serait retardé de 0,7 an, mais la probabilité de développer la démence chez les femmes de 75 ans dans le reste de leur vie augmenterait.

De plus, une augmentation de toutes les espérances de vie est observée entre 2015 et 2030. L'espérance de vie générale pour une non-démence à 75 ans ($LE_{0.}$) augmenterait de 1,25 an entre 2015 et 2030 pour atteindre une durée de près de 17 ans en 2030, mais l'espérance de vie sans démence n'augmenterait que de 0,47 an (d'où une augmentation de la durée moyenne passée en démence T_{11} de 0,78 an pour les femmes de 75 ans). En 2030, les femmes de 75 ans démentes vivraient en moyenne 10,41 années avec la démence et nous estimons que les non-démentes qui développeraient la démence plus tard passeraient en moyenne 4,60 années en démence. L'espérance de vie pour une femme de 75 ans démente augmenterait de 1,84 an entre 2015 et 2030.

Les mêmes tendances sont observées chez les hommes avec des valeurs systématiquement plus basses que chez les femmes.

Les IC que nous estimons sont relativement proches des estimations sauf pour l'espérance de vie en démence LE_{11} , certainement dû au fait que le nombre de cas déments de 75 ans n'est pas très élevé.

Finalement, nous observons une augmentation générale des indicateurs de santé

entre 2015 et 2030, ceci découle principalement du vieillissement de la population et de l'allongement de l'espérance de vie.

Les figures 5.11 et 5.12 permettent de visualiser l'impact des âges élevés. Le premier graphique représente la prévalence de la démence par âge estimée en 2015 et projetée en 2030 pour les hommes et les femmes, tandis que le second décrit le taux de prévalence de la démence par âge estimé en 2015 et projeté en 2030 pour les hommes et les femmes.

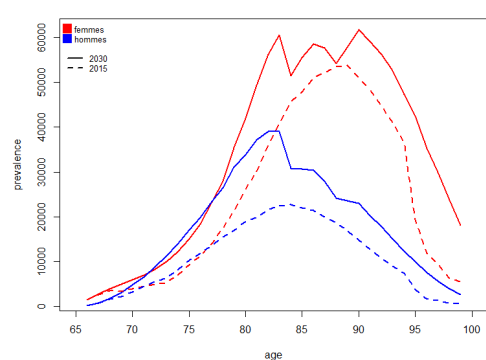


FIGURE 5.11 – Estimations de la prévalence de la démence par âge en 2015 et projections en 2030 (hommes et femmes; modèle M2).

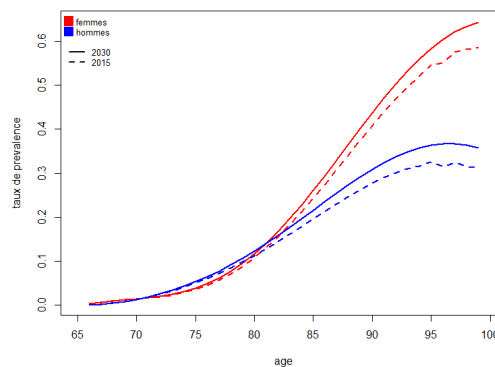


FIGURE 5.12 – Estimations du taux de prévalence de la démence par âge en 2015 et projections en 2030 (hommes et femmes; modèle M2).

Grâce à la figure 5.11, nous pouvons confirmer que la prévalence de la démence est plus forte chez les femmes que chez les hommes et qu'elle augmente fortement entre 2015 et 2030.

Néanmoins, le tracé de la prévalence de la démence par âge apporte de l'information supplémentaire à ce que nous avons pu constater en comparant les prévalences totales.

En effet, la répartition en fonction de l'âge n'est pas homogène, le nombre de cas prévalents (nulle à 65 ans selon notre hypothèse) augmente rapidement jusqu'à atteindre un maximum aux alentours de 84 ans chez les hommes en 2015 et 88 ans chez les femmes en 2015, suivi d'une baisse tout aussi rapide après ce pic.

La même tendance est observée en 2030 mise à part que le maximum est atteint plus tôt, aux alentours de 82 ans pour les hommes et les femmes, et que pour les femmes la prévalence par âge reste très élevée jusqu'à 91 ans avant de diminuer.

De plus, nous constatons que la prévalence de la démence chez les hommes et chez les femmes est très similaire entre 65 et 77 ans et c'est seulement après que la prévalence par âge devient beaucoup plus importante chez les femmes (cette observation est valable en 2015 et 2030).

Si nous observons maintenant les taux de prévalence de la démence par âge (figure 5.12), nous constatons qu'ils augmentent avec l'âge et sont très proches chez les hommes et les femmes jusqu'à 82 ans, puis augmentent nettement plus vite chez les femmes. Nous constatons également que la différence des taux entre 2015 et 2030 est presque nulle pour les âges les plus jeunes, puis l'écart augmente au fur et à mesure que nous avançons en âge. Par exemple, pour les femmes de 80 ans le taux de prévalence est estimé à 10,8% en 2015 vs 11,7% en 2030 (soit une augmentation d'environ 1% pour cet âge), alors que pour les femmes de 90 ans le taux est estimé à 40,9% en 2015 vs 43,7% en 2030 (soit une augmentation d'environ 3% pour cet âge).

Ces deux figures montrent bien que si le nombre de déments augmente beaucoup entre 2015 et 2030, c'est essentiellement due aux personnes les plus âgées (figure 5.12), et comme la population vieillie alors le nombre de déments très âgés augmente d'autant plus (figure 5.11).

5.4.3.4 Impacts de l'intervention sur les projections (modèle M2)

Dans ce paragraphe, nous présentons les résultats concernant l'impact de l'intervention sous différents scénarios, pour le modèle qui considère que la mortalité des déments évolue au cours du temps. La quantité de résultats étant importante, nous focalisons ce paragraphe sur les femmes et présentons les hommes en annexe A.3.2. Le tableau 5.4 donne donc les résultats des projections d'indicateurs de santé pour les femmes en France en 2030.

TABLE 5.4 – Projections après l'intervention de la prévalence de la démence (en milliers de sujets), de l'âge moyen de début de la démence AV_{01} et de la probabilité de développer la démence F_{01} à 75 ans, pour les femmes en France, (IC à 95% ; modèle M2).

Scenario										Prev	AV_{01}	F_{01}	
No.	$p_0(65)$	$p_0^*(65)$	ω	θ_{01}	θ_{01}^*	θ_{02}	θ_{02}^*	θ_{12}	θ_{12}^*	(N)	%#	(à 75 ans)	
Projections de référence en 2030										1 146	-	85,3	0,703
										[1 054;1 348]		[84,2;85,9]	[0,666;0,740]
Projections en 2030 après l'intervention en 2015													
1.a	0,4	0,2	0,5	1,5	1,5	2	2	2	2	1 410	+23,0	85,8	0,709
										[1 391;1 735]		[84,5;86,3]	[0,672;0,742]
1.b	0,4	0,4	1	1,5	1,27	2	1,74	2	1,74	1 149	+0,3	85,6	0,697
										[1 094;1 406]		[84,3;86,1]	[0,662;0,731]
2.a	0,4	0,2	0,5	2	2	1,5	1,5	1,5	1,5	1,297	+13,2	86,1	0,687
										[1 195;1 474]		[84,8;86,6]	[0,657;0,717]
2.b	0,4	0	0	2	2	1,5	1,5	1,5	1,5	1,447	+26,3	86,8	0,671
										[1 342;1 625]		[85,6;87,3]	[0,642;0,705]
3	0,2	0,2	1	3	1,5	1,5	1,5	1,5	1,5	1,050	-8,4	86,0	0,678
										[977;1 259]		[84,7;86,5]	[0,643;0,714]

% de variation comparé aux projections sans l'intervention.

En gras les valeurs de paramètres modifiées par l'intervention.

Ces résultats montrent que les scénarios pour l'intervention (ciblant tous les plus de 65 ans en 2015) peuvent avoir un fort impact sur la prévalence de la démence en 2030. Pour les scénarios 1.a, 2.a et 2.b qui réduisent la prévalence

du facteur de risque en 2015, nous observons une augmentation importante de la prévalence de 23%, 13,2% et 26,3% (respectivement), comparée aux projections sans intervention. L'augmentation du nombre de déments est plus importante quand le facteur de risque ciblé est plus fortement associé au décès qu'à la démence (scénarios 1.a *vs* 2.a). Pour le scénario 1.b où l'intervention diminue les risques relatifs associés à l'HBP de 13% et 15 %, la prévalence de la démence est presque inchangée. Enfin, pour le scénario 3 qui implique une intervention réduisant le risque de démence parmi les individus porteurs du gène ApoE4, les projections de la prévalence de la démence en 2030 pourraient diminuer de 8,4%.

Les projections de prévalence que nous obtenons contrastent avec les résultats de Jacqmin *et al.* [Jacqmin-Gadda 2013] pour les mêmes scénarios, mais avec une intervention ne ciblant que les individus de 65 ans au temps τ .

Cependant, nos résultats ne sont pas comparables tels quels, c'est pourquoi nous avons refait des projections sous l'intervention de Jacqmin *et al.* et nous présentons une comparaison des résultats avec notre intervention pour le scénario 1.a en Annexe A.3.3.

Tous les scénarios retardent l'âge moyen d'entrée en démence, celui-ci varie de 0,3 année pour le scénario 1.b à 1,5 année pour le scénario 2.b, le scénario extrême qui suppose une éradication du facteur de risque très fortement associé à la démence ($\theta_{01} = 2$) et modérément associé au décès ($\theta_{02} = \theta_{12} = 1.5$).

La figure 5.13 décrit la probabilité de développer la démence avant de décéder pour les femmes vivantes et non malades en fonction de leur âge, en 2015 et en 2030, tout en comparant le contexte sans intervention au contexte avec intervention sous les scénarios 1.a et 1.b (considérés comme les plus réalistes actuellement).

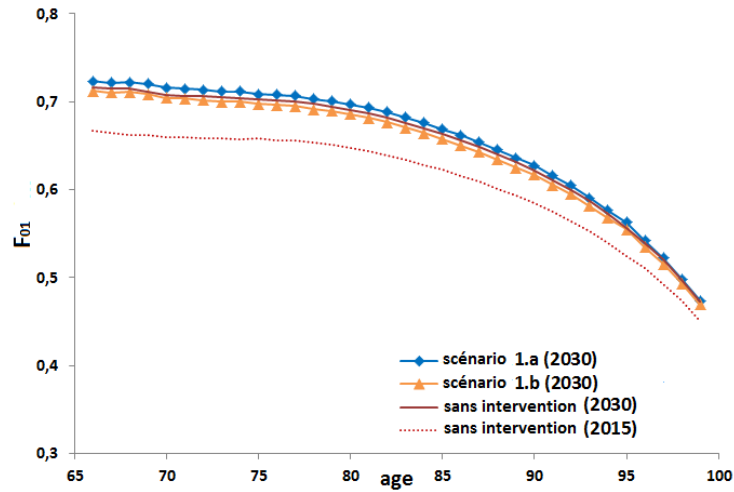


FIGURE 5.13 – Évolution de la probabilité de développer une démence par âge, pour les femmes vivantes et non-démences en 2015 et 2030, sans intervention ou avec les interventions 1.a et 1.b (modèle **M2**).

Ces probabilités augmentent entre 2015 et 2030 à travers l’âge, mais les interventions pour les scénarios 1.a et 1.b auraient un petit impact sur les projections de 2030. Cependant, le tableau 5.4 met en avant une petite réduction de la probabilité de développer la démence pour les femmes de 75 ans sous les scénarios 2.a, 2.b et 3.

Le Tableau 5.5 présente les résultats pour les autres indicateurs.

TABLE 5.5 – Projections après l’intervention des espérances de vie LE_{00} , $LE_{0.}$, LE_{11} , $LE_{.}$ et du nombre moyen d’années passées en démence T_{11} (pour des femmes non démentes) à 75 ans, pour les femmes en France, (IC à 95% ; modèle **M2**).

Scenario	No.	$p_0(65)$	$p_0^*(65)$	ω	θ_{01}	θ_{01}^*	θ_{02}	θ_{02}^*	θ_{12}	θ_{12}^*	LE_{00}	$LE_{0.}$	LE_{11}	$LE_{.}$	T_{11}
Projections de référence en 2030															
											12,44	17,04	10,41	16,77	4,60
											[12,19 ; 12,72]	[16,97 ; 17,25]	[9,17 ; 11,73]	[16,74 ; 16,83]	[4,25 ; 5,06]
Projections en 2030 après l’intervention en 2015															
1.a	0,4	0,2	0,5	1,5	1,5	2	2	2	2		13,01	17,89	11,73	17,66	4,88
											[12,61 ; 13,25]	[17,79 ; 18,07]	[10,56 ; 12,73]	[17,60 ; 17,70]	[4,54 ; 5,46]
1.b	0,4	0,4	1	1,5	1,27	2	1,74	2	1,74		12,78	17,41	11,22	17,19	4,63
											[12,39 ; 13,02]	[17,31 ; 17,60]	[10,03 ; 12,24]	[17,12 ; 17,22]	[4,29 ; 5,21]
2.a	0,4	0,2	0,5	2	2	1,5	1,5	1,5	1,5		13,14	17,71	11,15	17,49	4,57
											[12,79 ; 13,34]	[17,59 ; 17,87]	[9,95 ; 12,19]	[17,42 ; 17,51]	[4,25 ; 5,08]
2.b	0,4	0	0	2	2	1,5	1,5	1,5	1,5		13,82	18,36	12,02	19,11	4,54
											[13,46 ; 14,03]	[18,23 ; 18,50]	[10,78 ; 13,07]	[18,09 ; 18,21]	[4,20 ; 5,04]
3	0,2	0,2	1	3	1,5	1,5	1,5	1,5	1,5		12,91	17,25	10,82	17,05	4,34
											[12,53 ; 13,13]	[17,13 ; 17,43]	[9,61 ; 11,87]	[16,98 ; 17,09]	[4,00 ; 4,90]

En gras les valeurs de paramètres modifiées par l’intervention.

Nous observons que tous les scénarios mènent à une augmentation des espérances de vie (avec ou sans démence), car ils réduisent la mortalité, soit directement (scénarios 1 et 2) ou indirectement en diminuant l’incidence de la démence (scénario

3).

Apparemment, l'impact sur les espérances de vie est plus important quand la mortalité est directement modifiée (scénario 1.b). En revanche, le nombre moyen d'années passées en démence (T_{11}) décroît significativement seulement quand l'intervention réduit exclusivement l'incidence de la démence ($T_{11} = 4,34$ ans dans le scénario 3 *vs* 4,60 ans sans intervention) du fait d'une augmentation plus importante de l'espérance de vie sans démence (LE_{00}) par rapport à l'espérance de vie pour les non-déments (LE_0).

La figure 5.14 décrit les projections des espérances de vie LE_{00} , LE_{11} et LE_0 , ainsi que le nombre moyen d'années passées en démence T_{11} par âge, pour les femmes, sans intervention, pour les années 2015 et 2030, et avec intervention en 2030 sous les scénarios 1.a et 1.b.

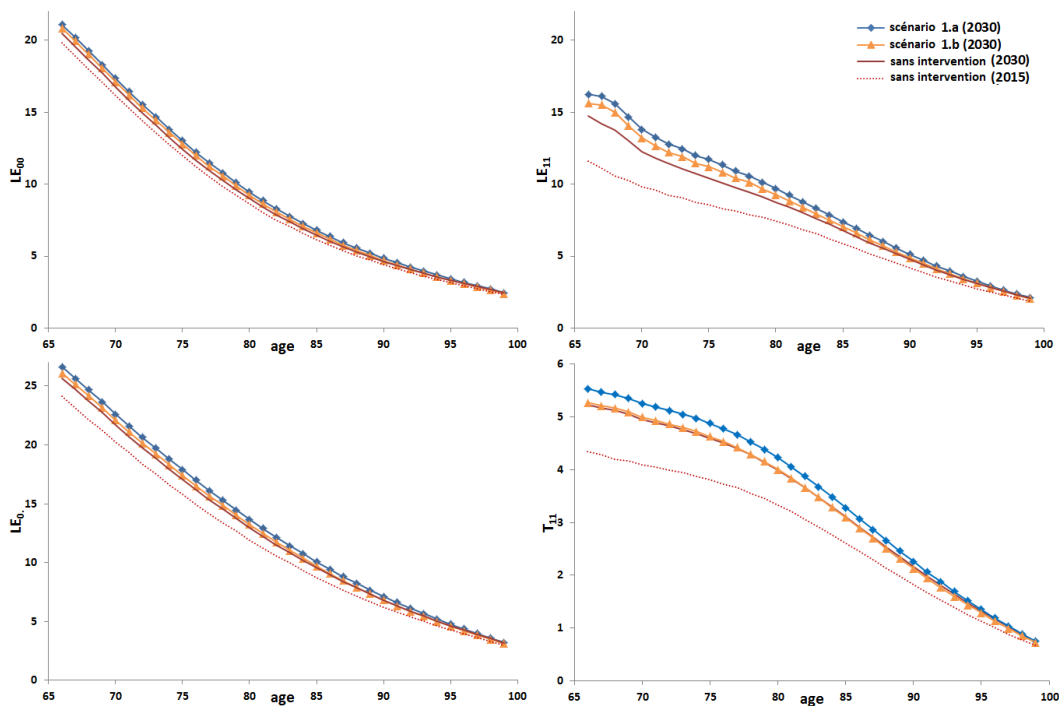


FIGURE 5.14 – Projections des espérances de vie et du nombre moyen d'années passées en démence par âge, pour les femmes, sans intervention et avec intervention (scénarios 1.a et 1.b; modèle **M2**).

L'impact de l'intervention apparaît plus important pour les sujets d'âges jeunes, car le risque de décès devient prédominant pour les sujets âgés quel que soit le contexte, ainsi, toutes les courbes convergent vers la même valeur à 99 ans.

Toutes les espérances de vie augmentent pour les deux scénarios d'intervention comparés aux projections sans intervention, mais la différence est plus grande pour l'espérance de vie des sujets déments (LE_{11}).

Le nombre moyen d'années passées en démence (T_{11}) change significativement pour le scénario 1.a, car l'intervention impacte plus fortement la mortalité.

En effet, sous ce scénario, l'intervention diminue de moitié les individus exposés au facteur de risque et ceux-ci avaient un fort risque de décéder, du coup, l'espérance de vie générale augmente et en conséquence, le nombre moyen d'années passées en démence également.

Notez que les figures 5.13 et 5.14 et les Tableaux 5.4 et 5.5 montrent clairement que les 5 hypothétiques scénarios d'intervention, choisis d'après les connaissances sur les facteurs de risque de la démence, auraient moins d'impacts sur toutes les mesures épidémiologique du fardeau de la démence (sauf pour l'âge de début de la démence) que la baisse naturelle de la mortalité attendue entre 2015 et 2030.

Les résultats pour les hommes se trouvent dans l'annexe A.3.2. Ils sont très similaires à ceux des femmes mis à part que pour la plupart des indicateurs de santé, les projections sont toujours inférieures à celles des femmes.

5.4.4 Sous le modèle semi-Markovien

Passons maintenant aux principaux résultats avec le modèle semi-Markovien non-homogène.

5.4.4.1 Estimations

À partir des données de la cohorte

Les données de la cohorte permettent d'estimer l'incidence en fonction de l'âge (présentée dans la figure 5.8 et comparée avec l'incidence des modèles Markoviens dans cette partie 5.4.1), que nous détaillons davantage en annexe A.3.4.

Les données de cohorte nous permettent également d'estimer le sur-risque de décès des déments par rapport aux non-déments en fonction du temps passé avec la démence, toujours pour le modèle **S2**. Cette estimation est décrite dans la figure 5.15 pour les hommes et les femmes avec les intervalles de confiance à 95%.

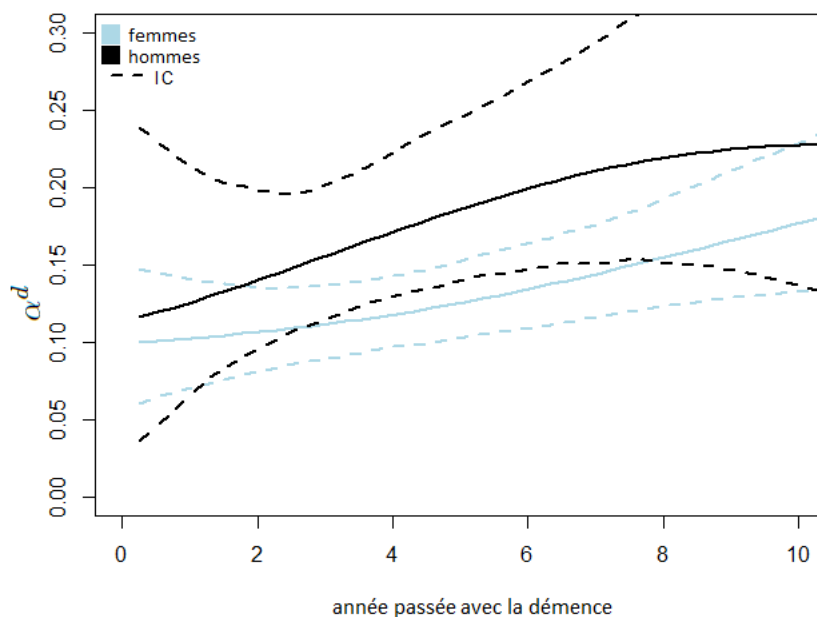


FIGURE 5.15 – Sur-risque de décès α^d pour les sujets déments en fonction du temps passé avec la démence, et IC à 95% pour les hommes et les femmes, cohorte PAQUID (modèle S2).

Nous remarquons que le sur-risque augmente avec le temps passé en démence, de plus, il est plus important chez les hommes par rapport aux femmes.

L'estimation de ce sur risque n'est pas à interpréter tel quel, il intervient dans l'équation qui permet d'estimer la mortalité des déments en fonction de la durée passée avec la maladie et de l'âge au temps t .

Après résolution d'équations différentielles : mortalité parmi les non-déments et les déments

Les projections de l'INSEE ont été utilisées pour ré-estimer la mortalité des non-déments α_{02} en prenant en compte l'âge, mais aussi l'évolution dans le temps calendaire. Nous avons vu (section 3.3.2) que cela nécessite la résolution d'une équation différentielle autant de fois que d'années t nécessaires. Tandis que, l'estimation sur la cohorte PAQUID nous donne l'impact de la durée passée avec la maladie à travers la fonction α^d . Ainsi, avec le modèle additif que nous avons proposé (section 3.3.2), la mortalité des déments, qui s'écrit comme la somme de ces deux fonctions, dépend de la durée passée avec la démence, de l'âge et implicitement du temps calendaire (puisqu'elle fait intervenir $\alpha_{02}(a, t)$).

Les figures 5.16 et 5.17 sont des représentations 3D de l'estimation des taux de mortalité parmi les sujets déments en fonction de l'âge et du temps passé avec la maladie pour les sujets qui ont développé la démence en 2015.

