

Santé travail

Maladie de Parkinson parmi les affiliés à la Mutualité sociale agricole dans cinq départements

Modèles prédictifs à partir des bases
de remboursement de médicaments

Frédéric Moisan, Alexis Elbaz

Sommaire

Abréviations	2
1. Introduction	3
2. Objectifs	3
3. La maladie de Parkinson	4
3.1 Caractéristiques cliniques	4
3.2 Diagnostic	4
3.3 Traitement antiparkinsonien	4
3.4 Données épidémiologiques sur la maladie de Parkinson	7
4. Méthodes	9
4.1 Identification des patients atteints de maladie de Parkinson à partir des bases de remboursement de médicaments de l'Assurance maladie	9
4.2 Définition des prédicteurs	10
4.3 Développement du modèle prédictif et validation interne	13
4.4 Prévalence de la maladie de Parkinson	14
4.5 Relation entre la prévalence de la maladie de Parkinson et les caractéristiques agricoles	14
5. Résultats	18
5.1 Inclusion des participants	18
5.2 Caractéristiques des participants	18
5.3 Développement du modèle prédictif	20
5.4 Analyses de sensibilité	24
5.5 Prévalence de la maladie de Parkinson	24
5.6 Caractéristiques agricoles et maladie de Parkinson	26
6. Discussion	31
6.1 Résumé des principaux résultats	31
6.2 Erreurs de classement	31
6.3 Facteurs influençant les performances du modèle prédictif	32
6.4 Possibilités d'utilisations des modèles prédictifs	33
6.5 Association avec les variables agricoles	34
6.6 Limites	34
6.7 Conclusions	35
Références bibliographiques	36
Annexes	42

Maladie de Parkinson parmi les affiliés à la Mutualité sociale agricole dans cinq départements

Modèles prédictifs à partir des bases de remboursement de médicaments

Rédacteurs

Frédéric Moisan, Institut national de la santé et de la recherche médicale (Inserm) U708 – Neuroépidémiologie ; Université Pierre et Marie Curie (UPMC, Paris 6)

Alexis Elbaz, Inserm U708 – Neuroépidémiologie ; UPMC, Paris 6 ; Département santé travail (DST), Institut de veille sanitaire (InVS)

Groupe de travail

Jean-Luc Dupupet (Caisse centrale de la Mutualité sociale agricole, MSA)

Marcel Goldberg (Inserm U1018 ; Université de Versailles Saint-Quentin-en-Yvelines ; Université Paris-Sud 11)

Jean Houssinot (Caisse centrale de la MSA)

Ellen Imbernon (DST, InVS)

Jean-Louis Mazurie (Caisse de MSA de Gironde)

Christophe Tzourio (Inserm U708 ; UPMC, Paris 6)

Remerciements

Nous remercions les médecins de la MSA (Drs Jacques Aïmedieu, Daniel Albert, Catherine Bolut, Christophe Fuzeau, Virginie Gaussères, Maryline Grandjean, Jean Houssinot, Marine Jeantet, Bernard Ladépêche, Didier Menu, Omar Tarsissi), les responsables des extractions de données à partir des bases de remboursements de médicaments de la MSA (Aude Alemant, Joël Gourgues, Sandrine Nogues, Hervé Pauly, Emilie Richard, Pierre Vannier), les enquêtrices (Véronique Dumay, Viviane Palleau, Frédérique Pellerin, Estelle Seguin, Sophie Sinibaldi), les neurologues (Irina Balaboi, Isabelle Benatru, Julien Dumurgier, Elsa Krim, Danièle Ranoux), Véronique Gourlet (analyse statistique), Yann Hamon (data manager), Aïcha Soumaré (coordination) et toutes les personnes qui ont participé à ce projet.

Financement

Alexis Elbaz bénéficie d'un contrat d'interface InVS-Inserm. Frédéric Moisan a bénéficié d'une bourse de thèse du ministère chargé de l'Enseignement supérieur et de la Recherche et de la Fondation pour la recherche médicale.

Ce projet a bénéficié d'un financement de l'Agence nationale de la recherche (ANR santé-environnement et santé-travail, SEST, 2005), de l'Agence française de sécurité sanitaire de l'environnement et du travail (Afsset) et de France Parkinson.

Abréviations

ALD	Affection de longue durée
ATC	Classification anatomique, thérapeutique et chimique
CIP	Code identifiant de présentation
Cnil	Commission nationale de l'informatique et des libertés
COMT	Catéchol-O-méthyltransférase
EDL	Equivalent de dose de lévodopa
ES	Erreur standard
ET	Ecart-type
IC	Intervalle de confiance
MBS	Marge brute standard
MP	Maladie de Parkinson
MPP+	1 - méthyle - 4 - phenylpyridinium
MPTP	1 - méthyle - 4 - phényl 1,2,3,6 - tétrahydro pyridine
MSA	Mutualité sociale agricole
ND	Non disponible
NR	Non retenu
OR	Odds ratio
ORL	Oto-rhino-laryngologique
Otex	Orientation technico-économique des exploitations
PF	Polynôme fractionnaire
PMSI	Programme de médicalisation des systèmes d'information
ROC	Receiver Operating Characteristics
SJSR	Syndrome des jambes sans repos