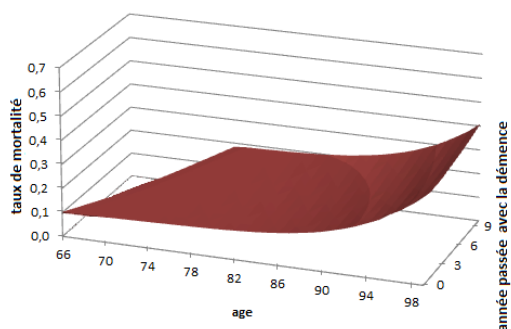


FIGURE 5.16 – Mortalité par âge et temps passé avec la démence pour les femmes développant la démence en 2015 (modèle **S2**).

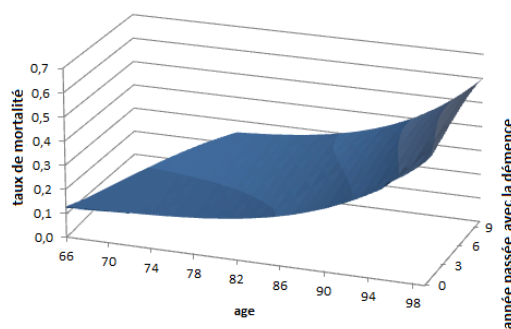


FIGURE 5.17 – Mortalité par âge et temps passé avec la démence pour les hommes développant la démence en 2015 (modèle **S2**).

Les estimations sont plus basses pour les femmes que pour les hommes et cet écart augmente avec l'âge. Cependant, les mêmes tendances sont observées pour les deux, à savoir que les taux de mortalité augmentent avec l'âge et avec le temps passé avec la maladie.

Remarque : les valeurs très hautes que nous observons aux âges élevés et pour une longue durée passée avec la démence sont rarement rencontrées. En effet, elles correspondent à des cas extrêmes, par exemple, nous observons une mortalité proche de 0,6 pour des hommes diagnostiqués à 95 ans qui restent 5 ans en démence et qui ont donc 100 ans à ce moment-là.

Pour une meilleure description de ces évolutions et dans le but d'observer les changements au cours du temps calendaire, la figure 5.18 montre l'évolution de la mortalité parmi les déments développant la maladie à 80 ans, en 2015 et en 2030.

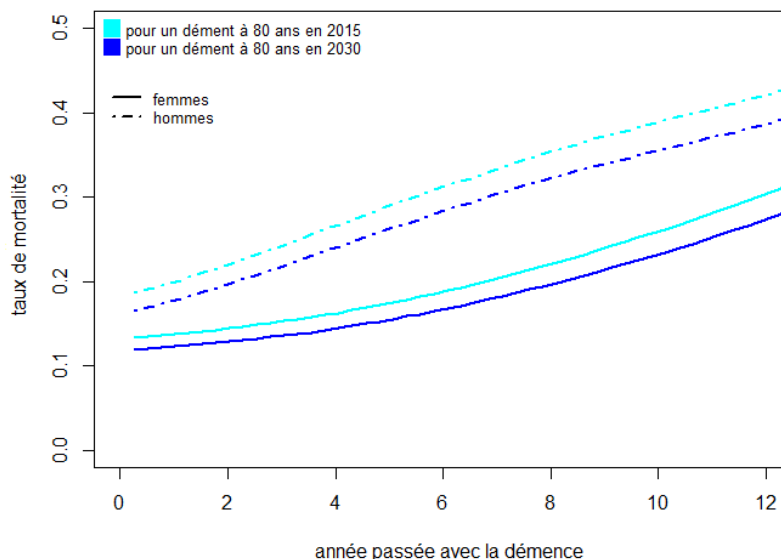


FIGURE 5.18 – Evolution de la mortalité des déments (α_{12}) en fonction du temps passé avec la maladie, pour des hommes et des femmes diagnostiqués à 80 ans, en 2015 et 2030 (modèle **S2**).

Tout d’abord, nous constatons que la mortalité des déments augmente avec la durée passée avec la maladie et qu’elle est plus forte chez les hommes que chez les femmes.

Nous observons également des tendances légèrement différentes : la mortalité des déments semble plus stable au début chez les femmes puis augmente plus rapidement au bout de 4 ans passés avec la démence, alors que pour les hommes, elle augmente constamment dès le début.

Enfin, comme nous nous y attendions, les projections du taux de mortalité pour 2030 sont plus faibles que celles de 2015, et la mortalité des déments reste supérieure chez les hommes.

5.4.4.2 Projections d’indicateurs de santé

Le tableau 5.6 présente les estimations en 2015 et les projections pour 2030 de divers indicateurs de santé avec leurs intervalles de confiance à 95% sous le modèle **S2** où la mortalité des déments dépend de l’âge et de la durée passée en démence.

TABLE 5.6 – Projections de la prévalence de la démence, de l'âge moyen d'entrée en démence AV_{01} , ainsi que de la probabilité de développer la démence F_{01} et les espérance de vie LE_{00} , $LE_{..}$ à 80 ans en France, avec les IC à 95%, sous le modèle **S2**.

	Prev (N)	AV_{01}	F_{01}	LE_{00} (à 80 ans)	$LE_{..}$
Femmes					
En 2015	754 000 [694 000 ; 878 000]	84,1 [82,6 ; 84,6]	0,63 [0,59 ; 0,68]	8,86 [8,65 ; 9,07]	11,36 [11,28 ; 11,43]
En 2030	987 000 [910 000 ; 1 157 000]	85,1 [83,7 ; 85,6]	0,70 [0,66 ; 0,75]	9,52 [9,29 ; 9,77]	12,40 [12,31 ; 12,47]
Hommes					
En 2015	297 000 [261 000 ; 337 000]	80,9 [79,8 ; 81,5]	0,36 [0,31 ; 0,41]	7,58 [7,37 ; 7,78]	8,44 [8,17 ; 8,58]
En 2030	455 000 [398 000 ; 520 000]	82,0 [81,0 ; 82,7]	0,41 [0,35 ; 0,46]	8,38 [8,12 ; 8,60]	9,34 [9,06 ; 9,49]

Du fait de la baisse attendue de la mortalité associée à une augmentation de l'espérance de vie générale, et bien que nous supposons une incidence de la démence homogène dans le temps, la prévalence, l'âge moyen d'entrée en démence AV_{01} et la probabilité de développer la démence F_{01} vont augmenter de 2015 à 2030 pour les deux sexes.

En 2015, nous estimons le nombre de cas de déments en France à environ 1 051 000 ([955 000 ; 1 215 000]), et nous projetons près de 1 442 000 ([1 308 000 ; 1 677 000]) cas pour 2030. La prévalence de la démence est toujours plus élevée chez les femmes que chez les hommes.

D'autre part, l'espérance de vie sans la démence pour les sujets non-déments à 80 ans augmenterait de 7,58 à 8,38 ans pour les hommes et de 8,86 à 9,52 ans pour les femmes entre 2015 et 2030, un peu moins que leur espérance de vie générale ($LE_{..}$). Des figures décrivant l'évolution de l'espérance de vie sans la démence en fonction de l'âge entre 2015 et 2030 sont présentées en annexe A.3.4.

5.5 Choix du modèle

Dans cette application à la démence, nous avons proposé plusieurs modèles pour donner des projections d'indicateurs de santé.

Les hypothèses faites sur ces modèles ne sont pas toujours les mêmes et la méthode d'estimation connaît donc quelques variantes.

Pour chaque modèle, nous estimons tout d'abord la prévalence en 2015, ce qui permet directement de savoir si nos estimations sous ces modèles sont proches de la prévalence attendue dans la population. Ensuite, nous donnons des projections pour 2030.

Le tableau 5.7 récapitule l'ensemble des résultats de projections de la prévalence de la démence que nous avons obtenus sous nos modèles (sans considérer l'impact d'interventions).

TABLE 5.7 – Récapitulatif des projections de la prévalence de la démence (IC à 95%) en 2030 en France, en fonction des modèles et hypothèses.

	Femmes	Hommes	Total
Modèle M1 : $\alpha_{12}(a, t) = \alpha_{12}(a)$			
Hypothèse : $\alpha_{02}(a, t) \neq \alpha_2(a, t)$			
	926 000	470 000	1 396 000
	[856 000 ; 1 054 000]	[408 000 ; 569 000]	[1 264 000 ; 1 623 000]
Hypothèse : $\alpha_{02}(a, t) = \alpha_2(a, t)$			
	787 000	416 000	1 203 000
	[717 000 ; 911 000]	[358 000 ; 511 000]	[1 075 000 ; 1 422 000]
Modèle M2 : $\alpha_{12}(a, t) = \alpha_{02}(a, t)g(a)$			
Incidence homogène dans le temps			
	1 146 000	604 000	1 750 000
	[1 054 000 ; 1 348 000]	[525 000 ; 743 000]	[1 579 000 ; 2 091 000]
"effet cohorte" baisse de 3% : $\alpha_{01}(a, t) = \alpha_{01}(a 1913) \cdot 0,97^{(t-a)-1913}$			
	478 000	218 000	696 000
	[440 000 ; 553 000]	[189 000 ; 262 000]	[629 000 ; 815 000]
"effet cohorte" baisse de 1% : $\alpha_{01}(a, t) = \alpha_{01}(a 1913) \cdot 0,99^{(t-a)-1913}$			
	866 000	431 000	1 297 000
	[798 000 ; 1 009 000]	[376 000 ; 527 000]	[1 174 000 ; 1 536 000]
"effet calendaire" baisse de 1% : $\alpha_{01}(a, t) = \alpha_{01}(a 2000) \cdot 0,99^{t-2000}$			
	857 000	445 000	1 302 000
	[788 000 ; 1 007 000]	[385 000 ; 547 000]	[1 173 000 ; 1 554 000]
Modèle S2 : $\alpha_{12}(a, t, d) = \alpha_{02}(a, t) + \alpha^d(d)$			
	987 000	455 000	1 442 000
	[910 000 ; 1 157 000]	[398 000 ; 520 000]	[1 308 000 ; 1 677 000]

Les projections varient d'un modèle à l'autre. Au sein d'un même modèle, nous avons déjà discuté les choix des hypothèses.

Pour le modèle **M1**, il est plus judicieux pour la démence de distinguer la mortalité des non-déments de la mortalité générale dans la population.

Pour le modèle **M2**, les hypothèses d'une évolution de l'incidence dans le temps que nous avons testées sont exploratoires, il est donc préférable de rester sur une hypothèse de l'incidence de la démence homogène au cours du temps.

Enfin, entre ces deux modèles, il semble évident que l'hypothèse de la mortalité des déments qui évolue au cours du temps est la plus réaliste. Les travaux que nous menons sur cette maladie chronique privilégient donc le modèle **M2**.

Pour le modèle **S2**, nous avons nous-même choisi de n'appliquer que ce modèle

semi-Markovien, en nous basant sur les conclusions des travaux déjà menés avec le modèle **S1** sur la cohorte PAQUID.

Comme nous avons accès aux données de la cohorte PAQUID, il nous est possible de tester l'adéquation de chacun de ces modèles. Cependant, nous n'avons aucune garantie que le meilleur modèle en 2000 soit le meilleur modèle pour l'année 2030.

Des travaux menés par l'équipe pour tester l'adéquation sur les données de PAQUID d'un modèle Markovien (**M1** ou **M2**) et un modèle semi-Markovien de type **S1** [Commenges 2007] ont conclu que le modèle Markovien était préférable au semi-Markovien.

Mais, nous n'avons pas comparé entre eux les trois modèles que nous proposons ici sur les données de PAQUID, car le but n'était pas de choisir un modèle pour PAQUID, mais pour des projections.

Pour construire et choisir le modèle de projections, il fallait, certes, que les estimations passées soient conformes aux observations de la cohorte et que la pertinence du modèle soit suffisante, mais il est davantage important de discuter les hypothèses, réfléchir aux évolutions potentielles concernant la maladie, mais aussi la population, les progrès...etc, plutôt que de choisir en fonction de la meilleure adéquation sur les données de cohorte.

5.6 Conclusion sur l'application à la démence

Pour cette application, la qualité des données que nous avons nous a permis de tester différents modèles tout en faisant varier les hypothèses.

Nous avons présenté deux modèles Markoviens et un modèle semi-Markovien, afin de comprendre l'influence de certaines hypothèses. Pour chacun des modèles, nous donnons des projections de la prévalence de la démence en 2030 afin de comparer et discuter des modèles.

Quel que soit le modèle, les estimations pour 2015 sont assez proches de celles attendues et leur précision/plausibilité est bonne. Quant aux projections pour 2030, elles vont systématiquement dans le sens d'une augmentation du nombre de déments et d'une augmentation du fardeau de la maladie, sauf si nous considérons une baisse de l'incidence de la démence au cours du temps, mais nous manquons encore d'informations sur ce point pour pouvoir confirmer des projections (nous pourrions le faire quand nous aurons plus de données).

Pour conclure sur cette application, nous allons discuter de l'intervention que nous avons proposée et de son impact, ainsi que des différentes hypothèses faites sur les paramètres du modèle illness-death à savoir sur l'incidence de la démence la mortalité des déments et des non-déments.

5.6.1 Intervention

L'intervention que nous proposons cible les facteurs de risque de la maladie plutôt que la maladie en elle-même car les maladies sur lesquelles nous travaillons sont incurables actuellement.

Pour la démence, bien que de nombreux facteurs de risque soient suspectés de nos jours, peu d'entre eux sont prouvés et leur impact sur la maladie est encore débattu. C'est pourquoi nous nous sommes focalisés sur le gène ApoE4 et l'HBP, qui sont des facteurs de risque bien établis de la démence et également du décès et qui ont déjà été considérés comme cibles potentielles d'une intervention. De plus, la prévalence de l'exposition est relativement connue dans la population [Tilvis 2004], hypothèse nécessaire pour que nous proposons des scénarios plausibles. Dans notre cas, nous avons pu considérer que l'HBP est un facteur de risque dit à "mid-life", car seule une HBP à "mid-life" a été trouvée associée au risque de démence.

Pour résumer, nos résultats, concernant l'impact de l'intervention sur le fardeau futur de la démence à travers tous les indicateurs de santé, suggèrent que les interventions hypothétiques, ciblant soit l'ApoE4 soit l'HBP, auraient moins d'effet qu'une baisse naturelle de la mortalité prévue entre 2015 et 2030.

De plus, d'après nos résultats, supposer qu'une intervention retarderait le début de la maladie de 2 ans, comme Brookmeyer *et al.* [Brookmeyer 2007] voir de 2 à 5 ans comme Vickland *et al.* [Vickland 2010] semble trop optimiste. En effet, le plus grand délai estimé dans nos scénarios est de 1,5 an pour un scénario extrême (2.b) où il s'agit d'un facteur de risque avec une forte prévalence (40% de cas exposés dans la population), qui double le risque de développer la démence et augmente le risque de décéder de 50% et qui serait éradiqué (donc la prévalence de l'exposition vaut 0) par l'intervention.

Aux vues des connaissances actuelles sur les facteurs de risque de la démence, il est difficilement raisonnable d'envisager des scénarios encore plus efficaces.

Il est intéressant de noter que les estimations des différents indicateurs de santé ne sont pas d'égale précision. Pour la prévalence, les intervalles de confiance sont larges, tandis que pour les espérances de vie, elles sont étroites. Ceci est en partie dû au fait que l'incertitude est considérée sur l'incidence de la démence, $\alpha_{01}(a, t)$, et sur le risque relatif de décès des déments par rapport aux non-déments, $g(a)$, mais pas sur la mortalité générale dans la population $\alpha_2(a, t)$ obtenue *via* l'INSEE. Cependant, il serait possible de considérer une incertitude avec les différents scénarios proposés par l'INSEE dont nous avons parlé dans la section 3.4.

Concernant les hypothèses que nous avons imposées pour cette application, elles sont flexibles. En effet, la proportion de sujets exposés au facteur de risque, ω , qui deviennent non-exposés après l'intervention et l'évolution des risques relatifs, θ_{ij} , après l'intervention reflètent l'efficacité de l'intervention, mais aussi la proportion

de sujets conformes et réellement impactés par l'intervention, et ce n'est pas nécessairement 100%. Nous pouvons donc jouer sur cela et décider que seule une partie des exposés sont impactés ou bien de façons différentes...etc.

De même, la méthode permet de cibler certaines tranches d'âge et, pourquoi pas, d'envisager des effets différents de l'intervention en fonction de classes d'âge.

Notez également que dans l'application les scénarios d'intervention proposés concernent un facteur de risque à la fois. Cependant, la méthode permet d'intervenir sur plusieurs facteurs de risque, pour cela, il suffit de déterminer une prévalence et des risques relatifs communs (moyennés et pondérés par exemple) à ces facteurs.

Néanmoins, notre intérêt reste de proposer des scénarios d'interventions qui soient réalistes et éthiques.

5.6.2 L'Hypothèse sur l'incidence de la démence

Bien que des études suggèrent une évolution de l'incidence de la démence au cours du temps [Fratiglioni 2000], [Rocca 2011], [Grasset 2015], ces résultats manquent encore de consistance et peuvent être modifiés par des changements de critères diagnostics de la démence.

C'est pourquoi, nous avons développé la méthode pour prendre en compte l'hypothèse d'une évolution de l'incidence au cours du temps, mais nous n'avons pas poussé nos recherches sur le type d'évolution (choix de la fonction f) et nous présentons donc des résultats seulement pour les projections de prévalence sous des hypothèses simples.

Néanmoins, la méthode permet d'envisager une fonction f beaucoup plus évoluée, non-monotone au cours du temps par exemple.

Nous pouvons tout de même observer qu'une baisse de 1% par an à partir de l'année 2000 induit une diminution de près de 25% des projections de prévalence de la démence en 2030 comparée aux projections sous l'hypothèse d'une incidence homogène dans le temps.

Penser que l'incidence de la démence évolue au cours du temps est sensé, notamment en connaissance de certains facteurs connus comme le niveau d'éducation ou l'alimentation, l'hygiène de vie..., mais des recherches restent à effectuer dans ce domaine pour déterminer le type d'évolution.

5.6.3 Les hypothèses sur la mortalité des déments et des non-déments

5.6.3.1 La mortalité des non-déments

La nécessité de distinguer la mortalité des non-déments de la mortalité générale dans la population semble évidente ici. En effet, nous avons montré graphiquement cette différence et nous avons constaté que les projections de la prévalence de la démence sont nettement sous-estimées si nous ne prenons pas en compte cette différence (tableau 5.7).

Cependant, ce n'est pas le cas pour toutes les maladies chroniques, notamment celles dont la prévalence est peu importante dans la population ou avec une faible sur-mortalité des malades.

Dans ce cas, considérer que la mortalité des non-malades est équivalente à la mortalité générale de la population se justifie et permet de faciliter l'estimation, mais aussi les projections, faites avec ce modèle.

5.6.3.2 La mortalité des déments : l'alternative semi-Markovienne

Le modèle Markovien **M2**, bien que très satisfaisant pour cette application, ne prend pas en compte la durée passée avec la démence. De ce fait, deux femmes démentes âgées de 80 ans auraient la même espérance de vie même si une des femmes est démente depuis 10 ans et l'autre seulement 1 an.

Pour la démence, il a été démontré que l'utilisation d'une hypothèse Markovienne simple sur les données de PAQUID est préférable à un modèle semi-Markovien simple [Commenges 2007]. Mais, actuellement, aucuns travaux ne permet de décider lequel des modèles **M2** et **S2** est le meilleur.

De plus, comme nous l'avons discuté (section 5.5), pour faire des projections il est préférable de choisir le modèle qui semble le plus adapté en fonction de l'évolution de la maladie, plutôt que le modèle avec la meilleure adéquation sur les données passées.

Un modèle qui permet de faire dépendre la mortalité des déments à la fois de l'âge et de la durée passée avec la maladie (tel que le modèle **S2**) est une alternative intéressante à étudier. Car en effet, nous avons montré que la mortalité des déments augmente avec la durée passée avec la maladie à la fois pour les hommes et pour les femmes.

Cependant, le modèle semi-Markovien semble peu adapté aux âges élevés où l'effet de l'âge sur la mortalité est le facteur de risque le plus important. Par contre, nous pensons que ce modèle peut être fort intéressant et même préférable dans le cas d'une pathologie qui touche aussi des âges jeunes.

Pour la démence, l'effet de l'âge ayant plus d'impact sur la mortalité que la durée passée avec la maladie, c'est donc le modèle **M2** que nous préférons utiliser et interpréter parmi les modèles que nous avons proposés.

Application à la maladie de Parkinson

Sommaire

6.1	Contexte et données	104
6.1.1	Parkinson : définition et état des lieux	104
6.1.2	Données	105
6.2	Modèles et hypothèses	107
6.3	Adaptation de la méthode	109
6.3.1	Pour le modèle Markovien	110
6.3.2	Pour le modèle semi-Markovien	110
6.4	Résultats	112
6.4.1	Estimations avec le modèle Markovien	112
6.4.2	Estimations avec le modèle semi-Markovien	116
6.4.3	Projections	117
6.5	Conclusion sur le Parkinson	118
6.5.1	Incidence et Prévalence observées	119
6.5.2	Perspectives	120

Cette application a été possible grâce à l'intérêt qu'a porté Alexis Elbaz, Directeur de recherche au sein du centre INSERM U 1018 à Villejuif (94) de l'équipe "Épidémiologie du vieillissement et des maladies liées à l'âge", et Frédéric Moisan, actuellement à l'INVS de Paris, à nos travaux.

Ils ont donc souhaité que nous utilisions notre approche dans le but de faire des projections de la prévalence du Parkinson, maladie sur laquelle porte leur recherche.

Dans ce chapitre, nous présentons, tout d'abord, le contexte et les données, puis les modèles que nous avons utilisés ainsi que les adaptations qui ont été nécessaires, pour ensuite, appliquer ces modèles et donner des projections de la prévalence de la maladie de Parkinson.

6.1 Contexte et données

6.1.1 Parkinson : définition et état des lieux

La maladie de Parkinson représente la cause la plus fréquente de syndrome Parkinsonien. Bien qu'elle soit la deuxième pathologie neuro-dégénérative après la maladie d'Alzheimer, elle demeure une maladie plus rare avec une incidence de l'ordre de 0,1 à 0,2 nouveau cas par an pour 1000 personnes [Bower 1999] et une prévalence généralement comprise entre 1 et 2 cas pour 1 000 personnes [von Campenhausen 2005]. Exceptionnelle avant 50 ans, la fréquence de la maladie augmente avec l'âge.

Toutefois, le nombre de personnes atteintes de la maladie de Parkinson en France n'est pas connu précisément. L'identification des patients Parkinsoniens représente l'une des principales difficultés rencontrées pour étudier l'épidémiologie de la maladie de Parkinson, notamment pour estimer sa fréquence.

Les approches classiques, comme les certificats de décès ou les statistiques d'hospitalisations, ne représentent pas une approche valide d'identification des patients Parkinsoniens. De même, les cohortes ne permettent pas forcément une bonne estimation de l'incidence de cette maladie, car les cas sont trop peu nombreux. Par exemple, dans la cohorte PAQUID, quelques cas de sujets atteints de la maladie de Parkinson sont répertoriés (100 patients diagnostiqués lors des suivis), mais leur nombre est insuffisant pour pouvoir généraliser des résultats.

6.1.1.1 Définition et diagnostic

Le Parkinson fait partie des maladies neuro-dégénératives les plus répandues (après la maladie d'Alzheimer). La destruction des neurones y est lente, progressive, et entraîne de nombreuses contraintes et symptômes pour le malade, qui va très souvent nécessiter des soins de longue durée.

La plupart des personnes touchées par le Parkinson sont âgées de 50 ans et plus. Cependant, le nombre de personnes de moins de 40 ans qui développent cette maladie est en hausse.

La carence en dopamine est la principale cause de survenue de cette maladie. Les cellules responsables de la production de la dopamine sont peu à peu détruites entraînant de lourds problèmes moteurs tels que des tremblements, troubles de la parole, raideurs musculaires dans les membres..., et d'autres symptômes secondaires comme la sénilité, les troubles de la parole, la perte de mémoire...

Le diagnostic de la maladie repose sur la détection de l'akinésie (difficulté à initier un mouvement) et d'au moins deux des principaux symptômes : bradykinésie (lenteur des mouvements), raideur ou tremblement au repos. L'asymétrie des symptômes est aussi très courante.

Cependant, beaucoup de cas ne sont pas évidents et demandent alors de rechercher la présence d'autres troubles secondaires. Les symptômes de la maladie évoluent de façon non prévisible et sont particulièrement difficiles à identifier au

début de la maladie. De plus, la plupart des symptômes sont communs à d'autres maladies neuro-dégénératives. Mis à part l'examen anatomo-pathologique du cerveau, il n'existe actuellement aucun examen complémentaire permettant d'établir le diagnostic de la maladie avec certitude.

C'est pourquoi, le diagnostic est souvent tardif, il peut se passer quelques années avant que la maladie soit réellement diagnostiquée chez un sujet.

6.1.1.2 Etat des lieux du Parkinson en France

En France, il y a peu de données sur la fréquence de la maladie de Parkinson. Une étude effectuée à partir des données de remboursement de l'Assurance Maladie, a estimé la prévalence de la maladie de Parkinson à 1 250 cas pour 100 000 personnes [Bertin 2005].

Généralement, cette maladie est considérée comme étant une maladie multifactorielle. Parmi les facteurs les plus récurrents, nous pouvons citer un facteur génétique, mais les études s'intéressent davantage aux facteurs environnementaux tels que les pesticides, le tabagisme et la consommation de café.

Jusqu'à aujourd'hui, la maladie de Parkinson est incurable, car on ne connaît toujours pas les causes de la dégradation des neurones. Il n'existe aucun traitement propre au Parkinson, les malades sont contraints à prendre une combinaison de divers médicaments et à faire des exercices. Une thérapie adaptée peut toutefois influencer de façon positive l'évolution de la maladie, en réduisant les symptômes et en atténuant les troubles.

6.1.2 Données

À notre connaissance, il existe peu d'estimations de l'incidence de la maladie de Parkinson en France. Une alternative aux données de cohorte, pour estimer cette incidence, consiste à estimer le nombre de nouveaux patients traités pour maladie de Parkinson à partir des données du SNIIRAM (Système national d'information inter-régimes de l'Assurance maladie), et plus précisément à partir des bases de remboursements de médicaments (profils de consommation de médicaments antiParkinsoniens spécifiques) de l'Assurance Maladie, disponibles dans le DCIR (Données de consommations inter-régimes).

Le SNIIRAM stocke les données de consommation sur une période de 3 ans plus l'année en cours. Le DCIR est une base de données individuelles et anonymes qui comprend les données de remboursement de soins des bénéficiaires de la grande majorité des régimes d'Assurance Maladie obligatoires, ayant consommés des soins ambulatoires pendant les trois dernières années plus l'année en cours. Les régimes qui contribuent au DCIR couvrent environ 98 % de la population française. Le DCIR comprend aussi des données médicales (données de mise en ALD (affection de longue durée), données des hospitalisations, PMSI (programme de médicalisation des systèmes d'information)).

À partir de diverses données du DCIR, en appliquant des méthodes d'estimations de corrections détaillées en annexe A.4.1, A. Elbaz et F. Moisan ont mis à notre disposition les données suivantes concernant la maladie de Parkinson en France :

- l'incidence du Parkinson en 2010, 2011 et 2012 par tranche d'âge de 5 ans et par sexe
- la prévalence du Parkinson en 2009, 2010 et 2011 par tranche d'âge de 5 ans et par sexe,
- les taux de mortalité des Parkinsoniens en 2010, 2011 et 2012 par tranche d'âge de 5 ans et par sexe.

Ces estimations ont été corrigées (pour la sensibilité) et re-estimées en fonction des valeurs attendues (en raison du sous-diagnostic supposé après 80 ans).

De plus, nous disposons des risques relatifs de décès des Parkinsoniens par rapport aux non-Parkinsoniens par tranches d'âge :

- soit issus de la littérature [Elbaz 2003] avec un risque relatif de 2,04 pour les sujets de moins de 67 ans ; 1,76 pour les sujets d'un âge compris entre 67 et 76 ans ; 1,48 pour les sujets de plus de 76 ans (sans distinction du sexe).
- soit nous les avons calculés en faisant le rapport des taux de mortalité des Parkinsoniens/non-Parkinsoniens par tranche d'âge de 5 ans et par sexe. (Notez que pour avoir ce taux, nous avons dû utiliser les données de l'INSEE concernant la population générale).

Ainsi, nous avons obtenu les risques relatifs contenus dans le tableau 6.1 qui suit :

TABLE 6.1 – Risques relatifs calculés des Parkinsoniens *vs* non-Parkinsoniens de **45 à 89 ans** en France.

classe d'âge	Femme	Homme
[45-50[2.146	2.662
[50-55[2.907	1.980
[55-60[2.727	2.129
[60-65[3.513	2.002
[65-70[3.018	1.859
[70-75[3.289	2.136
[75-80[2.777	2.172
[80-85[2.309	1.909
[85-90[1.734	1.498

Nous pouvons remarquer que les risques relatifs sont plus élevés chez les femmes que chez les hommes. De plus, nos calculs (en utilisant les données fournies) nous donnent globalement des risques plus élevés que ceux trouvés dans la littérature.

Enfin, *via* l'INSEE, nous récupérons exactement les mêmes données et de la même manière que ce que nous avons décrit pour l'application à la démence (section

5.1.3). Il s'agit donc, du taux de mortalité générale, α_2 , de la population française et de la population vivante, ν , par âge, sexe et année de naissance entre 1950 et 2060.

6.2 Modèles et hypothèses

Pour cette maladie, nous avons appliqué un modèle Markovien et un modèle semi-Markovien.

Les hypothèses communes aux deux modèles concernent l'incidence du Parkinson et la mortalité des non-Parkinsoniens.

Dans le cadre de la maladie de Parkinson, nous avons fixé $a_0 = 45$ ans, l'âge à partir duquel les individus sont à risques de développer la maladie de Parkinson. Le nombre de cas de Parkinson plus précoces est assez rare (ils représentent environ 1,3% des cas sur nos données).

De plus, sans autres hypothèses, nous considérons que l'incidence du Parkinson est homogène dans le temps $\alpha_{01}(a)$. Aucune tendance n'a été mise en évidence sur les trois années d'incidence que nous avons et il n'y pas d'hypothèses sûres dans la littérature.

En revanche, contrairement à la démence, il n'est pas nécessaire de distinguer la mortalité des non-malades de la mortalité générale dans la population, dans les modèles. En effet, la prévalence du Parkinson en France étant beaucoup moins importante, nous ne mettons pas en évidence une différence significative entre ces deux mortalités.

C'est pourquoi, pour faciliter la suite, nous préférons la négliger (même si la méthode permet de prendre en compte cette différence).

Il n'y aura donc pas d'équation différentielle à résoudre. Nous considérons directement que $\alpha_{02}(a, t) = \alpha_2(a, t)$.

La mortalité générale dans la population, α_2 , ainsi que la population à risque à 45 ans, $\nu(45, t)$, sont déterminées *via* les données de l'INSEE par sexe et pour chaque année t .

La figure 6.1 décrit donc le modèle mutli-états à partir duquel nous allons distinguer deux cas en fonction de l'hypothèse considérée pour la mortalité des Parkinsoniens.

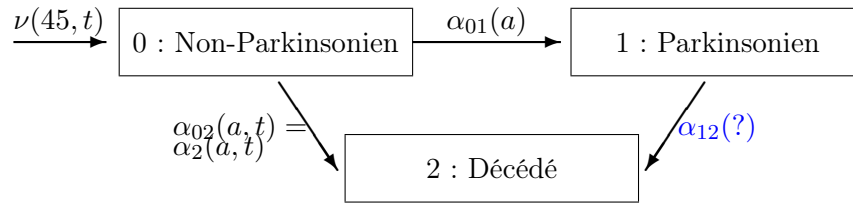


FIGURE 6.1 – Modèle illness-death.

- Modèle Markovien M2.

Le premier modèle que nous appliquons est un modèle Markovien où nous faisons l'hypothèse de non-homogénéité de la mortalité des Parkinsoniens. Ainsi, l'intensité de transition de l'état malade à l'état décédé dépend de l'âge et du temps calendaire : $\alpha_{12}(a, t)$.

D'autre part, nous supposons que cette mortalité, α_{12} , est proportionnelle à la mortalité des non-Parkinsoniens, α_{02} , que multiplie le risque relatif, g , de décès des Parkinsoniens par rapport aux non-Parkinsoniens. Les informations dont nous disposons sur g dépendent de certaines classes d'âge, mais pas du temps calendaire, donc nous considérons le risque relatif $g(a)$ homogène au cours du temps et dépendant seulement de l'âge a des sujets.

De ce fait, nous écrivons $\alpha_{12}(a, t) = \alpha_{02}(a, t)g(a)$, soit, de par l'hypothèse sur la mortalité des non-Parkinsoniens :

$$\alpha_{12}(a, t) = \alpha_2(a, t)g(a) \quad (6.1)$$

- Modèle semi-Markovien S2.

Le deuxième modèle appliqué conserve l'hypothèse de non-homogénéité de la mortalité des Parkinsoniens, mais cette fois-ci, il s'agit d'un modèle semi-Markovien. Ainsi, la mortalité des Parkinsoniens dépend de la durée passée avec la maladie et de l'âge au temps t : $\alpha_{12}(a, t, d)$.

Pour déterminer $\alpha_{12}(a, t, d)$ nous utilisons le modèle additif présenté dans la section 3.3.2 (l'équation 3.5) qui fait intervenir $\alpha_{02}(a, t)$, la mortalité des non-Parkinsoniens, et $\alpha^d(d)$, le sur risque de décès des Parkinsoniens par rapport aux non-Parkinsoniens.

Du fait de l'hypothèse faite sur la mortalité des non-Parkinsoniens, cette équation

devient :

$$\alpha_{12}(a, t, d) = \alpha_2(a, t) + \alpha^d(d) \quad (6.2)$$

6.3 Adaptation de la méthode

La méthode présentée pour estimer les paramètres de ces modèles, dans le chapitre 3, ne peut pas être appliquée telle quelle. En effet, nous n'avons pas de données individuelles sur une cohorte pour pouvoir estimer l'incidence de la maladie de Parkinson ainsi que la mortalité des Parkinsoniens ou des non-Parkinsoniens.

En revanche, nous avons l'incidence (de 2010 à 2012), la prévalence (de 2009 à 2011), et des risques relatifs (qui servent au modèle Markovien que nous allons appliquer).

Nous avons donc adapté la méthode d'estimation des paramètres du modèle. Par contre, la méthode pour faire des projections reste inchangée, nous ne la détaillerons pas de nouveau.

Le travail effectué sur les données démographiques reste le même, à savoir que nous récupérons les données passées et futures de α_2 pour chaque âge a et temps t nécessaires, et de ν à 45 ans pour chaque années t nécessaires. Ensuite, nous lisons les données de la mortalité générale (loi Gompertz) pour avoir une fonction continue sur l'âge.

Pour l'incidence, nous utilisons les trois années dont nous disposons, celles-ci sont lissées sur l'âge (les 3 années confondues) avec des fonctions splines dans le but d'avoir une fonction continue.

Les données de prévalence observées en 2009, 2010 et 2011 sont également lissées avec des fonctions splines pour chaque année (2009, 2010 et 2011) individuellement. Ceci permet d'avoir des valeurs de prévalence pour chaque âge pour ces trois années. Ces données de prévalence vont nous servir de références pour les estimations que nous feront avec nos modèles.

Remarque 1 : pour la démence, nous n'avons pas de données de prévalence, en revanche, nous avons la mortalité des déments et des non-déments estimées sur des données individuelles.

Remarque 2 : nous avons testé différentes méthodes de lissage, autres que les fonctions splines, telles que la loi exponentielle ou la loi de Weibull. Nous n'avons pas relevé de différences, nous avons donc gardé les splines puisque c'est la méthode que nous avons utilisé jusqu'à présent pour estimer les fonctions.

6.3.1 Pour le modèle Markovien

Pour ce modèle, il s'agit donc d'estimer la mortalité des Parkinsoniens et des non-Parkinsoniens. Par hypothèse, la mortalité des non-Parkinsoniens est égale à la mortalité générale α_2 . Nous utilisons donc l'équation 6.1 pour estimer α_{12} . Dans cette équation, il nous reste à déterminer le risque relatif $g(a)$.

Nous avons testé différentes façons de procéder :

- tout d'abord, nous avons choisi d'utiliser les risques relatifs de la littérature, séparés en trois classes d'âge et sans distinction en fonction du sexe,
- puis nous avons utilisé les risques relatifs que nous avons recalculés, par classes d'âge de 5 ans et en distinguant les sexes.

Pour ces deux cas, la démarche est semblable à celle utilisée dans la section 5.2.1 pour le modèle **M2**, mais elle ne nécessite pas de résoudre une équation différentielle pour estimer α_{02} .

Une fois les estimations faites, nous utilisons le modèle pour estimer la prévalence du Parkinson (par âge et/ou totale) en 2009, 2010, et/ou 2011 afin de comparer avec la prévalence observée pour ces années-là.

Les constats faits sur ces premiers essais nous ont poussés à considérer des valeurs différentes pour le risque relatif $g(a)$.

En effet, non satisfaits des résultats aboutissant à une surestimation de la prévalence pour les années connues, nous avons choisi de modifier complètement la phase d'estimation en y intégrant les prévalences observées (lissées pas âge) et en considérant que $g(a)$ est inconnu.

Puis, nous avons mis en place un algorithme de recherche, qui cherche à estimer paramétriquement $g(a)$ avec des fonctions splines au sein du modèle illness-death, de façon à retrouver les valeurs de prévalence observées par âge pour les 3 années dont nous disposons. A chaque itération, $g(a)$ est estimé, puis sur cette base la prévalence est estimée pour 2009, 2010 et 2011 et comparée à la prévalence observée sur ces mêmes années pour chaque âge.

Si les résultats sont trop différents l'algorithme continue et pour minimiser l'écart entre nos prévalences estimées et les prévalences observées, nous utilisons la méthode des moindres carrés avec une pondération sur l'âge. Cette pondération est simplement le rapport du nombre de cas par âge sur le nombre de cas total.

6.3.2 Pour le modèle semi-Markovien

Pour le modèle semi-Markovien, l'objectif est également d'estimer la mortalité des Parkinsoniens et des non-Parkinsoniens. Par hypothèse, la mortalité des non-Parkinsoniens est égale à la mortalité générale α_2 . Nous utilisons l'équation 6.2 pour estimer α_{12} . Dans cette équation, $\alpha^d(d)$, considéré comme le sur-risque de décès des Parkinsoniens par rapport aux non-Parkinsoniens, est inconnu.

La phase d'estimations est modifiée de la même manière que pour le modèle Markovien (en mettant en place un algorithme de recherche), sauf que dans ce cas, c'est le paramètre $\alpha^d(d)$ que nous cherchons à estimer paramétriquement avec des fonctions splines, en fonction de la durée passée avec la maladie de Parkinson, au sein du modèle illness-death.

Le but est, là encore, que nos prévalences estimées par âge avec notre modèle soient comparables aux prévalences par âges observées pour les années 2009, 2010 et 2011. Dans l'algorithme, nous minimisons l'écart entre ces prévalences avec la méthode des moindres carrés pondérée sur l'âge.

Remarque : que se soit pour le modèle Markovien ou semi-Markovien, c'est une fois les paramètres estimés que nous avons fait des projections de la prévalence du Parkinson en 2030 en France.

6.4 Résultats

Dans cette partie résultats, nous nous concentrons uniquement sur les femmes à l'exception de quelques résultats pour présenter la prévalence totale. Le même travail, ainsi que les mêmes observations et constats, ont été effectués pour les deux sexes.

De plus, nous avons fixé un âge maximum de 92 ans pour toutes les estimations et projections. Ceci vient du fait que nos données sources sont très peu informatives après cet âge. Rappelons que ceux qui nous ont fourni les données ont déjà fait un gros travail pour corriger les estimations après 80 ans, à cause du sous-diagnostic très important du Parkinson chez les personnes âgées (entre autres raisons, car la consommation de médicaments antiParkinsonien n'est pas connue chez les malades institutionnalisés), mais ils restent incertains des données sur ces grands âges.

Enfin, pour ce qui est des projections, nous nous sommes limités à la prévalence et nous n'avons pas simulé d'interventions, car nous n'avions pas de scénarios à proposer.

6.4.1 Estimations avec le modèle Markovien

Tout d'abord, nous supposons que le risque relatif de décès des Parkinsoniens par rapport aux non-Parkinsoniens, $g(a)$ (fonction de l'âge), est connu. Nous avons différencié deux façons de le considérer, l'une *via* les données de la littérature, l'autre par des calculs que nous avons effectués. Ainsi, pour chacun de ces cas, nous avons estimé la prévalence avec notre modèle.

Le tableau 6.2 compare la prévalence totale du Parkinson estimée par notre modèle, à la prévalence observée, en 2010 entre 45 et 92 ans, en distinguant deux façons de considérer le risque relatif g .

Les estimations des taux de mortalité par âge des Parkinsoniens sont représentées dans la figure 6.4, pour les femmes, en fonction de la manière dont $g(a)$ est considéré.

TABLE 6.2 – Estimation de la prévalence du Parkinson par sexe en France en 2010, pour des âges compris entre **45 et 92 ans**.

Année	Femme	Homme	Total
Prévalences estimées			
$g(a)$ de la littérature	120 000	101 000	221 000
$g(a)$ calculés	104 500	93 500	198 000
Prévalences observées			
	76 000	76 500	152 500

Nous constatons que l'utilisation de ces risques relatifs, issus de la littérature ou calculés par nous-même, surestiment la prévalence en 2010 à la fois pour les hommes et les femmes. Les risques relatifs de la littérature ne semblent pas adaptés pour estimer la prévalence en fonction de l'incidence donnée.

Ceci peut s'expliquer par le fait que ces risques relatifs ne sont pas propres à la population française, ne distinguent pas les hommes des femmes et sont pour des classes d'âge trop larges. Ils sont donc peut être trop faibles.

C'est pourquoi, nous avons souhaité calculer $g(a)$ en utilisant les données françaises à disposition (les taux de mortalité de l'INSEE et les taux de mortalité des Parkinsoniens), en distinguant les hommes des femmes et en faisant des classes d'âge de 5 ans. Nous avons déjà pu constater que les risques relatifs calculés sont plus élevés que ceux de la littérature.

Les résultats obtenus donnent des prévalences plus proches, mais restent tout de même trop supérieures aux prévalences observées.

En conséquence, nous avons supposé que le problème venait des risques relatifs puisqu'ils ne permettent pas de lier les incidences données aux prévalences observées.

C'est pourquoi nous avons modifié la méthode d'estimation et mis en place l'algorithme de recherche de $g(a)$ décrit dans la section 6.3.1.

La figure 6.2 représente le risque relatif de décès des Parkinsoniens par rapport aux non-Parkinsoniens en fonction de l'âge à l'issue de cette nouvelle phase d'estimation. Il s'agit des estimations pour les femmes entre 70 et 90 ans.

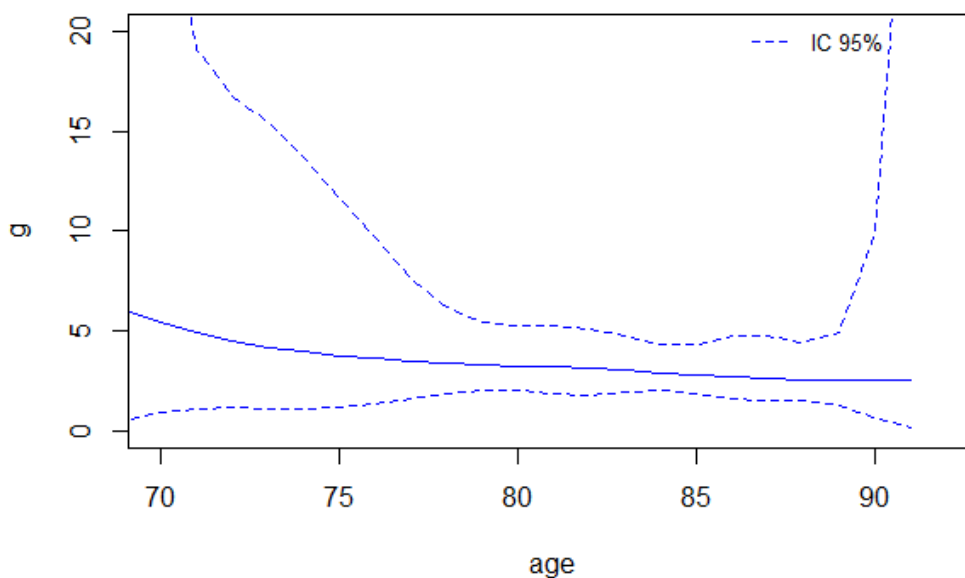


FIGURE 6.2 – Estimations du risque relatif de décès des Parkinsoniens par rapport aux non-Parkinsoniens par âge pour les femmes et IC à 95%, (modèle **M2**).

Nous observons que d'après cette procédure, les risques relatifs nécessaires pour faire correspondre les incidences et prévalences que l'on nous a données sont beaucoup plus élevés que ceux utilisés jusqu'à présent. Les résultats sont même démesurés avant 70 ans (voir l'annexe A.4.2).

Cependant pour ces âges jeunes il n'y a pas beaucoup de cas, ce qui peut expliquer ces résultats, et les valeurs à ces âges ont peu d'impact pour le calcul de la prévalence.

Pour le vérifier, nous avons décidé de mettre une contrainte sur $g(a)$ pour tous les âges inférieurs à 70 ans. Si le $g(a) > 20$ alors nous imposons $g(a) = 20$ (qui est déjà une valeur trop élevée).

La figure 6.3 représente l'estimation de la prévalence par âge en distinguant un cas avec $g(a)$ qui vient de la méthode d'estimation et l'autre cas lorsqu'on ajoute la contrainte.

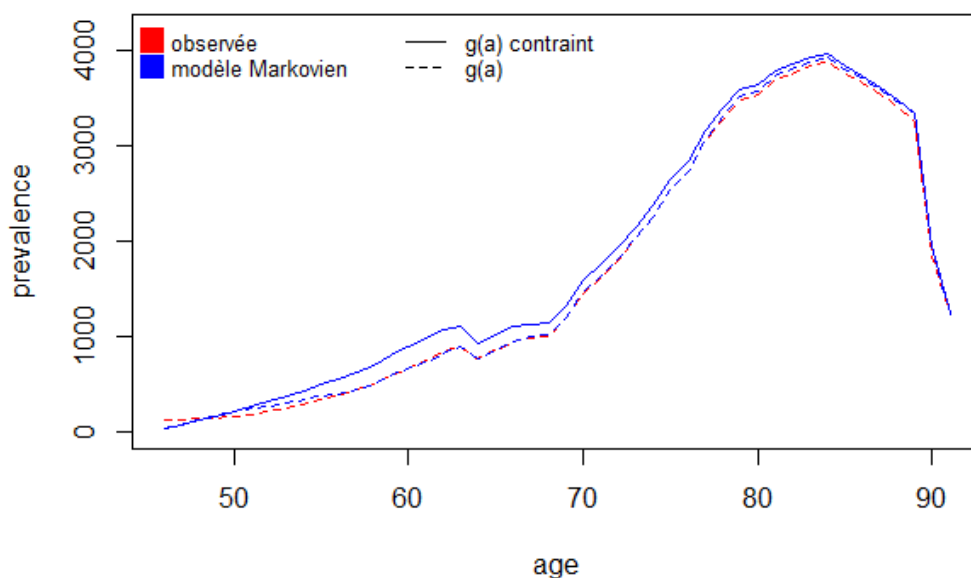


FIGURE 6.3 – Estimations de la prévalence du Parkinson en fonction de l'âge pour les femmes en 2010 en fonction de l'hypothèse sur le risque relatif $g(a)$.

Cette figure montre que la contrainte modifie relativement peu l'estimation de la prévalence en 2010. En effet, la prévalence totale chez les femmes sans la contrainte sur $g(a)$ est proche de 77 500 cas contre 81 000 cas avec la contrainte (rappelons que la prévalence totale observée des femmes s'élèvent à environ 76 000 cas en 2010, sur cette même tranche d'âge). Nous verrons plus tard que les

projections sont également peu modifiées.

Cette dernière figure 6.4 résume l'estimation de la mortalité des Parkinsonien avec le modèle en distinguant les quatres manières de considérer le risque relatif $g(a)$ que nous venons de détailler.

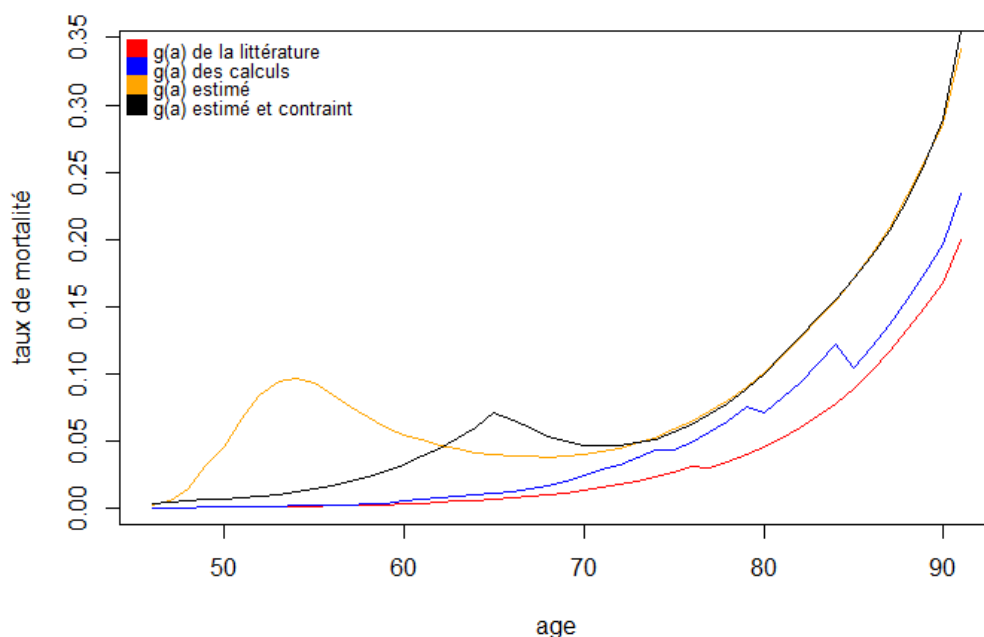


FIGURE 6.4 – Estimations de la mortalité des Parkinsoniens pour les femmes, modèle **M2**.

Un premier constat est que la mortalité des Parkinsoniens, estimée avec des risques relatifs issus de la littérature ou calculés sur les données fournies, est plus faible que la mortalité des Parkinsoniens lorsque les risques relatifs sont supposés inconnus et sont estimés *via* l'algorithme de recherche. Cela explique donc que pour les deux premiers cas, nous sur-estimons la prévalence en 2010 puisque la mortalité des Parkinsoniens est beaucoup plus faible.

D'autre part, nous constatons un pic de mortalité pour les âges les plus jeunes lorsque nous estimons $g(a)$ *via* l'algorithme (contraint ou non), ceci n'est pas une erreur puisqu'il reflète le fait que nous avons estimé des risques relatifs très élevés pour ces âges.

Puisque, clairement, nous avons un problème de modélisation de la mortalité pour les sujets jeunes, il a paru important de prendre en compte la durée passée avec la maladie, en plus de l'âge, dans les taux de décès des malades. C'est pourquoi,

utiliser un modèle semi-Markovien nous a semblé être la solution.

6.4.2 Estimations avec le modèle semi-Markovien

Pour ce modèle, il s'agit dans un premier temps d'estimer le sur-risque de décès $\alpha^d(d)$ des Parkinsoniens par rapport aux non-Parkinsoniens qui dépend de la durée passée avec la maladie. Pour cela, nous utilisons l'algorithme de recherche que nous avons développé (section 6.3.2) afin de donner une estimation de la prévalence du Parkinson en 2010, avec notre modèle, proche de la prévalence observée cette même année. Cette fonction nous permet donc d'estimer la mortalité des Parkinsoniens en fonction de la durée passée avec la maladie et de l'âge au temps t via l'équation 6.2.

Nous avons rencontré les mêmes problèmes d'estimations qu'avec le risque relatif $g(a)$ pour le modèle Markovien. En effet, $\alpha^d(d)$ est très faible pour les premières années passées avec la maladie puis "explose" après une durée de 10 ans.

Ainsi, ce problème se retrouve dans l'estimation de la mortalité des Parkinsoniens qui va rapidement devenir très importante, comme le montre la figure 6.5 qui représente la mortalité des Parkinsonien pour les femmes de 60 ans, en fonction de la durée passée avec la maladie.

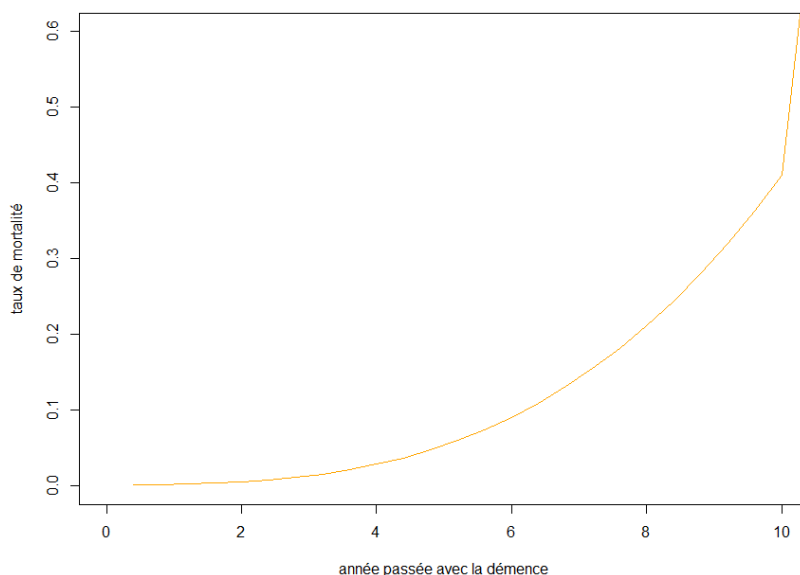


FIGURE 6.5 – Estimations de la mortalité des Parkinsoniens en fonction de la durée passée avec la maladie, pour les femmes de 60 ans.

Cette courbe laisse penser qu'un sujet ne vivrait pas plus de 10 ans avec la

maladie de Parkinson.

Cependant, bien que ces estimations soient peu conformes à ce qui est connu pour cette pathologie, la prévalence du Parkinson estimée sous ce modèle est de 76500 cas pour les femmes en 2010, ce qui est semblable à la prévalence observée la même année.

6.4.3 Projections

Bien que les estimations ne soient pas encore satisfaisantes, nous avons fait des projections de la prévalence du Parkinson en 2030 pour les femmes.

Nous présentons les résultats pour le modèle semi-Markovien et le modèle Markovien avec $g(a)$, non contraint, considéré comme inconnu et recherché par l'algorithme, car c'est celui-là qui estime le mieux la prévalence observée en 2010.

La figure 6.6 décrit la prévalence du Parkinson par âge chez les femmes en 2010 et en 2030, estimée avec deux modèles et comparée à la prévalence observée en 2010.

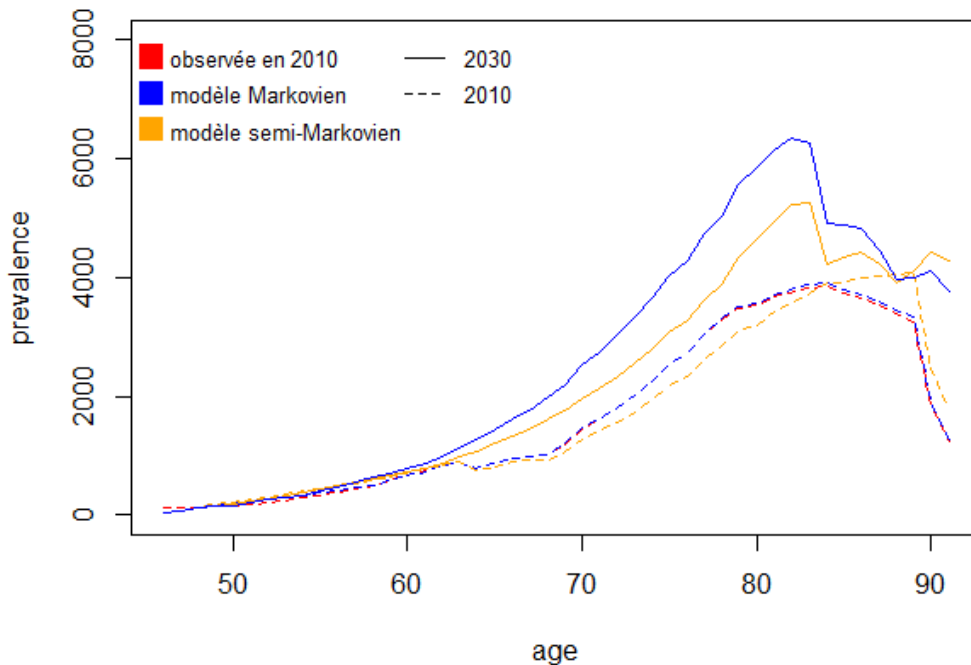


FIGURE 6.6 – Estimations de la prévalence du Parkinson en 2010 et projections en 2030 en fonction des modèles, pour les femmes, modèles **M2** et **S2**.

Sur cette figure, nous constatons que la prévalence par âge observée en 2010 et presque identique à la prévalence estimée sous le modèle Markovien en 2010. Le modèle semi-Markovien semble un peu moins proche des données en 2010 bien que nous ayons vu que la prévalence totale estimée est égale à la prévalence totale observée pour cette année.

Concernant les projections pour 2030, le modèle Markovien donne presque systématiquement des prévalences par âge plus importantes que le modèle semi-Markovien.

Le tableau 6.3 donne les prévalences totales pour les estimations en 2010 et les projections en 2030 pour les femmes

TABLE 6.3 – Estimations de la prévalence du Parkinson en 2010 et projections 2030, pour les femmes en France âgées entre **45 et 92 ans**.

Prévalence	2010	2030
Observée	76 500	-
Estimée : Modèle Markovien	77 500	117 000
Estimée : Modèle semi-Markovien	76 500	100 500

Si l'on s'intéresse à la prévalence totale, alors nous observons que le modèle semi-Markovien donne une meilleure estimation en 2010. De plus, les projections en 2030 de la prévalence du Parkinson chez les femmes diffèrent de façon importante en fonction du modèle. Sous le modèle Markovien, nous nous attendons à une augmentation de 50% du nombre de cas de Parkinson en 2030 contre une augmentation de 30% de cas sous le modèle semi-Markovien.

6.5 Conclusion sur le Parkinson

Pour cette application, les données mises à notre disposition ne permettaient pas d'utiliser la méthode (qui a pu être appliquée à la démence) telle quelle. Une adaptation était nécessaire, de nouvelles stratégies ont été développées, et cela nous a permis d'envisager différemment la méthode, tout en gardant le même objectif de fournir des projections.

Ainsi, pour la maladie de Parkinson, nous avons testé deux types de modèle, un Markovien et un semi-Markovien, sans pour autant avoir trouvé le modèle optimal pour l'instant.

Nous avons donné des projections de la prévalence de la maladie en 2030 et constaté que celle-ci augmente davantage plus avec le modèle Markovien qu'avec le modèle semi-Markovien.

Cependant, dans l'état actuel, nous ne sommes pas satisfaits des modèles lorsque nous estimons la prévalence par âge et totale de la maladie en 2010 et que

nous la comparons aux données de prévalence observée pour cette même année.

C'est pourquoi, en conclusion sur cette application, nous allons discuter des problèmes et des pistes à explorer.

6.5.1 Incidence et Prévalence observées

Pour le modèle Markovien, nous avons essayé plusieurs manières de considérer le risque relatif de décès des Parkinsoniens par rapport aux non-Parkinsoniens. Tout d'abord, nous avons utilisé des risques relatifs issus de la littérature, puis des risques relatifs par classe d'âge et en fonction du sexe que nous avons calculés grâce aux données fournies. Ensuite, nous avons donné des projections de la prévalence de la maladie, mais aucune de ces deux façons de considérer le risque relatif n'a permis d'estimer une prévalence proche de la prévalence observée. Ces risques relatifs semblent trop faibles pour permettre de faire correspondre l'incidence et la prévalence fournies.

Nous avons donc décidé de considérer les risques relatifs par âge inconnus et nous avons cherché ces valeurs de façon à obtenir (à partir de l'incidence donnée) une prévalence estimée proche de la prévalence observée.

Bien que nous arrivions à approcher de façon satisfaisante la prévalence observée, les risques relatifs que cela implique sont beaucoup trop élevés, surtout pour les âges les plus jeunes.

Les différents essais et résultats obtenus nous ont amenés à penser qu'avec le modèle Markovien le problème venait d'une mauvaise adéquation pour les âges jeunes.

Pour le modèle semi-Markovien, nous avons rencontré le même type de problème lorsque nous cherchions à estimer le sur-risque de décès des Parkinsoniens en fonction de la durée passée avec la maladie, dans le but d'obtenir une prévalence estimée proche de la prévalence observée. Or, dans ce cas, il semble que se sont les âges les plus élevés que nous représentons moins bien avec ce type de modèle (figure 6.6).

Finalement, pour cette application, nous avons supposé que les données d'incidence et de prévalence observées (fournies par A. Elbaz) étaient correctes, mais nos résultats remettent cette hypothèse en doute.

En effet, en utilisant les données observées, nous avons tracé le rapport de la prévalence sur l'incidence et obtenu les courbes de la figure 6.7

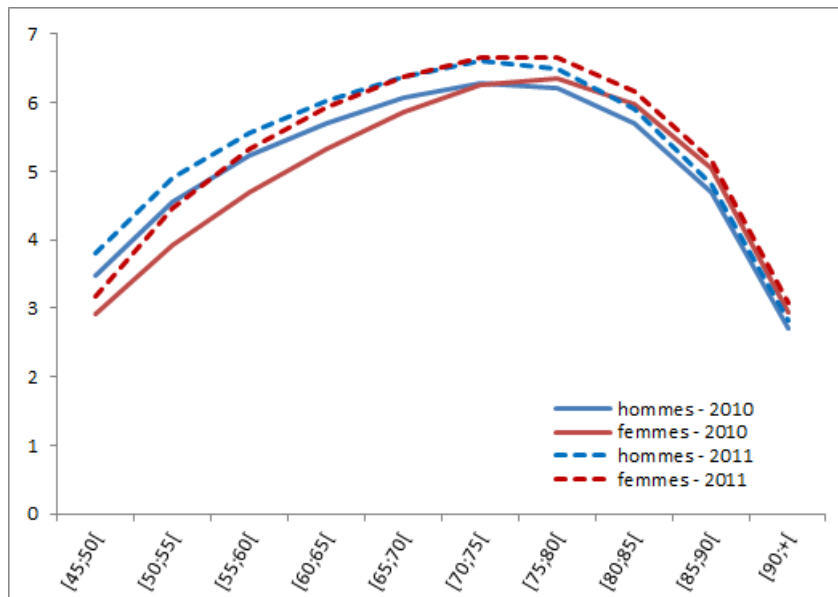


FIGURE 6.7 – Rapport de la prévalence sur l'incidence des données observées pour le Parkinson, en 2010 et 2011, pour les hommes et les femmes.

Quels que soient l'année ou le sexe, ce rapport suppose une mortalité élevée et une espérance de vie courte pour les âges les plus jeunes. C'est aussi le cas pour les âges supérieurs à 85 ans, mais là il n'y a rien d'étonnant à ces âges.

Ceci explique que lorsque nous avons cherché à faire coïncider l'incidence fournie et la prévalence observée, nous avons trouvé des risques relatifs très élevés pour les âges jeunes avec le modèle Markovien, et un sur-risque de décès qui augmente très rapidement en fonction de la durée passée avec la maladie, avec le modèle semi-Markovien.

Finalement, il est possible que soit l'incidence soit la prévalence observée n'ait pas été correctement estimée.

6.5.2 Perspectives

6.5.2.1 Nouvelle stratégie

Bien que les données mises à notre disposition ont été beaucoup travaillées par A. Elbaz et F. Moisan afin de corriger les biais, sous diagnostic,...etc, nous pensons quand même que cela vaut la peine de changer de stratégie.

En tenant compte de la difficulté du diagnostic de la maladie et du peu de recul sur les données (3 ans de suivis), il est crédible de penser qu'il y a un problème dans l'estimation de l'incidence de la maladie de Parkinson.

Nous pensons que certains cas "incident" sont en fait des cas "prévalents".

Ainsi, une des pistes que nous envisageons pour le vérifier est de considérer que

l'incidence est inconnue. Il s'agira, donc d'utiliser les risques relatifs donnés (soit par la littérature soit ceux que nous avons calculés) ainsi que la prévalence observée et d'en déterminer l'incidence de la maladie de Parkinson.

6.5.2.2 Autre modèle

Suite aux discussions que nous avons eu avec A.Elbaz sur les résultats et sur l'épidémiologie et l'évolution de la maladie de Parkinson, nous avons conclu que pour cette pathologie, il est important de prendre en compte la durée passée avec la maladie dans l'estimation de la mortalité des Parkinsoniens. Un modèle semi-Markovien serait donc plus adapté.

Quant aux hypothèses d'une incidence homogène au cours du temps et d'une mortalité des non-Parkinsoniens équivalente à la mortalité générale de la population, elles n'ont pas besoin d'être modifiées.

Le modèle semi-Markovien que nous avons proposé incluant un modèle additif pour estimer la mortalité des Parkinsoniens est préféré à un modèle semi-Markovien simple, car nous faisons aussi intervenir l'âge des sujets qui est également important pour cette pathologie.

Nous avons testé plusieurs façons de modéliser la mortalité des Parkinsoniens : modèle additif, multiplicatif, exponentiel..., et nous avons retenu ce modèle additif qui semble plus intuitif et plus facilement interprétable. De plus, en fonction du modèle, la résolution de l'équation différentielle n'est pas toujours simple.

C'est pourquoi nous proposons de continuer à modéliser la mortalité des Parkinsoniens avec un modèle additif, mais à la vue de nos résultats, nous pensons que l'âge et la durée passée avec la maladie n'impactent pas de la même façon la mortalité des Parkinsoniens.

L'effet de la durée passée avec la maladie serait plus important que l'effet de l'âge pour les classes d'âge les plus jeunes, alors que pour les classes d'âge les plus âgées ce serait l'inverse.

Nous envisageons donc un modèle qui couple à la fois le modèle Markovien et le modèle semi-Markovien, par exemple :

$$\alpha_{12}(a, t, d) = \alpha_2(a, t)g(a) + \alpha^d(d)$$

auquel on ajouterait un poids qui augmente en fonction de l'âge. Sur notre exemple on aurait donc un modèle de ce type :

$$\alpha_{12}(a, t, d) = \rho(a)\alpha_2(a, t)g(a) + (1 - \rho(a))\alpha^d(d) \quad (6.3)$$

Bien sûr, cette réflexion reste à mûrir afin de déterminer un modèle qui soit estimable et interprétable par la suite.

Conclusions et perspectives

Sommaire

7.1	Résumé	124
7.2	Discussion	125
7.2.1	Modélisation d'une intervention	125
7.2.2	Développement d'un modèle semi-Markovien	127
7.2.3	Modèle Markovien ou semi-Markovien	128
7.3	Perspectives	129
7.3.1	Développements des modèles	129
7.3.2	Approche médico-économique	130
7.3.3	Généralisation et utilisation	131

7.1 Résumé

L'objectif principal de cette thèse était de développer des modèles multi-états pour faire des projections concernant plusieurs indicateurs de santé pour les maladies chroniques.

Pour se faire, nous avons fixé plusieurs objectifs secondaires que nous allons brièvement rappeler, il s'agissait de :

- relâcher certaines hypothèses sur les paramètres/fonctions de la méthode afin de rendre le modèle plus flexible,
- rendre notre modèle capable d'évaluer l'impact d'une intervention, ciblant les facteurs de risque d'une maladie chronique, sur les projections des indicateurs de santé,
- élargir les projections à divers indicateurs de santé,
- appliquer l'approche proposée à d'autres maladies chroniques pour vérifier son adaptabilité.

Ainsi, dans un premier temps, nous avons commencé par assouplir l'hypothèse sur l'incidence de la maladie qui jusque-là était systématiquement considérée homogène dans le temps, donc estimée en fonction de l'âge seulement. Dorénavant, nous pouvons considérer une évolution de l'incidence également au cours du temps et faire des projections. Pour le moment, nous manquons d'hypothèses sur l'évolution possible de l'incidence dans le temps. Nous avons donc proposer deux évolutions, soit due à un "effet cohorte" (conséquence d'une amélioration du niveau d'éducation par exemple), soit due à un "effet calendaire" (conséquence d'une prise en charge par exemple), pour illustrer la méthode.

Ensuite, nous avons développé une autre modélisation de la mortalité des malades de façon à ce qu'elle puisse dépendre de l'âge et de la durée passée avec la maladie. Nous avons proposé une modélisation semi-Markovienne qui permet de faire dépendre la mortalité des malades à la fois de la durée passée avec la maladie et de l'âge des sujets, tout en prenant en compte une évolution naturelle de la mortalité au cours du temps. La méthode permet donc maintenant de considérer la mortalité des malades soit en fonction de l'âge, soit en fonction de l'âge et du temps calendaire, soit en fonction de l'âge et de la durée (et implicitement du temps calendaire).

En outre, les travaux que nous avons menés permettent à présent d'évaluer l'impact d'une intervention bien plus générale que celle qui était proposée jusqu'à présent sur ce modèle.

En effet, nous considérons l'impact d'une intervention en prenant en compte le fait que le facteur de risque a un effet sur le risque de développer la maladie, mais peut également avoir un effet sur la mortalité, qui plus est, l'effet peut être différent chez les malades et chez les non-malades.

De plus, nous ne sommes pas restrictifs sur la population ciblée par l'intervention, il peut s'agir de toute la population (des plus de a_0 ans), d'une tranche d'âge ou d'un âge spécifique.

Enfin, la méthode permet de tester un large panel de scénarios et les hypothèses concernant les paramètres de l'intervention sont relativement flexibles, ils peuvent dépendre de l'âge, de la période ou de la génération.

D'autre part, nous avons élargi l'éventail des indicateurs de santé sur lesquels nous donnons des projections futures, afin de mieux évaluer le fardeau de la maladie. Les travaux précédents proposaient uniquement des projections de prévalence, mais une augmentation du nombre de malade n'est pas forcément négative si elle s'accompagne d'un allongement de l'espérance de vie sans la maladie ou d'un âge moyen d'entrée dans la maladie plus tardif. De surcroît, il est d'autant plus important d'avoir des projections de plusieurs indicateurs de santé lorsqu'il s'agit de comparer différents scénarios d'intervention.

Enfin, dans le but de généraliser la méthode en l'appliquant à d'autres maladies, nous avons modifié la méthode d'estimation pour pouvoir appliquer notre modèle même si les données initiales sont différentes de celles dont nous avons l'habitude.

Dans le cadre de cette thèse, les modèles ont été développés pour la démence, mais ont également été appliqués à la maladie de Parkinson. Les principaux résultats de ces applications ont été discutés respectivement dans la section 5.6 pour la démence et 6.5 pour la maladie de Parkinson.

Nous allons donc pour conclure, discuter de ce travail de manière plus générale, énoncer les avantages et les limites de ce que nous avons développé, puis ouvrir sur des perspectives ou des développements que nous avons envisagés, mais pas eu le temps de mettre en place.

7.2 Discussion

7.2.1 Modélisation d'une intervention

Pour prendre en compte l'impact de l'intervention que nous proposons dans la méthode, la difficulté venait du fait que cela demande de modifier à la fois la phase d'estimation et la phase de projection du modèle. De plus, il est nécessaire de déterminer l'état des sujets au moment de l'intervention. En effet, nous devons considérer que la proportion d'exposés/non-exposés au facteur de risque peut être modifiée après l'intervention, mais également, nous devons distinguer ces deux groupes, car l'impact de l'intervention est différent selon le statut d'exposition (il n'y a un effet que sur les exposés au facteur de risque). De la même façon, nous devons distinguer les malades des non-malades au moment de l'intervention, car l'impact de l'intervention est différent selon le statut de la maladie des sujets exposés.

En ce qui concerne les modifications apportées à la méthode, il a été proposé par Joly *et al.* [Joly 2013] d'utiliser un modèle à intensités proportionnelles pour

chaque intensité de transition du modèle illness-death. Celui-ci, décrit en section 4.2, permet de différencier les exposés des non-exposés au facteur de risque dans la cohorte, ainsi que de modifier l'effet du facteur de risque après l'intervention sur chaque intensité de transition de manière indépendante. *In fine*, les projections seront donc une combinaison d'exposés et de non-exposés.

Cependant, cette méthode a été pensée pour une intervention qui vise un âge, et plus précisément un âge où tous les individus sont non-malades. Or, ce n'est pas le cas de notre intervention, c'est pourquoi, nous avons dû, sur cette base, généraliser la méthode.

Nous avons donc fait apparaître dans les expressions mathématiques, le paramètre ω qui permet de prendre en compte la modification de la proportion d'exposés/non-exposés après l'intervention, ainsi que le temps τ de l'intervention qui permet de connaître l'âge des sujets au moment de l'intervention et de distinguer l'effet de l'intervention en fonction de l'état malades/non-malades des sujets à ce moment-là. Les formules ont donc toutes été complexifiées dans le but d'avoir une intervention beaucoup plus générale.

L'approche proposée a de nombreux avantages.

La mortalité y est "soigneusement" modélisée, car il est possible de distinguer la mortalité des non-malades de la mortalité générale dans la population, ce qui s'avère essentiel pour les maladies chroniques qui ont une forte prévalence.

Notez également qu'en fonction du modèle illness-death à appliquer, nous sommes capables de prendre en compte l'âge des sujets et le temps calendaire, pour chaque intensité de transition, et même la durée passée avec la maladie pour la mortalité des malades.

À notre connaissance, tout cela n'est pas pris en compte dans les autres études qui modélisent l'impact d'une intervention.

De plus, l'impact de l'intervention est implémenté dans le but de considérer à la fois l'effet du facteur de risque sur la mortalité (en distinguant malades et non-malades) et sur l'incidence de la maladie. Or, la plupart des études qui cherchent à évaluer l'impact d'interventions ne considèrent pas l'effet sur la mortalité.

Enfin, comparé aux précédents travaux [Brookmeyer 2007], [Barnes 2011], [Jacqmin-Gadda 2013], nous avons proposé et implémenté une intervention plus réaliste qui permet de cibler des sujets de tous âges, ou de réduire à certaines classes d'âge. En effet, pour la démence, il était plus approprié et plus éthique que l'intervention touche toute la population à risque de développer la maladie.

D'autre part, nous avons élargi les projections en variant les indicateurs de santé ce qui donne une meilleure évaluation du fardeau futur de la maladie.

Bien que l'intervention soit assez souple, il reste des limites, notamment, nous ne pouvons pas encore considérer un facteur de risque qui serait acquis après l'âge a_0 (âge à partir duquel les sujets sont à risque de développer la maladie). Cela impliquerait un modèle beaucoup plus complexe.

De plus, l'utilisation d'un modèle à intensités proportionnelles pour prendre en

compte l'impact d'une intervention, suppose que nous avons un effet proportionnel pour chaque risque relatif θ_{ij} , mais nous pouvons mettre en place une modélisation différente si une autre hypothèse s'avère plus pertinente.

L'impact d'une intervention a été appliqué à des données françaises sur la démence sous le modèle **M2**. Les hypothèses que nous avons imposées pour l'application dans ce contexte sont flexibles, et il est évidemment possible d'évaluer l'impact de l'intervention que nous avons proposée sur les autres modèles **M1**, **S1** et **S2** (la méthode est la même, mais elle n'a été implémentée que pour les modèles Markoviens pour l'instant).

Ce travail a fait l'objet d'un premier article [[Wanneveich 2016](#)].

7.2.2 Développement d'un modèle semi-Markovien

Pour que notre modèle puisse s'adapter à d'autres pathologies, il était nécessaire de prendre en compte dans la méthode une évolution de la mortalité des malades en fonction de la durée passée avec la maladie. En effet, pour ces pathologies, il est assez fréquent de constater une hausse de la mortalité des malades proportionnelle au nombre d'années passées avec la maladie.

Jusqu'à présent, la méthode était développée dans le cadre d'un modèle illness-death Markovien (non-homogène), nous devons donc modifier la méthode pour ensuite faire des projections dans le cadre d'un modèle illness-death semi-Markovien.

Pour commencer, nous avons choisi une modélisation simple, en ne faisant dépendre la mortalité des malades que de la durée passée avec la maladie. Mais ceci n'était pas une finalité, car nous savions pertinemment que l'impact de l'âge sur la mortalité est important. C'est pourquoi, ceci n'était qu'une étape pour développer un modèle plus adapté où la mortalité des malades dépend à la fois de l'âge et de la durée passée avec la maladie.

Notre première difficulté était donc de choisir ce modèle.

Nous en avons testé plusieurs, des modèles additifs, des modèles multiplicatifs ou des modèles plus complexes. Cependant, le développement de modèles plus complexes a été principalement gêné par les données qui ne permettaient pas toujours de les estimer, à cause de problèmes d'identifiabilité, ceci même pour un modèle relativement simple comme $\alpha_{12}(a, t, d) = \alpha_{12}(a, t) + \alpha_{12}(d)$.

Pour chaque modèle où nous le pouvions, nous avons fait des essais sur les données de la cohorte PAQUID en estimant les paramètres par vraisemblance pénalisée et approximation par fonctions splines ou bien avec une loi de Weibull.

Tout ceci a fait l'objet d'un long travail de comparaison et de recherche des paramètres de lissage, qui n'est pas détaillé dans cette thèse.

Rappelons, de plus, que le processus d'estimation des paramètres du modèle combine des estimations de données de cohorte (ou autre *cf.* Parkinson) et de données démographiques nationales, avec des projections de la mortalité générale et de

la taille de la population, et que cela nécessite de résoudre une équation différentielle.

Finalement, nous avons choisi de modéliser la mortalité des malades par un modèle additif. Dans ce modèle, nous faisons l'hypothèse que l'âge a le même effet sur la mortalité des non-malades et des malades, mais nous ajoutons un sur-risque de décès pour les malades qui est fonction de la durée passée avec la maladie. Ce modèle prend également en compte une évolution au cours du temps, mais semblable à la population générale. En plus du fait que ce modèle estime correctement les données de la cohorte, il reste intuitif et interprétable, et la résolution de l'équation différentielle est analytique.

Encore une fois la méthode permet de distinguer la mortalité des non-malades de la mortalité générale dans la population si la maladie le requiert. Ainsi, quand la prévalence de la maladie est faible ou que l'excès de mortalité est modeste, des simplifications sont possibles en négligeant cette différence.

Cependant, nous observons que même avec ce modèle intuitif, ajouter un paramètre de temps, pour modéliser la mortalité des malades, complexifie son interprétation. En effet, nous devons combiner à la fois un effet de l'âge, de la durée et du temps. De plus, comme nous utilisons différentes sources de données, certains types de corrélations ou dépendances ne sont pas prises en compte.

Ce modèle semi-Markovien a été appliqué à la fois pour la démence et pour la maladie de Parkinson pour des projections en 2030.

Ce travail est l'objet d'un second article, actuellement en révision dans la revue *Population Health Metrics*.

7.2.3 Modèle Markovien ou semi-Markovien

Tout au long de ce travail, nous avons constamment distingué ces deux types de modèles, nous avons mené les recherches en parallèle et fait les applications sur l'un et l'autre. Dans la méthode d'estimation, seule la façon de considérer la mortalité des malades diffère réellement entre ces deux modèles. En effet, les hypothèses que nous pouvons faire sur l'incidence de la maladie et la mortalité des non-malades sont les mêmes. De même, la prise en compte de l'impact d'une intervention et la possibilité de donner des projections de divers indicateurs de santé sont possibles pour chacun des modèles. Cependant, nous avons pu noter que la modélisation différente de la mortalité des malades entraîne une estimation différente des autres fonctions du modèle, notamment de l'incidence (cf. figure 5.8).

Quoiqu'il en soit, nous ne comparons pas les modèles sur des critères statistiques, au contraire, comme nous l'avons développé dans la section 5.5, le choix du modèle relève davantage de la pathologie sur laquelle nous voulons faire des projections et de son évolution d'un point de vue épidémiologique. Ainsi, le choix du modèle dépend essentiellement du paramètre de temps qui influence majoritairement l'évolution de la maladie.

Le modèle Markovien que nous avons développé est adapté aux maladies où l'âge et le temps calendaire jouent un rôle majeur. Alors que, le modèle semi-Markovien que nous proposons est davantage adapté aux maladies pour lesquelles la durée passée dans l'état malade impacte fortement la mortalité. Donc, ces modèles sont comparables en terme de flexibilité, mais accordent de l'importance à des temps différents.

Cependant, nous devons également considérer les données dont nous disposons pour choisir un modèle. Si nous manquons d'informations ou si les données sont incomplètes ou trop peu informatives cela limite le choix du modèle. En effet, si nous envisageons des modélisations plus complexes de la mortalité des malades, nous pouvons être freinés par les données qui ne permettent pas de tout estimer.

Les résultats des applications montrent l'importance de considérer plusieurs temps (âge, durée passée avec la maladie, temps calendaire) pour modéliser la mortalité des malades.

L'ensemble de ces travaux met en évidence l'importance du choix du modèle et des hypothèses, car en fonction de cela, les variations dans les projections peuvent être de grandes ampleurs.

7.3 Perspectives

7.3.1 Développements des modèles

Au cours de cette thèse, les différents développements entrepris et achevés ont permis d'améliorer l'utilisation et les possibilités de ces modèles afin de fournir des projections. Il n'en demeure pas moins que nos recherches nous ont amenés à envisager d'autres développements, principalement sur la prise en compte d'une intervention et sur la modélisation de la mortalité des malades.

Concernant la prise en compte de l'impact de l'intervention, nous envisageons de développer davantage la méthode ou d'y apporter des modifications ou des ajustements, car même si nous proposons une intervention plus générale et plus flexible que ce qui était fait jusqu'à présent, nous restons limités sur certains points (cf section 7.2.1). De plus, l'intervention a lieu à un moment donné et ses effets perdurent dans le temps, mais nous pourrions imaginer quelle soit récurrente ou autre.

D'autre part, nous pourrions envisager différemment la modélisation de la mortalité des malades, notamment dans le contexte du modèle semi-Markovien.

Nous avons déjà testé plusieurs façons de modéliser la mortalité des malades, en fonction de l'âge et de la durée passée avec la maladie, et nous avons présenté seulement le modèle **S2**. Ce choix a été expliqué de par le fait qu'il s'agissait d'un modèle intuitif, interprétable et avec une résolution analytique de l'équation différentielle

pour estimer la mortalité des non-malades. En effet, pour beaucoup d'autres modèles envisagés, l'équation différentielle n'avait pas de solution simple, ce qui rendait la procédure d'estimation beaucoup plus lourde (dans le cas où la différence, entre la mortalité des non-malades et la mortalité générale dans la population, ne peut pas être négligée) ; et pour d'autres, nous rencontrons des problèmes d'identifiabilité.

Cependant, il serait intéressant de continuer de développer ces modèles qui modélisent la mortalité des malades. En effet, suite à l'application sur la maladie de Parkinson, nous avons constaté que les effets de l'âge et de la durée passée avec la maladie, sur la mortalité des malades, semblent être différents en fonction de l'âge du malade. Ceci pourrait s'avérer être le cas également pour d'autres pathologies.

C'est pourquoi, une modélisation de la mortalité des malades qui propose de mettre plus de poids sur l'effet de l'âge ou sur l'effet de la durée passée avec la maladie, en fonction de l'âge du malade ou en fonction du temps, serait intéressante à développer. Nous envisageons donc une nouvelle modélisation de la mortalité des malades qui couple à la fois un modèle Markovien et un modèle semi-Markovien. Nous ne pouvons pas parler simplement d'un modèle semi-Markovien, car dans ce cas, nous pourrions avoir un modèle qui alterne les hypothèses (Markovienne/semi-Markovienne) et par exemple si le poids attribué à l'effet de la durée est nul pour certains âges, nous serons sous une hypothèse Markovienne.

7.3.2 Approche médico-économique

Les conséquences socio-économiques des maladies chroniques, telles que la démence, sont importantes en termes de coûts médicaux, de dépenses sociales et de frais engagés pour les soins. Par exemple, en 2010, l'OMS estimait le coût sociétal total de la démence dans le monde à 604 milliards de dollars¹, soit 1% du produit intérieur brut (PIB) mondial, ce coût étant plus important dans les pays à revenus élevés.

Avec notre modèle de projection, nous travaillons beaucoup en fonction des temps de séjours du sujet dans les différents états. Ainsi, il est possible d'envisager une approche économiste afin de réaliser des projections de coûts selon des scénarios. Pour cela, il faudrait définir de nouveaux indicateurs, avoir une estimation du coût en fonction de différents paramètres, et proposer des scénarios quant à l'évolution de ces coûts pour les années à venir. De plus, nous pourrions faire des projections avec et sans interventions.

D'autre part, il serait intéressant de pondérer nos projections d'espérances de vie avec ou sans maladie par la qualité de vie. Nous pourrions développer un indicateur connu de survie ajustée sur la qualité de vie, QALY pour *Quality Adjusted Life Years*, qui nous donnerait un arbitrage entre quantité et qualité de vie. En effet, ce critère est souvent utilisé dans les études médico-économiques et permet de juger plus facilement entre différentes politiques de santé publique.

1. <http://www.who.int/mediacentre/factsheets/fs362/fr/>

Nous avons déjà discuté de la faisabilité d'un tel travail avec des professionnels du domaine. C'est un travail qui serait donc réalisé en collaboration avec une équipe d'économistes de la santé.

7.3.3 Généralisation et utilisation

Des perspectives quant à l'utilisation d'autres sources de données sont envisageables.

Une autre cohorte française, l'étude 3 Cités (3C), contient des données comparables à celles recueillies dans PAQUID [Alpérovitch 2002], avec un soin tout particulier apporté aux données médicales comme les facteurs vasculaires.

Au niveau européen, plusieurs cohortes sur la démence existent et sont utilisées, nous pouvons citer les études suédoises (Kungsholmen study) [Fratiglioni 2000], anglaises (MRC-cfas) et hollandaises (Rotterdam study) [Schrijvers 2012].

Il s'agit de cohortes populationnelles, avec un recueil initial d'un certain nombre de données (socio-démographiques, facteurs de risque, état de santé, ...), dans lesquelles une détection systématique de la démence est réalisée au cours du suivi.

En élargissant l'utilisation de notre modèle à des sources de données variées, mais se concentrant sur la même pathologie, nous pourrions, pourquoi pas, envisager de faire des projections à plus grande échelle, Européenne par exemple.

Dans cette thèse, nous avons déjà initié la généralisation de la méthode en relâchant des hypothèses et en rendant le modèle de projection plus flexible. Nous avons également entrepris d'adapter la méthode à d'autres maladies chroniques en vue de fournir des projections. Cependant, comme nous avons pu le constater, ceci n'est pas encore abouti et du travail reste à faire pour que la méthode s'adapte sans problèmes à d'autres pathologies.

Par ailleurs, il existe d'autres indicateurs de santé qu'il pourrait être utile de proposer pour juger de l'impact des différents scénarios de l'intervention. Par exemple, la fraction étiologique (ou risque attribuable), mesurant la proportion de cas de la maladie que nous pouvons attribuer au facteur de risque parmi les sujets exposés à ce facteur.

De plus, les modèles et la méthode sont implémentés en Fortran, ils nécessitent donc forcément un utilisateur qui connaît ce langage.

De ce fait, une fois la méthode généralisée, il serait fort utile d'uniformiser les programmes, de cibler les approches les plus utiles (en d'autres termes, d'identifier les modèles Markoviens/semi-Markoviens les plus pertinents) et de proposer un panel de modèles qui pourrait être implémentés sous la forme d'un package. Cela nécessitera également d'être assez flexible quant à la structure des données en entrée.

Bien évidemment, il est utopique de penser que nous pourrions proposer tous les modèles possibles. Mais la création d'un package sous le logiciel² R, par exemple,

2. <https://www.r-project.org/>

serait l'aboutissement d'un long travail sur cette approche et ces modèles et rendrait son utilisation plus simple et accessible à un large public.

Annexes

Sommaire

A.1 Contributions à la vraisemblance pour un modèle illness-death	134
A.2 Résolution des équations différentielles	136
A.2.1 Modèle Markovien M2 : méthode numérique, Runge-Kutta d'ordre 4	136
A.2.2 Modèle semi-Markovien S2	137
A.3 Résultats supplémentaires sur l'application à la démence	139
A.3.1 Mortalités estimées sur PAQUID	139
A.3.2 Impacts de l'intervention pour les hommes (modèle M2) . .	139
A.3.3 Comparaison des projections de prévalence : notre intervention <i>vs</i> l'intervention de Jacqmin <i>et al.</i> (modèle M2) . .	142
A.3.4 Autres résultats pour le modèle S2	144
A.4 Résultats supplémentaires sur l'application à la maladie de Parkinson	145
A.4.1 Estimation de l'incidence du Parkinson sur les données de la DCIR	145
A.4.2 Estimation du risque relatif, modèle M2	148

A.1 Contributions à la vraisemblance pour un modèle illness-death

Nous détaillons ici les différentes contributions à la vraisemblance intervenant dans un modèle illness-death, reprises et adaptées à nos notations de la thèse de C. Tourraine.

Nous nous plaçons dans un schéma d'observation des dates de décès en temps continu et des dates de maladie en temps discret et nous prenons en compte la censure par intervalle.

Nous distinguons quatre cas de trajectoires possibles des individus qui sont présentés dans la figure A.1.

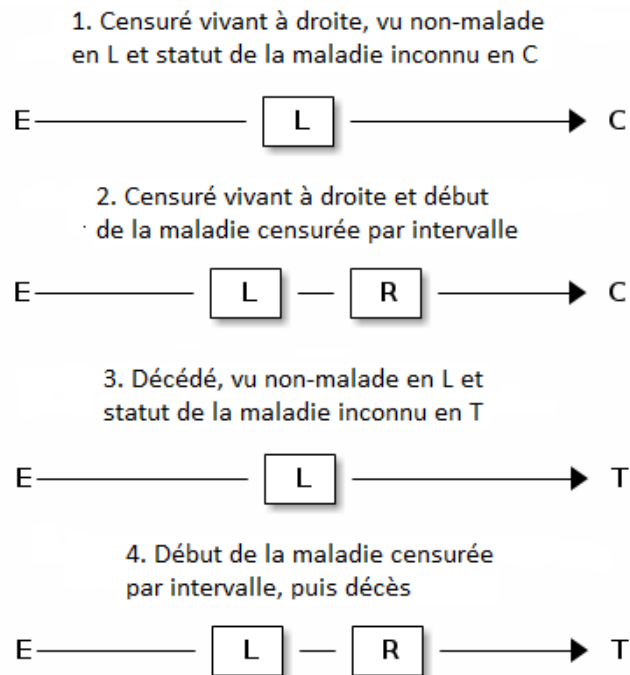


FIGURE A.1 – Trajectoires observées dans un modèle illness-death. Les lettres représentent différents temps, avec E l'entrée dans l'étude, C la fin des suivis, T le décès, L la dernière visite vu non-malade et R la première visite vu malade.

Pour un modèle Markovien non-homogène, nous avons les contributions suivantes :

1. Si le sujet i est toujours vivant à l'âge C de dernières nouvelles et n'était pas

A.1. Contributions à la vraisemblance pour un modèle illness-death 135

dément à l'âge L auquel il a été vu pour la dernière fois :

$$L_i = \frac{1}{e^{-A_{01}(0,a)-A_{02}(0,a)}} \left\{ e^{-A_{01}(0,C)-A_{02}(0,C)} + \int_L^C e^{-A_{01}(0,u)-A_{02}(0,u)} \alpha_{01}(u) \frac{e^{-A_{12}(0,C)}}{e^{-A_{12}(0,u)}} du \right\} \quad (\text{A.1})$$

Souvent, la date de dernières nouvelles correspond à la date de dernière visite. Dans ce cas, la contribution s'écrit :

$$L_i = \frac{e^{-A_{01}(0,L)-A_{02}(0,L)}}{e^{-A_{01}(0,a)-A_{02}(0,a)}}$$

2. Si le sujet i a été diagnostiqué malade à l'âge R et est toujours vivant à l'âge de dernières nouvelles C :

$$L_i = \frac{1}{e^{-A_{01}(0,a)-A_{02}(0,a)}} \left\{ \int_L^R e^{-A_{01}(0,u)-A_{02}(0,u)} \alpha_{01}(u) \frac{e^{-A_{12}(0,C)}}{e^{-A_{12}(0,u)}} du \right\} \quad (\text{A.2})$$

où L est son âge à la visite précédant la visite de diagnostic.

3. Si le sujet i est décédé à l'âge T et n'était pas dément à l'âge L auquel il a été vu pour la dernière fois :

$$L_i = \frac{1}{e^{-A_{01}(0,a)-A_{02}(0,a)}} \left\{ e^{-A_{01}(0,T)-A_{02}(0,T)} \alpha_{02}(T) + \int_L^T e^{-A_{01}(0,u)-A_{02}(0,u)} \alpha_{01}(u) \frac{e^{-A_{12}(0,T)}}{e^{-A_{12}(0,u)}} \alpha_{12}(T) du \right\} \quad (\text{A.3})$$

4. Si le sujet i a été diagnostiqué malade à l'âge R et est décédé à l'âge T :

$$L_i = \frac{1}{e^{-A_{01}(0,a)-A_{02}(0,a)}} \left\{ \int_L^R e^{-A_{01}(0,u)-A_{02}(0,u)} \alpha_{01}(u) \frac{e^{-A_{12}(0,T)}}{e^{-A_{12}(0,u)}} \alpha_{12}(T) du \right\} \quad (\text{A.4})$$

Pour un modèle semi-Markovien homogène. Soit \tilde{T} un âge, qui correspond à l'âge de dernières nouvelles C , si le sujet était vivant à la dernière visite, ou bien à l'âge de décès T et δ_2 est l'indicateur de décès (1 si décédé, 0 sinon).

Nous avons donc les contributions suivantes :

1. et 3. Si le sujet i n'était pas malade à l'âge L auquel il a été vu pour la dernière fois :

$$L_i = e^{-A_{01}(a,L)-A_{02}(a,L)} \left\{ e^{-A_{01}(L,\tilde{T})-A_{02}(L,\tilde{T})} \alpha_{02}(\tilde{T})^{\delta_2} + \int_L^{\tilde{T}} e^{-A_{01}(L,u)-A_{02}(L,u)} \alpha_{01}(u) e^{-A_{12}(0,\tilde{T}-u)} \alpha_{12}(\tilde{T}-u)^{\delta_2} du \right\} \quad (\text{A.5})$$

2. et 4. Si le sujet i a été diagnostiqué malade à l'âge R (et vu non-malade pour la dernière fois en L) :

$$L_i = e^{-A_{01}(a,L)-A_{02}(a,L)} \left\{ \int_L^R e^{-A_{01}(L,u)-A_{02}(L,u)} \alpha_{01}(u) e^{-A_{12}(0,\tilde{T}-u)} \alpha_{12}(\tilde{T}-u)^{\delta_2} du \right\} \quad (\text{A.6})$$

A.2 Résolution des équations différentielles

A.2.1 Modèle Markovien M2 : méthode numérique, Runge-Kutta d'ordre 4

La méthode de Runge-Kutta est une méthode d'analyse numérique qui permet d'approximer les solutions d'équations différentielles du premier ordre, en suivant une procédure itérative, selon un pas d'itération h .

Il existe plusieurs méthodes de Runge-Kutta, mais la plus utilisée pour obtenir une solution numérique reste la méthode d'ordre 4, car elle se distingue par sa "simplicité" de programmation, la stabilité de sa solution et la modification simple du pas.

Le principe de base de cette méthode est de calculer une valeur approchée de la solution de l'équation différentielle $y'(s) = f(s, y)$, en un certain nombre de points t_1, t_2, \dots, t_n en utilisant l'équation suivantes :

$$y_{n+1} = y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

Avec :

$k_1 = f(s_n; y_n)$, la pente au début de l'intervalle,

$k_2 = f(s_n + \frac{h}{2}; y_n + \frac{h}{2}k_1)$, la pente au milieu de l'intervalle en considérant k_1 ,

$k_3 = f(s_n + \frac{h}{2}; y_n + \frac{h}{2}k_2)$, la pente au milieu de l'intervalle en considérant k_2 ,

$k_4 = f(s_n + h; y_n + hk_3)$, la pente à la fin de l'intervalle en considérant k_3 .

Ainsi, la valeur qui suit, soit y_{n+1} , est approchée par la somme de la valeur actuelle, y_n , et du produit de la taille de l'intervalle, h , par la pente estimée. La pente est obtenue par une moyenne pondérée de pentes. Notons que dans notre cas, le pas d'itération h vaut 1 puisque nous devons calculer y pour chaque âge où le sujet est à risque de développer la maladie.

Exemple d'application : si nous souhaitons estimer la prévalence d'une maladie en t . Il s'agit de déterminer toutes les valeurs de y pour les sujets à risque de développer la maladie, c'est-à-dire ayant entre a_0 et 100 ans (que nous fixons pour l'exemple) en t , donc nés entre $t - 100$ et $t - a_0$.

Tout d'abord, il s'agit de se placer pour une année de naissance $t - a$ (avec $a_0 < a < 100$), puis de faire appel au programme Runge-Kutta autant de fois que les sujets nés cette année-là sont à risque de développer la maladie, soit $100 - a$ fois, afin de déterminer la valeur de y pour chaque âge entre a_0 et a .

Puis, nous réitérons l'opération pour chaque années de naissances concernées, en faisant varier a de a_0 à 100.

Nous obtiendrons donc autant de résultats que décrit dans la figure A.2 qui suit :

		Ages					
		a_0				100
Années de naissance	t-100	y	y	y	y	y	y
		y	y	y	y	y	
	⋮	y	y	y	y		
		y	y	y			
	⋮	y	y				
	t- a_0	y					

FIGURE A.2 – Récapitulatif des résultats donnés par le programme de Runge-Kutta.

Comme tout principe d'itération, il faut fixer une valeur initiale, soit $y(a_0) = A_{02}(a_0, a_0) = 0$.

Enfin, toutes les autres quantités de l'équation étant connues, par cette méthode, nous allons pouvoir calculer toutes les valeurs de α_{02} par âge et année de naissance.

A.2.2 Modèle semi-Markovien S2

L'équation différentielle à résoudre est la suivante :

$$y'(s) = c(s) - e^{-y(s)} a(s)$$

avec $y'(s) = \alpha_{02}(a)$, $y(s) = A_{02}(a_0, a)$, $c(s) = \alpha_2(a)$ et

$$a(s) = \frac{\int_{a_0}^a e^{-A_{01}(a_0, u)} \alpha_{01}(u) e^{-A^d(0, a-u)} \alpha^d(a-u) du}{P_v(a_0, a)}$$

Celle-ci a une solution analytique, il faut d'abord passer par un changement de variable.

Notons $x(s) = e^{y(s)}$, alors $x'(s) = y'(s)e^{y(s)}$ ce qui nous amène à :

$$\frac{x'(s)}{x(s)} = \frac{-a(s)}{x(s)} + c(s)$$

$$x'(s) = -a(s) + c(s)x(s)$$

Il s'agit d'une équation différentielle linéaire (1^{er} ordre) à coefficient non constant.

La résolution de l'équation homogène se fait en posant $a(s) = 0$, ce qui nous ramène à :

$$x'(s) - c(s)x(s) = 0$$

Ainsi, $x(s) = 0$ est une solution, ainsi que $\frac{x'(s)}{x(s)} = c(s)$, d'où $\ln(x(s)) = C(s) + k$.
Alors la solution est $x(s) = k_0 e^{C(s)}$ avec $C(s)$ la primitive de $c(s)$ et k_0 une constante.

Résolution de l'équation avec second membre $x'(s) - c(s)x(s) = -a(s)$.
En utilisant la méthode de la variation de la constante, nous cherchons une solution telle que $x(s) = k_0(s)e^{C(s)}$, donc :

$$x'(s) - c(s)x(s) = k_0'(s)e^{C(s)} + k_0(s)c(s)e^{C(s)} - c(s)x(s) = -a(s)$$

$$k_0'(s) = [-a(s) - k_0(s)c(s)e^{C(s)} + c(s)x(s)]e^{-C(s)} = -a(s)e^{-C(s)}$$

ce qui nous amène à : $k_0(s) = -\int^s a(u)e^{-C(u)} du + k_1$ avec k_1 une constante.

Maintenant, si nous remplaçons $k_0(s)$ dans $x(s) = k_0(s)e^{C(s)}$ nous obtenons :

$$x(s) = e^{C(s)}[-\int^s a(u)e^{-C(u)} du + k_1]$$

Ainsi, nous pouvons écrire $y(s)$ et $y'(s)$:

$$y(s) = C(s) + \ln(k_1 - \int^s a(u)e^{-C(u)} du)$$

$$y'(s) = c(s) - \frac{a(s)e^{-C(s)}}{k_1 \int^s a(u)e^{-C(u)} du}$$

Il reste à remplacer $c(s)$, $a(s)$ et $y(s)$ par leurs expressions pour déterminer A_{02} et α_{02}

$$A_{02}(a_0, a) = A_2(a_0, a) \tag{A.7}$$

$$+ \ln(k_1 - \int_{a_0}^a \frac{\int_{a_0}^u e^{-A_{01}(a_0, v)} \alpha_{01}(v) e^{-A^d(0, u-v)} \alpha^d(u-v) dv}{P_v(a_0, u)} e^{-A_2(a_0, u)} du)$$

$$\alpha_{02}(a) = \alpha_2(a)$$

$$- \frac{e^{-A_2(a_0, a)} \int_{a_0}^a e^{-A_{01}(a_0, u)} \alpha_{01}(u) e^{-A^d(0, a-u)} \alpha^d(a-u) du}{P_v(a_0, a)}$$

$$- \frac{\int_{a_0}^a k_1(s) - \int_{a_0}^u e^{-A_{01}(a_0, v)} \alpha_{01}(v) e^{-A^d(0, u-v)} \alpha^d(u-v) dv}{P_v(a_0, u)} e^{-A_2(a_0, u)} du$$

(Pour déterminer $k_1(s)$, nous supposons que $a = 65$ dans l'équation A.7, et nous trouvons $k_1(s) = 1$)

A.3 Résultats supplémentaires sur l'application à la démence

A.3.1 Mortalités estimées sur PAQUID

La figure A.3 représente la mortalité des déments, des non-déments et globale estimées sur la cohorte PAQUID en fonction de l'âge pour les femmes.

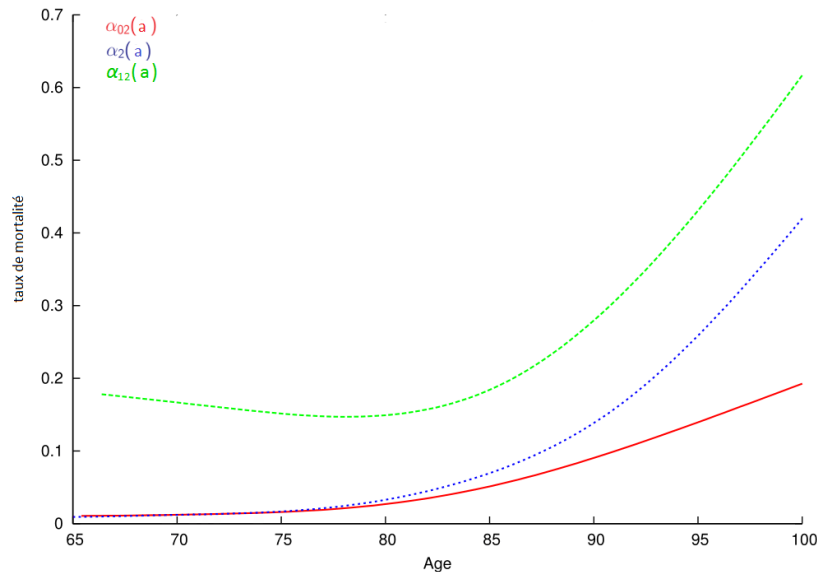


FIGURE A.3 – Estimations de la mortalité des déments, des non déments et globale en fonction de l'âge, pour les femmes de la cohorte PAQUID.

On observe que la mortalité des déments influence peu la mortalité globale estimée sur la cohorte pour les âges antérieurs à 80 ans car il y a peu de cas, de ce fait la mortalité des non-déments et la mortalité globale se superposent. Par contre, après 80 ans la mortalité globale de PAQUID se distingue significativement de la mortalité des non-déments du fait de l'augmentation du nombre de déments dans la cohorte.

Les mêmes observations sont faites concernant les hommes (pas représentés ici) avec une distinction entre la mortalité des non-déments et globale un peu plus tardive (vers 85 ans) mais tout de même significative.

Cette figure soutient l'hypothèse que nous devons différencier la mortalité des non-déments de la mortalité globale.

A.3.2 Impacts de l'intervention pour les hommes (modèle M2)

Les tableaux A.1 et A.2 concernent les projections d'indicateurs de santé pour les hommes, en fonction des différents scénarios d'intervention, avec les IC à 95%,

pour le modèle **M2**.

TABLE A.1 – Projections après l'intervention de la prévalence de la démence (en millier de sujets), de l'âge moyen de début de la démence AV_{01} et de la probabilité de développer la démence (F_{01}) à 75 ans, pour les hommes en France, (IC à 95%; modèle **M2**).

Scénario										Prev		AV_{01}	F_{01}
No.	$p_0(65)$	$p_0^*(65)$	ω	θ_{01}	θ_{01}^*	θ_{02}	θ_{02}^*	θ_{12}	θ_{12}^*	(N)	%#	(at age 75)	(at age 75)
Projections de référence en 2030										604	-	82,8	0,478
										[525;743]		[81,3;83,6]	[0,425;0,531]
Projections en 2030 après l'intervention en 2015													
1.a	0,4	0,2	0,5	1,5	1,5	2	2	2	2	733	+21,4	83,3	0,485
										[640;862]		[82,0;84,1]	[0,438;0,538]
1.b	0,4	0,4	1	1,5	1,27	2	1,74	2	1,74	605	+0,2	83,0	0,472
										[531;721]		[81,6;83,8]	[0,427;0,523]
2.a	0,4	0,2	0,5	2	2	1,5	1,5	1,5	1,5	664	+9,9	83,6	0,460
										[594;799]		[82,2;84,4]	[0,413;0,512]
2.b	0,4	0	0	2	2	1,5	1,5	1,5	1,5	725	+20,0	84,2	0,442
										[658;864]		[82,9;85,1]	[0,398;0,498]
3	0,2	0,2	1	3	1,5	1,5	1,5	1,5	1,5	539	-10,8	83,4	0,447
										[472;595]		[82,1;84,3]	[0,401;0,499]

% de variation comparé aux projections sans l'intervention.

En gras les valeurs de paramètres modifiées par l'intervention.

TABLE A.2 – Projections après l'intervention des espérances de vie LE_{00} , LE_0 , LE_{11} , $LE_{..}$ et du nombre moyen d'années passées en démence T_{11} (pour des hommes non-déments) à 75 ans, pour les hommes en France, (IC à 95%; modèle **M2**).

Scénario										LE_{00}	LE_0	LE_{11}	$LE_{..}$	T_{11}
No.	$p_0(65)$	$p_0^*(65)$	ω	θ_{01}	θ_{01}^*	θ_{02}	θ_{02}^*	θ_{12}	θ_{12}^*					
Projections in 2030 without intervention										11,48	13,90	8,07	13,55	2,41
										[11,15;11,77]	[13,82;14,02]	[6,96;9,07]	[13,51;13,59]	[2,05;2,87]
Projections in 2030 with intervention in 2015														
1.a	0,4	0,2	0,5	1,5	1,5	2	2	2	2	12,10	14,68	9,09	14,39	2,58
										[11,84;12,39]	[14,59;14,81]	[8,01;10,08]	[14,39;14,42]	[2,20;2,97]
1.b	0,4	0,4	1	1,5	1,27	2	1,74	2	1,74	11,77	14,19	8,65	13,90	2,42
										[11,52;12,05]	[14,10;14,33]	[7,57;9,65]	[13,89;13,93]	[2,05;2,81]
2.a	0,4	0,2	0,5	2	2	1,5	1,5	1,5	1,5	12,11	14,49	8,66	14,21	2,38
2.b	0,4	0	0	2	2	1,5	1,5	1,5	1,5	12,71	15,06	9,35	14,83	2,35
										[11,86;12,39]	[14,39;14,64]	[7,55;9,68]	[14,19;14,26]	[2,00;2,78]
3	0,2	0,2	1	3	1,5	1,5	1,5	1,5	1,5	11,80	14,03	8,39	13,78	2,23
										[11,55;12,07]	[13,93;14,17]	[7,28;9,40]	[13,75;13,83]	[1,86;2,62]

En gras les valeurs de paramètres modifiées par l'intervention.

Nous observons les mêmes tendances que celles décrites pour les femmes, sauf que tous les indicateurs de santé sont systématiquement moins élevés pour les hommes. Ainsi, les espérances de vie des hommes sont plus faibles que celles des femmes, mais les mêmes changements sont constatés : entre 2015 et 2030 ; entre les projections avec et sans intervention ; entre les différentes espérances de vie.

De plus, les intervalles de confiance sont peu larges exceptées pour LE_{11} , l'espérance de vie d'un malade, à cause du nombre de déments diagnostiqués à 75

ans.

La figure A.4 présente la probabilité de développer la démence estimée pour les hommes, par âge en 2015 et en 2030, sans intervention et avec l'intervention sous les scénarios 1.a et 1.b.

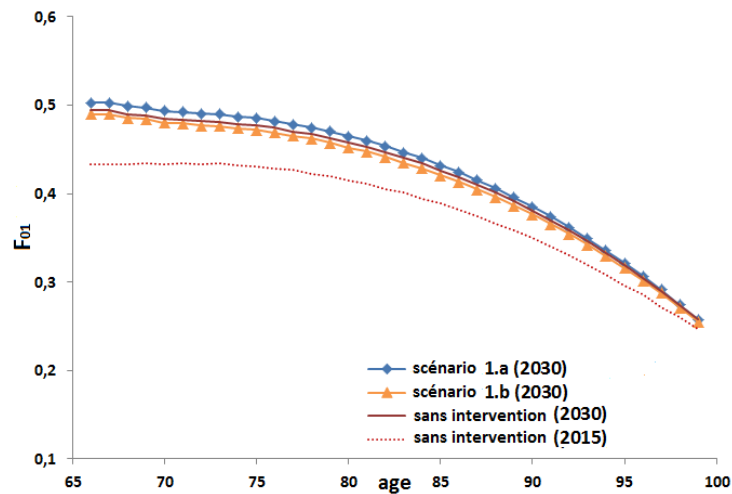


FIGURE A.4 – Evolution de la probabilité de développer la démence par âge, pour les hommes vivants et non-déments en 2015 et 2030, sans intervention ou avec les interventions 1.a et 1.b (modèle M2).

Comparée à la figure présentée pour les femmes 5.13, cette probabilité est toujours plus faible chez les hommes, mais suit la même évolution en fonction de l'âge.

La figure A.5 décrit les projections des espérances de vie LE_{00} , LE_{11} et LE_0 , ainsi que le nombre moyen d'années passées en démence T_{11} , pour les hommes, sans intervention en 2015 et 2030 et avec intervention sous les scénarios 1.a et 1.b en 2030.

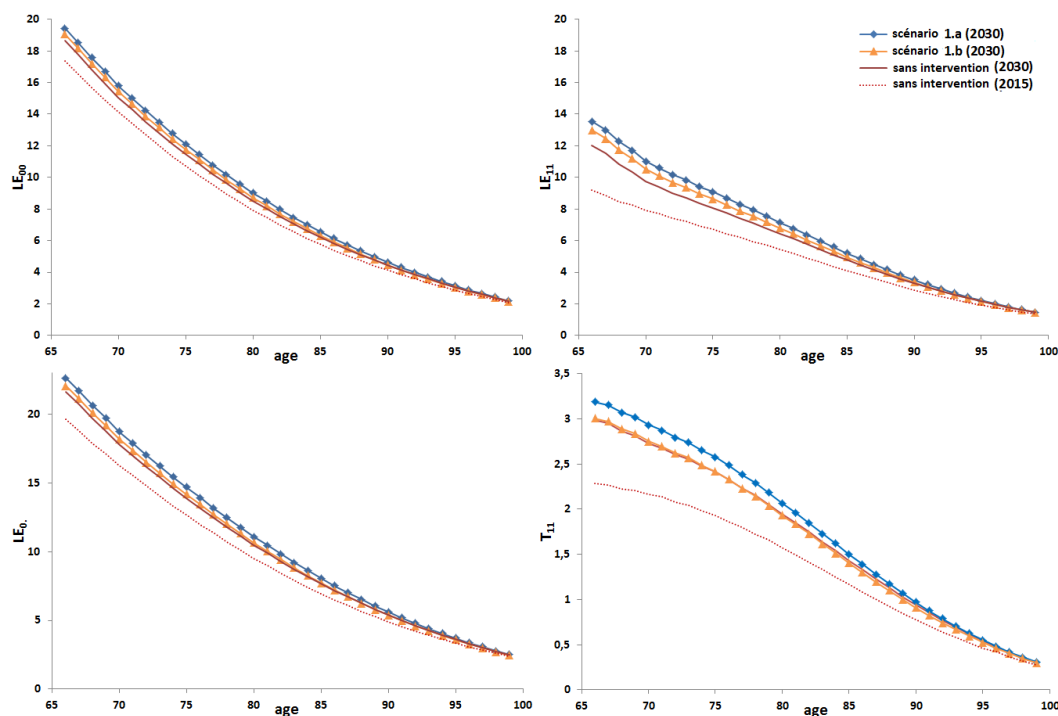


FIGURE A.5 – Projections des espérances de vie et du nombre moyen d’années passées en démence par âge, pour les hommes, sans intervention et avec intervention (scénarios 1.a et 1.b; modèle M2).

Nous pouvons faire les mêmes commentaires que ce que nous avons observé pour les femmes, figure 5.14.

A.3.3 Comparaison des projections de prévalence : notre intervention *vs* l’intervention de Jacqmin *et al.* (modèle M2)

Nos projections de prévalence de la démence sont beaucoup plus importantes pour 2030 que celles faites avec l’intervention de Jacqmin *et al.* [Jacqmin-Gadda 2013].

La figure A.6 décrit le pourcentage de variation par année entre 2030 et 2050, qu’il y a entre les projections de prévalence totale (hommes et femmes) avec notre intervention ou celle de Jacqmin *et al.*, par rapport aux projections de prévalence totale sans intervention.

Rappelons que ce scénario 1.a suppose un effet du facteur de risque plus fort sur le risque de décès que sur le risque de développer la démence.

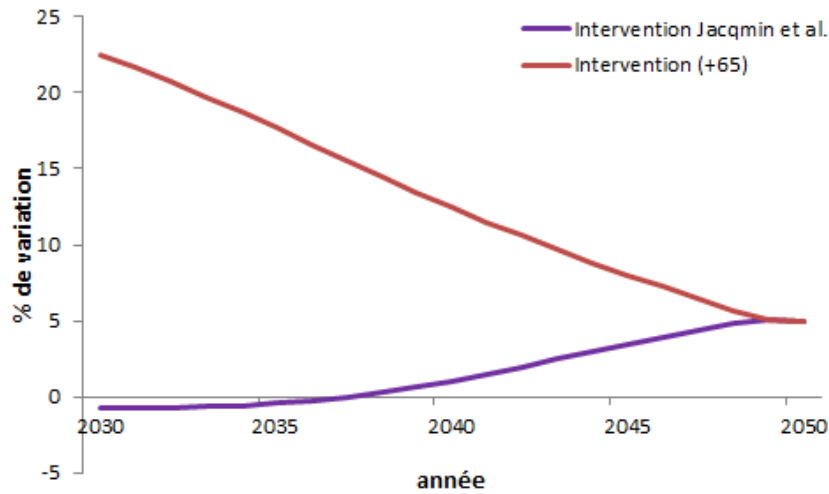


FIGURE A.6 – Pourcentage de variation par année, des projections de la prévalence de la démence en France, entre le contexte avec intervention en 2015 par rapport à sans intervention (modèle **M2**).

Ainsi, pour ce scénario, nous projetons environ 22% de cas de déments en plus en 2030, sous notre intervention par rapport à un contexte sans intervention, tandis qu'avec l'intervention de Jacqmin *et al.* il n'y aurait pas d'augmentation du nombre de cas par rapport aux projections sans intervention.

Mais lorsque nous faisons des projections plus lointaines, ce pourcentage de variation diminue avec notre intervention et augmente avec l'intervention de Jacqmin *et al.* jusqu'à se rejoindre en 2050 avec des projections supérieures, d'environ 6% de cas de déments en plus par rapport aux projections dans un contexte sans intervention.

Rappelons que la principale différence entre les deux interventions est que : la nôtre affecte tous les individus de plus de 65 ans dès 2015 alors que celle de Jacqmin *et al.* ne cible que les 65 ans en 2015.

Pour les sujets de moins de 65 ans en 2015, nous les traitons de la même façon, à savoir que les effets de l'intervention s'appliqueront sur eux dès lors qu'ils atteindront 65 ans.

Ainsi, l'intervention de Jacqmin *et al.* n'a un effet que sur les individus de 65 à 80 ans en 2030, tandis que notre intervention affecte les sujets entre 65 et 100 ans.

En revanche, en 2050 les deux interventions ont affecté la même population (c'est à dire les sujets de 65 à 100 ans).

La nature du scénario, à savoir une réduction de la prévalence de l'exposition au facteur de risque qui à un effet plus fort sur le décès que sur la démence, explique donc les pourcentages de variation observés, pour les projections de prévalence de la démence sous les deux différentes interventions, par rapport au contexte sans intervention.

Par exemple, il faut comprendre que l'intervention réduit principalement le risque de décéder (plus que le risque de développer la démence), or en 2030 cela concerne beaucoup moins d'individus pour l'intervention de Jacqmin *et al.* que pour la nôtre.

C'est pourquoi, sous notre intervention, comme le risque de décès diminue plus que le risque de développer la démence, nous avons une augmentation du nombre de déments attendus en 2030.

A.3.4 Autres résultats pour le modèle S2.

La figure A.7 décrit l'estimation de l'incidence de la démence *via* la cohorte PAQUID en fonction de l'âge avec les IC à 95% chez les hommes et chez les femmes.

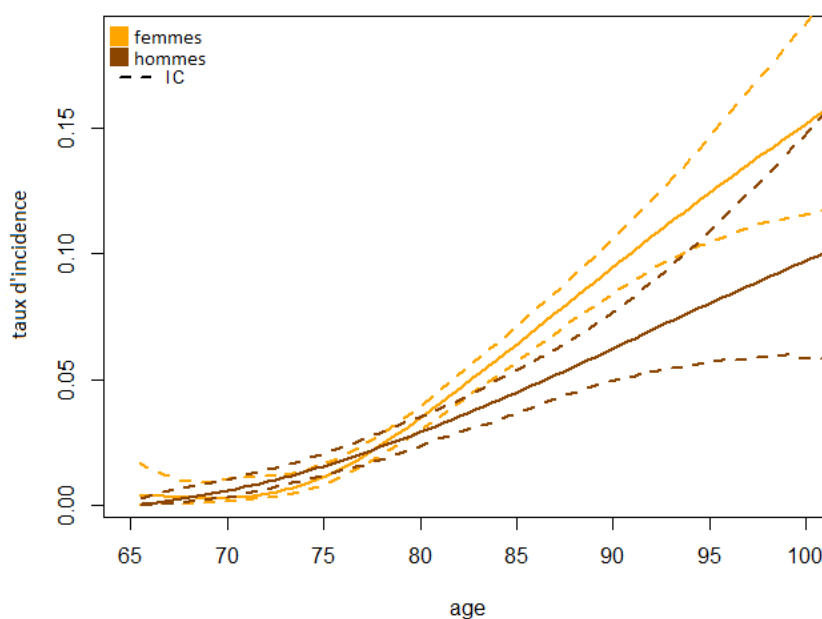


FIGURE A.7 – Estimations de l'incidence de la démence, α_{01} , par âge, pour les hommes et les femmes et IC à 95% (IC), cohorte PAQUID (modèle S2).

Nous observons une augmentation de l'incidence de la démence. Pour les âges les plus jeunes, les taux d'incidence pour les femmes et les hommes sont proches, mais à partir de 75 ans, ils augmentent rapidement et plus pour les femmes que pour les hommes.

Les figures A.8 et A.9 décrivent l'espérance de vie sans la démence en 2015 et 2030 en fonction de l'âge, pour les hommes et les femmes, avec les intervalles de confiance à 95% (modèle S2).

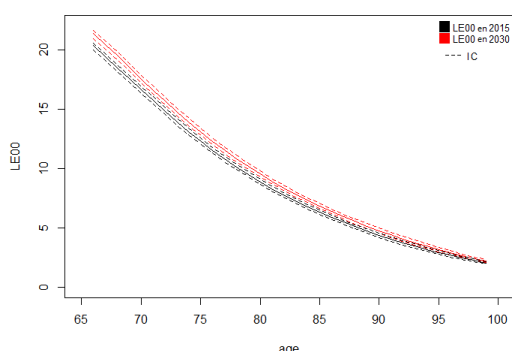


FIGURE A.8 – Espérance de vie sans démence (IC à 95%) pour les femmes.

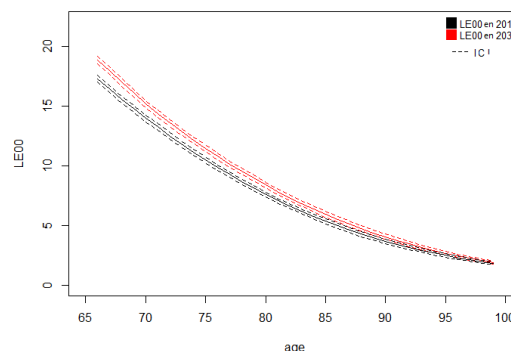


FIGURE A.9 – Espérance de vie sans démence (IC à 95%) pour les hommes.

Nous observons une augmentation de l'espérance de vis sans démence entre 2015 et 2030, surtout pour les âges les plus jeunes, et cette hausse est plus forte chez les hommes.

A.4 Résultats supplémentaires sur l'application à la maladie de Parkinson

A.4.1 Estimation de l'incidence du Parkinson sur les données de la DCIR

Ce travail a été effectué par F. Moisan, A. Elbaz *et al.* et est l'objet d'un article [Moisan 2011]. Les données du DCIR qui ont servi à estimer l'incidence du Parkinson sont les suivantes :

- délivrances de médicaments pouvant être utilisés pour traiter la maladie de Parkinson (médicaments antiparkinsoniens ; codes N04, G02CB01 et G02CB02 de la classification anatomique, thérapeutique et chimique (ATC) des médicaments) entre le 1^{er} janvier et le 31 décembre des années 2009, 2010 et 2011 avec pour chaque délivrance :
 - spécialité médicale des prescripteurs,
 - code ATC détaillé des médicaments,
 - dosage des comprimés et nombre de comprimés par boîte,
 - nombre de boîtes délivrées,
 - date de la délivrance.
- pour les personnes ayant consommé au moins une fois un médicament antiparkinsonien sur la période, l'ensemble des prescriptions de médicaments neuroleptiques (code ATC N05A) sur la même période ;
- caractéristiques démographiques en 2009, 2010, 2011 (âge ; sexe ; département de résidence) ;

- si décès sur cette période, le mois du décès.

Ensuite les prédicteurs nécessaires à l'application d'un modèle prédictif de type logistique ont été générés année par année. Le logit de la probabilité de recevoir un traitement antiparkinsonien pour la maladie de Parkinson s'écrit en fonction du sexe, du nombre de prescriptions dans l'année, du temps du traitement, des doses annuelles des divers médicaments antiparkinsonien...

Les personnes atteintes de la maladie de Parkinson sont identifiées à partir de leur profil de consommation de médicaments. Sachant que beaucoup des médicaments listés peuvent être prescrits pour d'autres maladies, un modèle statistique qui estime la probabilité qu'une personne a effectivement reçu un traitement antiparkinsonien pour une maladie de Parkinson, en fonction du type et/ou de la dose totale annuelle des différents médicaments délivrés, de la spécialité du prescripteur et de la régularité avec laquelle cette personne a été traitée sur l'année, a été utilisé [Moisan 2011].

Afin d'estimer l'incidence de la maladie de Parkinson en 2010 les données de trois années consécutives ont été utilisées : 2009 et 2010 pour identifier les cas et 2011 pour réaliser une correction. Puis le modèle a été appliqué aux consommateurs de médicaments antiparkinsoniens, rentrant dans les critères, de chacune des trois années. Ensuite, pour chacune des trois années, la probabilité que les consommateurs de médicaments antiparkinsoniens aient été traités pour une maladie de Parkinson a été estimée (sont considérés comme atteints ceux dont la probabilité était supérieure ou égale à 0,255).

L'incidence de la maladie de Parkinson en 2010 est donc estimée de la façon suivante : les cas incidents de maladie de Parkinson en 2010 sont définis comme les cas prédits Parkinsoniens par le modèle en 2010 et dont la première prescription de médicaments antiparkinsoniens a eu lieu en 2010 (aucune prescription en 2009).

Puis le nombre de cas incidents est divisé par une estimation fournie par l'INSEE des personnes-années cumulées pour la population française en 2010 par âge (défini comme l'âge atteint dans l'année).

Ce calcul a été fait par sexe, classe d'âge et département de résidence.

Par la suite, l'incidence estimée a été corrigée suivant deux approches différentes :

- Méthode 1 : les estimations de la sensibilité et de la spécificité du modèle prédictif permettent d'obtenir un nombre corrigé de patients parkinsoniens incidents en 2010 pour tenir compte des erreurs de classement (faux positifs et négatifs) sous l'hypothèse que les performances de modèle sont similaires quels que soient le sexe et la classe d'âge.
- Méthode 2 : en utilisant les données de 2011 pour corriger le nombre de

cas incidents. Il s'agit de considérer que les consommateurs de médicaments antiparkinsoniens en 2010 avaient une maladie de Parkinson incidente en 2010 quand : le traitement a été débuté en 2010 ; la consommation des médicaments antiparkinsoniens se poursuit en 2011 ; le modèle prédit ces sujets comme étant Parkinsoniens en 2011.

Pour finir, il s'avère que la méthode d'identification des patients parkinsoniens conduit à une sous-estimation du nombre de personnes atteintes chez les individus les plus âgées, en raisons de malades non diagnostiqués et non traités à cause d'un mauvais accès aux soins ou parce que le diagnostic est plus difficile chez ces personnes (symptômes liés au vieillissement normal).

Mais la sous-estimation vient également du fait que les consommateurs de médicaments, institutionnalisés dans des établissements avec des pharmacies à usage interne, ne sont pas identifiés par le DCIR.

En conséquence, les estimations sont plus fiables pour les personnes âgées de moins de 80 ans. Au-delà, la diminution de l'incidence est vraisemblablement expliquée en grande partie par une moins bonne identification des patients. Une analyse de sensibilité a été réalisée.

Note : La même démarche a été suivie pour estimer l'incidence en 2011 et 2012.

A.4.2 Estimation du risque relatif, modèle M2

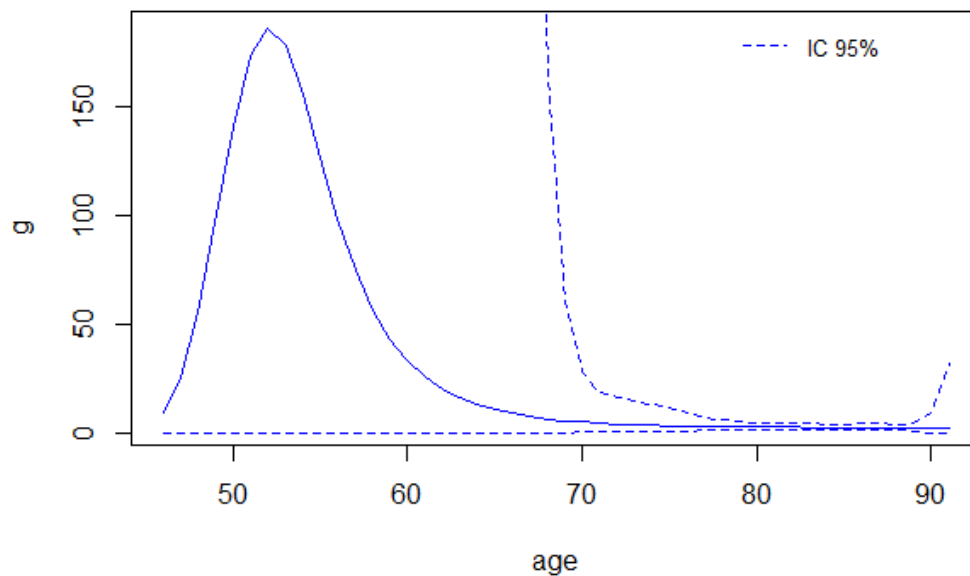


FIGURE A.10 – Estimation du risque relatif de décès des Parkinsoniens par rapport aux non-Parkinsoniens par âge pour les femmes et IC à 95%.

Cette figure A.10 fournit les estimations, issues de l'algorithme de recherche, du risque relatif de décès des Parkinsoniens par rapport aux non-Parkinsoniens par âge pour les femmes. Comme nous l'avons constaté, l'estimation de ce risque pour les âges jeunes est largement trop importante. Cependant, l'analyse de sensibilité que nous avons effectuée (en contraignant le risque relatif pour les âges les plus jeunes) montre que l'impact sur les projections de prévalence est moindre.

Bibliographie

- [Aalen 1978] O.O Aalen et S. Johansen. *An empirical transition matrix for non-homogeneous Markov chains based on censored observations*. Scandinavian Journal of Statistics, vol. 5, pages 141–150, 1978. (Cit  en page 30.)
- [Alho 2000] J.M. Alho. *Discussion of "The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications"*. North American Actuarial Journal., vol. 4, no. 1, pages 91–93, 2000. (Cit  en page 15.)
- [Alioum 2001] A. Alioum et D. Commenges. *MKVPCI : a computer program for Markov models with piecewise constant intensities and covariates*. Computer Methods and Programs in Biomedicine, vol. 64, pages 109–119, 2001. (Cit  en page 41.)
- [Alp rovitch 2002] A. Alp rovitch, P. Amouyel, J.F. Dartigues, P. Ducimet re, B. Mazoyer, K. Ritchie et C. Tzourio. *Les  tudes  pid miologiques sur le vieillissement en France : de l  tude Paquid   l  tude des Trois Cit s*. Comptes Rendus Biologies, vol. 325, pages 665–672, 2002. (Cit  en page 131.)
- [Andersen 1993] P.K. Andersen, O. Borgan, R.D. Gill et N. Keiding. *Statistical models based on counting processes*. Springer Series in Statistics, 1993. (Cit  en pages 18 et 19.)
- [Barnes 2011] D. E. Barnes et K. Yaffe. *The projected effect of risk factor reduction on Alzheimer’s disease prevalence*. The Lancet Neurology, vol. 10, no. 9, pages 819–828, 2011. (Cit  en pages 7, 52 et 126.)
- [Berr 2007] C. Berr, T. N. Akbaraly, F. Nourashemi et S. Andrieu. *Epid miologie des d mences*. La Presse M dicale, vol. 36, no. 10, pages 1430–1441, 2007. (Cit  en pages 6, 66 et 67.)
- [Bertin 2005] N. Bertin, M. Chantelou, N. Vallier, A. Weill, P. Fender et H. Allemant. *The prevalence of Parkinson’s disease and its cost to the French national health fund in metropolitan France in 2000*. Rev Med Ass Maladie, vol. 36, pages 113–122, 2005. (Cit  en page 105.)
- [Blanpain 2010] N. Blanpain et O. Chardon. *Projections de population 2007-2060 pour la france m ropolitaine : m thode et principaux r sultats*. Document de travail, INSEE, 2010. (Cit  en page 70.)
- [Bower 1999] J.H. Bower, D.M. Maraganore et S.K. McDonnell ans W.A. Rocca. *Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976-1990*. Neurology, vol. 52, pages 1214–1220, 1999. (Cit  en page 104.)
- [Box 1976] G.E.P. Box et G.M. Jenkins. *Time series analysis : Forecasting and control*. Holden-Day, San Francisco, 1976. (Cit  en page 14.)
- [Brookmeyer 1998] R. Brookmeyer, S. Gray et C. Kawas. *Projections of Alzheimer’s disease in the United States and the public health impact of delaying disease*

- onset*. American Journal of Public Health, vol. 88, no. 9, pages 1337–1342, 1998. (Cit  en page 6.)
- [Brookmeyer 2007] R. Brookmeyer, E. Johnson, K. Ziegler-Graham et H. M. Arighi. *Forecasting the global burden of Alzheimer’s disease*. Alzheimer’s & Dementia, vol. 3, no. 3, pages 186–191, 2007. (Cit  en pages 7, 52, 98 et 126.)
- [Brouhns 2002a] N. Brouhns et M. Denuit. *Risque de long vit  et rentes viag res II. Tables de mortalit  prospectives pour la population belge*. Belgian Actuarial Bulletin. Ghent University, Ghent., vol. 2, no. 1, pages 49–63, 2002. (Cit  en page 12.)
- [Brouhns 2002b] N. Brouhns, M. Denuit et J.K. Vermunt. *A poisson log-bilinear regression approach to the construction of projected lifetables*. Insurance : Mathematics and Economics., vol. 31, pages 373–393, 2002. (Cit  en page 15.)
- [Burgio 2015] A. Burgio et L. Frova. *Projections de mortalit  par cause de d c s : Extrapolation tendancielle ou mod le  ge-p riode-cohorte*. Population, vol. 50, pages 1031–1051, 2015. (Cit  en page 17.)
- [Clayton 1978] D.G. Clayton. *A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence*. Biometrika, vol. 65, pages 145–151, 1978. (Cit  en page 18.)
- [Clayton 1987] D. Clayton et E. Schifflers. *Models for temporal variation in cancer rates. I : Age-period and age-cohort models*. Statistics in Medicine, vol. 6, pages 449–481, 1987. (Cit  en page 15.)
- [Commenges 1998] D. Commenges, L. Letenneur, P. Joly, A. Alioum et J.F. Dartigues. *Modelling age-specific risk : application to dementia*. Statistic in medicine, vol. 17, pages 1973–1988, 1998. (Cit  en page 18.)
- [Commenges 1999] D. Commenges. *Multi-state models in Epidemiology*. Lifetime Data Analysis, vol. 5, pages 315–327, 1999. (Cit  en page 18.)
- [Commenges 2002] D. Commenges. *Inference for multi-state models from interval-censored data*. Statistical methods in medical research, vol. 44, pages 162–182, 2002. (Cit  en page 42.)
- [Commenges 2007] D. Commenges, P. Joly, A. G gout-Petit et B. Liqueur. *Choice between semi-parametric estimators of Markov and non-Markov multi-state models from generally coarsened observations*. Scandinavian Journal of Statistics, vol. 34, no. 1, pages 33–52, 2007. (Cit  en pages 75, 97 et 100.)
- [Commenges 2015] D. Commenges et H. Jacqmin-Gadda. *Dynamical biostatistical models.*, volume 86. CRC Press, 2015. (Cit  en page 53.)
- [Cox 1972] D.R. Cox. *Regression models and life-tables*. Journal of the Royal Statistical Society. Series B (Methodological), vol. 34, no. 2, pages 187–220, 1972. (Cit  en page 18.)
- [Dartigues 1991] J. F. Dartigues, M. Gagnon, P. Michel, L. Letenneur, D. Commenges, P. Barberger-Gateau, S. Auriacombe, B. Rigal, R. Bedry, A. Alperovitch, J. M. Orhohozo, P. Henry, P. Loiseau et R. Salamon. *Le programme*

- de recherche PAQUID sur l'épidémiologie de la démence. Méthodologie et résultats initiaux.* Revue Neurologique, vol. 147, no. 3, pages 225–230, 1991. (Cité en pages 67 et 69.)
- [Decarli 1987] A. Decarli et C. la Vecchia. *Age, period and cohort models : review of knowledge and implementation in GLIM.* Rivista di Statistica, vol. 20, no. 4, pages 398–410, 1987. (Cité en page 15.)
- [Duron 2008] E. Duron et O. Hanon. *Hypertension, cognitive decline and dementia.* Archives of Cardiovascular Diseases, vol. 101, no. 3, pages 181–189, 2008. (Cité en pages 67 et 76.)
- [Durrleman 1989] S. Durrleman et R. Simon. *Flexible regression models with cubic splines.* Statistic in medicine, vol. 8, no. 5, pages 551–561, 1989. (Cité en page 16.)
- [Elbaz 2003] A. Elbaz, J.H. Bower, B.J. Peterson, D.M. Maraganore, S.K. McDonnell, J.E. Ahlskog, D.J. Schaid et W.A. Rocca. *Survival study of Parkinson disease in Olmsted county, Minnesota.* Archneuro, vol. 60, pages 91–96, 2003. (Cité en page 106.)
- [Farrer 1997] L. A. Farrer, L. A. Cupples, J. L. Haines, B. Hyman, W. A. Kukull, R. Mayeux, R. H. Myers, M. A. Pericak-Vance, N. Risch et C. M. van Duijn. *Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium.* Journal of the American Medical Association, vol. 278, no. 16, pages 1349–1356, 1997. (Cité en page 77.)
- [Fratiglioni 2000] L. Fratiglioni, L. J. Launer, K. Andersen, M. M. Breteler, J. R. Copeland, J. F. Dartigues, A. Lobo, J. Martinez-Lage, H. Soininen et A. Hofman. *Incidence of dementia and major subtypes in Europe : A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group.* Neurology, vol. 54, no. 11 suppl 5, pages S10–S15, 2000. (Cité en pages 99 et 131.)
- [Frydman 1999] H. Frydman. *Nonparametric estimation of a Markov illness-death process from interval-censored observations, with application to diabetes survival data.* Biometrika, vol. 82, no. 4, pages 773–789, 1999. (Cité en page 30.)
- [Frydman 2009] H. Frydman et M. Szarek. *Nonparametric estimation in a Markov illness-death process from interval censored observations with missing intermediate transition status.* Biometrics, vol. 65, no. 1, pages 143–151, 2009. (Cité en page 30.)
- [Frydman 2010] H. Frydman et M. Szarek. *Estimation of overall survival in an illness-death model with application to the vertical transmission of HIV-1.* Statistics in medicine, vol. 29, no. 19, pages 2045–2054, 2010. (Cité en page 30.)
- [Genin 2011] E. Genin, D. Hannequin, D. Wallon, K. Sleegers, M. Hiltunen, O. Combarros, M. J. Bullido, S. Engelborghs, P. de Deyn, C. Berr, F. Pasquier,

- B. Dubois, G. Tognoni et *others*. *Apoe and alzheimer disease : a major gene-with semidominant inheritance*. *Molecular Psychiatry*, vol. 16, no. 9, pages 903–907, 2011. (Cité en page 77.)
- [Gompertz 1825] B. Gompertz. *On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies*. *Philosophical Transactions of the Royal Society*, vol. 115, pages 513–583, 1825. (Cité en pages 42 et 71.)
- [Grasset 2015] L. Grasset, C. Brayne, P. Joly, H. Jacqmin-Gadda, K. Perez, A. Foubert-Samier, J. F. Dartigues et C. Helmer. *Trends in dementia incidence : Evolution over a 10-year period in France*. *Alzheimer's & Dementia*, vol. in press, 2015. (Cité en page 99.)
- [Greenland 1995] S. Greenland et W.D. Finkle. *A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses*. *American journal of epidemiology*, vol. 142, no. 12, pages 1255–1264, 1995. (Cité en page 17.)
- [Gutterman 1999] S. Gutterman et I.T. vanderhoof. *Forecasting changes in mortality : a search for a law of causes and effects*. *North American Actuarial Journal.*, vol. 3, no. 2, pages 135–138, 1999. (Cité en page 14.)
- [Hebert 2003] L E. Hebert, P A. Scherr, J L. Bienias, D A. Bennett et D A. Evans. *Alzheimer disease in the United States (2010-2050) estimated using the 2010 census*. *Neurology*, vol. 60, no. 8, pages 1119–1122, 2003. (Cité en page 6.)
- [Heutte 2001] N. Heutte, C. Huber et O. Pons. *Semi-Markovian models applied to aids with censoring*. *Statistics in medicine*, vol. 21, pages 103–113, 2001. (Cité en page 22.)
- [Hougaard 1999] P. Hougaard. *Multi-state models : a review*. *Lifetime Data Analysis*, vol. 5, pages 239–264, 1999. (Cité en pages 18 et 19.)
- [Jacqmin-Gadda 2013] H. Jacqmin-Gadda, A. Alperovitch, C. Montlathuc, D. Commenges, K. Leffondre, C. Dufouil, A. Elbaz, C. Tzourio, J. Ménard, J. F. Dartigues et P. Joly. *20-years prevalence projections for dementia and impact of preventive policy about risk factors*. *European Journal of Epidemiology*, vol. 28, no. 6, pages 493–502, 2013. (Cité en pages 7, 52, 67, 77, 88, 126 et 142.)
- [Janssen 1997] J. Janssen, R. Manca et E. Volpe. *Markov and semi-Markov options pricing models with arbitrage possibility*. *Applied stochastic models and data analysis*, vol. 13, no. 2, pages 103–113, 1997. (Cité en page 22.)
- [Janssen 1999] J. Janssen et N. Limnios. *Semi-markov models and applications*. Kluwer Academic Publishers, 1999. (Cité en page 27.)
- [Joly 1999] P. Joly et D. Commenges. *A penalized likelihood approach for a progressive three-state model with censored and truncated data : Application to AIDS*. *Biometrics.*, vol. 55, pages 887–890, 1999. (Cité en page 30.)

- [Joly 2002] P. Joly, D. Commenges, C. Helmer et L. Letenneur. *A penalized likelihood approach for an illness-death model with interval-censored data : application to age-specific incidence of dementia*. *Biostatistics*, vol. 3, no. 3, pages 433–443, 2002. (Cité en pages 7, 30, 42 et 57.)
- [Joly 2013] P. Joly, C. Touraine, A. Georget, J. F. Dartigues, D. Commenges et H. Jacqmin-Gadda. *Prevalence projections of chronic diseases and impact of public health intervention*. *Biometrics*, vol. 69, pages 109–117, 2013. (Cité en pages 6, 23, 52, 58, 67 et 125.)
- [Keiding 1991] N. Keiding. *Age-specific incidence and prevalence : A statistical perspective*. *Journal of the Royal Statistical Society.A*, vol. 154, pages 371–396, 1991. (Cité en page 18.)
- [Kupper 1985] L.L. Kupper, J.M. Janis, A. Karmous et B.G. Greenberg. *Statistical age-period-cohort analysis : A review and critique*. *Journal of chronic disease*, vol. 38, pages 811–830, 1985. (Cité en page 15.)
- [Lee 1992] R.D. Lee et L.R. Carter. *Modeling and forecasting U.S. mortality*. *Journal of the American Statistical Association*, vol. 87, no. 419, pages 659–671, 1992. (Cité en page 12.)
- [Lee 1994] R.D. Lee et R. Rofman. *Modelling and forecasting mortality in Chile*. *Notas Poblacion*, vol. 59, no. 22, pages 182–213, 1994. (Cité en page 12.)
- [Majer 2013] I.M. Majer, R. Stevens, W.J. Nusselder, J.P. Mackenbach et P.H.M. van Baal. *Modeling and forecasting health expectancy ; Theoretical framework and application*. *Demography*, vol. 50, no. 2, pages 673–697, 2013. (Cité en page 14.)
- [Moisan 2011] F. Moisan, V. Gourlet, J.L. Mazurie, J.L. Dupupet, J. Houssinot, M. Goldberg, E. Imbernon, C. Tzourio et A. Elbaz. *Prediction model of Parkinson's disease based on antiparkinsonian drug claims*. *American Journal of Epidemiology*, vol. 174, pages 354–363, 2011. (Cité en pages 145 et 146.)
- [Negri 1990] E. Negri, C. la Vecchia, A. Decarli et P. Boyle. *Projections to the end of the century of mortality from major cancer sites in Italy*. *Tumori*, no. 76, pages 420–428, 1990. (Cité en page 18.)
- [Osmond 1985] C. Osmond. *Using Age, Period and Cohort Models to Estimate Future Mortality Rates*. *International journal of epidemiology*, vol. 14, no. 1, pages 124–129, 1985. (Cité en page 15.)
- [Ouhbi 1999] B. Ouhbi et N. Limnios. *Nonparametric Estimation for Semi-Markov Processes Based on its Hazard Rate Functions*. *Statistical Inference for Stochastic Processes*, vol. 2, no. 2, pages 151–173, 1999. (Cité en page 30.)
- [Pearl 2009] J. Pearl. *Causality*. Cambridge university Press, 2009. (Cité en page 55.)
- [Pérez-Ocón 2002] R. Pérez-Ocón et I. Torres-Castro. *A reliability semi-Markov model involving geometric processes*. *Applied stochastic models in business and industry*, vol. 18, no. 2, pages 157–170, 2002. (Cité en page 22.)

- [Peters 2008] R. Peters, N. Beckett, F. Forette, J. Tuomilehto, R. Clarke, C. Ritchie, A. Waldman, I. Walton, R. Poulter, S. Ma, M. Comsa, L. Burch, A. Fletcher et C. Bulpitt. *Incident dementia and blood pressure lowering in the hypertension in the very elderly trial cognitive function assessment (hyvet-cog) : a double-blind, placebo controlled trial*. *The Lancet neurology*, vol. 7, no. 8, pages 683–689, 2008. (Cité en page 77.)
- [Polidori 2012] M. C. Polidori, L. Pientka et P. Mecocci. *A Review of the Major Vascular Risk Factors Related to Alzheimer's Disease*. *Journal of Alzheimer's Disease*, vol. 32, pages 521–530, 2012. (Cité en pages 7 et 67.)
- [Power 2011] M. C. Power, J. Weuve, J. J. Gagne, M. B. McQueen, A. Viswanathan et D. Blacker. *The association between blood pressure and incident Alzheimer disease : a systematic review and meta-analysis*. *Epidemiology*, vol. 22, no. 5, pages 646–659, 2011. (Cité en page 76.)
- [Putter 2007] H. Putter, M. Fiocco et R.B. Geskus. *Tutorial in biostatistics : Competing risks and multi-state models*. *Statistics in medicine*, vol. 26, pages 2389–2430, 2007. (Cité en page 31.)
- [Ramaroson 2003] H. Ramaroson, C. Helmer, P. Barberger-Gateau, L. Letenneur et J. F. Dartigues. *Prevalence of dementia and Alzheimer's disease among subjects aged 75 years or over : updated results of the PAQUID cohort*. *Revue Neurologique*, vol. 159, no. 4, pages 405–411, 2003. (Cité en page 67.)
- [Renshaw 2003] A. Renshaw et S. Haberman. *Lee-Carter mortality forecasting : a parallel generalized linear modelling approach for England and Wales mortality projections*. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, vol. 52, no. 1, pages 119–137, 2003. (Cité en page 15.)
- [Renshaw 2008] A. Renshaw et S. Haberman. *Extensions to the lee-carter model, including risk measurement in the age-period-cohort versions of the model*. Cass Business School, City University, London, UK, 2008. (Cité en page 15.)
- [Rocca 2011] W. A. Rocca, R. C. Petersen, D. S. Knopman, L. E. Hebert, D. A. Evans, K. S. Hall, S. Gao, F. W. Unverzagt, K. M. Langa, E. B. Larson et L. R. White. *Trends in the incidence and prevalence of Alzheimer's disease, dementia, and cognitive impairment in the United States*. *Alzheimers Dement*, vol. 7, no. 1, pages 80–93, 2011. (Cité en pages 68, 74, 82 et 99.)
- [Rosenberg 1995] P.S. Rosenberg. *Hazard function estimation using B-splines*. *Biometrics*, vol. 51, no. 3, pages 874–887, 1995. (Cité en page 30.)
- [Royston 2000] P. Royston. *A strategy for modelling the effect of a continuous covariate in medicine and epidemiology*. *Statistics in medicine*, vol. 19, no. 14, pages 1831–1847, 2000. (Cité en page 16.)
- [Saint-Pierre 2005] P. Saint-Pierre. *Modèles multi-états de type Markovien et application à l'asthme*. PhD thesis, Université Montpellier I, 2005. (Cité en page 30.)
- [Schrijvers 2012] E.M.C. Schrijvers, B.F.J. Verhaaren, P.J.Koudstaal, A. Hofman, M.A. Ikram et M.M.B Breterler. *Is dementia incidence declining ? Trends in*

- dementia incidence since 1990 in the Rotterdam Study*. *Neurology*, vol. 78, no. 19, pages 1456–1463, 2012. (Cit  en pages 68, 74, 82 et 131.)
- [Staessen 2000] J. A. Staessen, J. Gasowski, J. G. Wang, L. Thijs, E. Den-Hond, J. P. Boissel, J. Coope, T. Ekbohm, F. Gueyffier, L. Liu, K. Kerlikowski, S. Pocock et R. H. Fagard. *Risks of untreated and treated isolated systolic hypertension in the elderly : meta-analysis of outcome trials*. *Lancet*, vol. 355, no. 9207, pages 865–872, 2000. (Cit  en page 78.)
- [Tabeau 1999] E. Tabeau, P. Ekamper, C. Huisman et A. Bosch. *Improving Overall Mortality Forecasts by Analysing Cause-of-Death, Period and Cohort Effects in Trends*. *European Journal of Population*, vol. 15, no. 2, pages 153–183, 1999. (Cit  en page 18.)
- [Therneau 2000] T. Therneau et P.M. Grambsch. *Modeling Survival Data : Extending the Cox Model*. *Statistics for Biology and Health*, 2000. (Cit  en pages 19 et 27.)
- [Tilvis 2004] R. S. Tilvis, M. H. K h nen-V re, J. Jolkkonen, J. Valvanne, K. H. Pitkala et T. E. Strandberg. *Predictors of cognitive decline and mortality of aged people over a 10-year period*. *Dementia and Geriatric cognitive disorders*, vol. 59, no. 3, pages M268–M274, 2004. (Cit  en page 98.)
- [Titman 2015] A.C. Titman. *Transition probability estimates for non-Markov multi-state models*. *Biometrics*, vol. 71, no. 4, pages 1034–1041, 2015. (Cit  en page 30.)
- [Tolppanen 2012] A. M. Tolppanen, A. Solomon, H. Soininen et M. Kivipelto. *Mid-life Vascular Risk Factors and Alzheimer’s Disease : Evidence from Epidemiological Studies*. *Journal of Alzheimer’s Disease*, vol. 32, pages 531–540, 2012. (Cit  en pages 67, 76 et 77.)
- [Touraine 2013a] C. Touraine. *Mod les illness-death pour donn es censur es par intervalle : Application   l’ tude de la d mence*. PhD thesis, Universit  de Bordeaux II, 2013. (Cit  en page 33.)
- [Touraine 2013b] C. Touraine, C. Helmer et P. Joly. *Predictions in an illness-death model*. *Statistical Methods in Medical Research*, vol. 10, 2013. (Cit  en page 36.)
- [van den Hout 2008] A. van den Hout et F. E. Matthews. *Multi-state analysis of cognitive ability data : a piecewise-constant model and a Weibull model*. *Statistics in Medicine*, vol. 27, no. 26, pages 5440–5455, 2008. (Cit  en page 37.)
- [van den Hout 2010] A. van den Hout et F. E. Matthews. *Estimating stroke-free and total life expectancy in the presence of non-ignorable missing values*. *Journal of the Royal Statistical Society Series A Statistics in Society*, vol. 173, no. 2, pages 331–349, 2010. (Cit  en page 37.)
- [Verdecchia 1989] A. Verdecchia, R. Capocaccia, V. Egidi et A. Golini. *A method for the estimation of chronic disease morbidity and trends from mortality data*. *Statistic in medicine*, vol. 8, no. 2, pages 201–216, 1989. (Cit  en page 15.)

- [Vickland 2010] V. Vickland, G. McDonnell, J. Wern, D. Draper, L. F. Low et H. Brodaty. *A computer model of dementia prevalence in Australia : Forecasting outcomes of delaying dementia onset, slowing disease progression, and eradicating dementia types*. *Dementia and Geriatric cognitive disorders*, vol. 29, pages 123–130, 2010. (Cité en page 98.)
- [von Campenhausen 2005] S. von Campenhausen, B. Bornschein, R. Wick, K. Botzel, C. Sampiao, W. Poewe, W. Oertel, U. Siebert, K. Berger et R. Dodel. *Prevalence and incidence of Parkinson's disease in Europe*. *Eur Neuropsychopharmacol*, vol. 15, pages 473–490, 2005. (Cité en page 104.)
- [Wanneveich 2016] M. Wanneveich, H. Jacqmin-Gadda, J. F. Dartigues et P. Joly. *Impact of intervention targeting risk factors on chronic disease burden*. *Statistical Methods in Medical Research*, vol. in press, 2016. (Cité en pages 67 et 127.)
- [Wilmoth 1990] J. Wilmoth, J. Vallin et G. Caselli. *When Does a Cohort's Mortality Differ from What we Might Expect ?*. *Population*, vol. 2, pages 93–126, 1990. (Cité en page 18.)
- [Winnock 2002] M. Winnock, L. Letenneur, H. Jacqmin-Gadda, J. Dallongeville, P. Amouyel et J. F. Dartigues. *Longitudinal analysis of the effect of apolipoprotein E $\epsilon 4$ and education on cognitive performance in elderly subjects : the PAQUID study*. *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 72, no. 6, pages 794–797, 2002. (Cité en page 78.)
- [Xu 2010] J. Xu, J.D. Kalbfleisch et B. Tai. *Statistical analysis of Illness-Death processes ans semicompeting risks data*. *Biometrics*, vol. 66, no. 3, pages 716–725, 2010. (Cité en page 22.)

Résumé : De nos jours, la Santé Publique porte de plus en plus d'intérêts aux maladies chroniques neuro-dégénératives liées au vieillissement telles que la démence ou la maladie de Parkinson. Ces pathologies ne peuvent ni être évitées ni guéries et occasionnent une détérioration progressive de la santé des malades requérant donc des soins spécifiques. Le contexte démographique actuel laisse entrevoir un vieillissement de la population et un allongement de l'espérance de vie continue, ce qui aura pour conséquence majeure d'aggraver le fardeau économique, social et démographique de ces maladies dans les années à venir. C'est pourquoi, dans l'optique d'anticiper les problèmes à venir, il est important de développer des modèles statistiques permettant de faire des projections et d'évaluer le fardeau futur de ces maladies via divers indicateurs de santé. De plus, il est intéressant de pouvoir donner des projections, en fonction de scénarios potentiels, qui pourraient être mis en place (*e.g.* un nouveau traitement), afin d'évaluer l'impact qu'aurait une telle intervention de Santé Publique, mais aussi de prendre en compte une variation de l'incidence due, par exemple, à des changements de comportement.

Une approche utilisant le modèle illness-death sous une hypothèse Markovienne a été proposée par Joly *et al.* 2013 pour répondre à cet objectif. Cette approche est particulièrement adaptée dans ce contexte, car elle permet, notamment, de prendre en compte le risque compétitif qui existe entre devenir malade ou décéder (pour un individu non-malade). L'utilisation de ce modèle pour faire des projections a l'avantage, d'une part, de faire intervenir des projections démographiques nationales pour mieux capter l'évolution de la mortalité au cours du temps, et d'autre part, de proposer une modélisation adaptée de la mortalité (en distinguant la mortalité des malades, des non-malades et générale).

C'est sur la base de ce modèle que découle tout le travail de cette thèse. Dans une première partie, les hypothèses du modèle existant ont été améliorées ou modifiées dans le but de considérer l'évolution de l'incidence de la maladie au cours du temps, puis de passer d'un modèle Markovien à un modèle semi-Markovien afin de modéliser la mortalité des malades en fonction de la durée passée avec la maladie. Dans une deuxième partie, la méthode initiale, permettant de considérer l'impact d'une intervention mais avec des hypothèses restrictives, a été développée et généralisée pour prendre en compte des interventions plus flexibles. Puis, les expressions mathématiques/statistiques d'indicateurs de santé pertinents ont été développées dans ce contexte dans le but d'avoir un panel de projections permettant une meilleure évaluation du fardeau futur de la maladie. L'application principale de ce travail a porté sur des projections concernant la démence. Cependant, en appliquant ces modèles à la maladie de Parkinson, nous avons proposé des méthodes permettant d'adapter notre approche à d'autres types de données.

Mots clés : Modèle illness-death ; Modèle de Markov/semi-Markov ; Projections ; Maladies Chroniques ; Santé Publique.

Abstract :

Nowadays, Public Health is more and more interested in the neuro-degenerative chronic diseases related to the ageing such as Dementia or Parkinson's disease. These pathologies cannot be prevented or cured and cause a progressive deterioration of health, requiring specific cares. The current demographic situation suggests a continuous ageing of the population and a rise of the life expectancy. As consequence of this, the economic, social and demographic burden related to these diseases will worsen in years to come. That is why, developing statistical models, which allow to make projections and to estimate the future burden of chronic diseases via several health indicators, is becoming of paramount importance. Furthermore, it would be interesting to give projections, according to hypothetical scenarios, which could be set up (*e.g.* a new treatment), to estimate the impact of such Public health intervention, but also to take into account modifications in disease incidence due, for example, to behavior changes.

To attend this objective, an approach was proposed by Joly *et al.* 2013, using the illness-death model under Markovian hypothesis. Such approach has shown to be particularly adapted in this context, since it allows to consider the competitive risk existing between the risk of death and the risk to develop the disease (for a non-diseased subject). On one hand, projections made by using this model take advantage of including national demographic projections to better consider the mortality trends over time. Moreover, the approach proposes to carefully model mortality, distinguishing overall mortality, non-diseased and diseased mortality trends.

All the work of this thesis have been developed based on this model. In a first part, the hypotheses of the existing model are improved or modified in order to both : consider the evolution of disease incidence over time ; pass from a Markov to a semi-Markov hypothesis, which allows to model the mortality among diseased subjects depending on the time spent with the disease. In a second part, the initial method, allowing to take into account the impact of an intervention, but with many restrictive assumptions, is developed and generalized for more flexible interventions. Then, the mathematical/statistical expressions of relevant health indicators are developed in this context to have a panel of projections giving a better assessment of the future disease burden.

The main application of this work concerns projections of Dementia. However, by applying these models to Parkinson's disease, we propose methods which allows to adapt our approach to other types of data.

Keywords : Illness-death model ; Markov/semi-Markov model ; Projections ; Chronic Diseases ; Public Health.