1. Introduction

La surveillance épidémiologique nécessite de pouvoir identifier de manière fiable et répétée au cours du temps des patients atteints des pathologies à surveiller. En fonction des pathologies et du contexte, différentes possibilités sont envisageables : certificats de décès, codage des hospitalisations, registres spécifiques, déclarations de maladie obligatoires, etc. Pour de nombreuses pathologies, ces sources présentent cependant des limites et manquent de sensibilité ou de spécificité. Il est également possible de réaliser des enquêtes avec un recueil individuel d'information, mais le nombre de patients identifiés est souvent limité, en particulier lorsqu'il s'agit de maladies peu fréquentes.

En France, comme dans de nombreux pays, les bases de remboursement de médicaments de l'Assurance maladie sont maintenant accessibles au niveau national et représentent un moyen supplémentaire d'identifier des patients à partir de leur profil de consommation médicamenteuse. Ces bases permettent de recueillir des informations détaillées sur les médicaments consommés, mais aussi sur la fréquence et la régularité d'achat, sur le dosage des médicaments achetés, la spécialité du prescripteur ou certaines caractéristiques démographiques des assurés comme l'âge ou le sexe. Si les bases de remboursement de médicaments de l'Assurance maladie étaient une source valide, c'est-à-dire suffisamment sensible et spécifique, elles représenteraient un moyen d'identifier des patients atteints des pathologies d'intérêt peu coûteux et pouvant être mis en œuvre de manière répétée.

2. Objectifs

Avant d'utiliser les bases de remboursement de médicaments pour identifier des patients atteints d'une maladie, il est au préalable nécessaire d'étudier quelle est leur validité. Compte tenu de notre intérêt pour la maladie de Parkinson chez les agriculteurs [1], nous avons pris comme exemple cette pathologie et essayé de répondre aux questions suivantes :

- peut-on identifier des patients parkinsoniens à partir des bases de remboursement de médicaments ?
- des informations telles que la dose de médicament, la régularité des prescriptions, la spécialité du prescripteur ou des caractéristiques démographiques sont-elles utiles à prendre en compte ?
- quelle est la validité de cette approche en termes de sensibilité et de spécificité ?
- comment estimer la prévalence de la maladie de Parkinson à partir des bases de remboursement de médicaments et étudier sa relation avec une exposition environnementale ?

Pour tenter de répondre à ces questions, nous avons identifié sur une période d'un an (2007) les personnes ayant consommé au moins un médicament pouvant être utilisé pour traiter la maladie de Parkinson parmi les affiliés à la Mutualité sociale agricole (MSA) dans cinq départements français et nous avons recueilli auprès de ces personnes des informations cliniques afin de déterminer si elles étaient effectivement atteintes de la maladie de Parkinson.

3. La maladie de Parkinson

La maladie de Parkinson est la cause la plus fréquente de syndrome parkinsonien [2]. Elle est liée à la perte progressive des neurones dopaminergiques de la *substantia nigra pars compacta*, qui est une structure mésencéphalique impliquée dans la régulation de l'activité des noyaux gris centraux. Cette perte neuronale s'accompagne de la présence d'inclusions neuronales caractéristiques (corps de Lewy).

3.1 Caractéristiques cliniques

La maladie de Parkinson est caractérisée par l'association de quatre symptômes cardinaux qui définissent le syndrome parkinsonien : tremblement de repos, rigidité extrapyramidale (avec « en roue dentée »), akinésie (lenteur du mouvement), instabilité posturale. Ces signes, qui ont fréquemment un caractère asymétrique, sont le plus souvent associés, mais certains d'entre eux peuvent être absents, en particulier au début de l'évolution de la maladie ; de plus, le tremblement de repos n'est pas constant.

Des symptômes non moteurs sont également fréquents (par exemple, troubles de l'humeur, constipation, troubles du sommeil paradoxal, perte de l'odorat, déclin cognitif) [3]. Certains d'entre eux peuvent apparaître au cours de l'évolution de la maladie (par exemple, déclin cognitif), tandis que d'autres (par exemple, anosmie, constipation, troubles de l'humeur ou du sommeil paradoxal), peuvent être présents d'emblée, ou même exister des années avant l'apparition des signes moteurs [4].

En revanche, dans la maladie de Parkinson, le syndrome parkinsonien n'est pas associé à d'autres anomalies de l'examen neurologique qui témoigneraient d'une atteinte plus étendue du système nerveux. Il n'existe pas de troubles de l'oculomotricité (en faveur de la paralysie supranucléaire progressive), de syndrome cérébelleux ou d'hypotension orthostatique (en faveur de l'atrophie multisystématisée), ou de détérioration cognitive précoce (en faveur de la démence à corps de Lewy). Il est important d'éliminer la prise de neuroleptiques avant le début des symptômes (en faveur d'un syndrome parkinsonien iatrogène).

3.2 Diagnostic

Le diagnostic de maladie de Parkinson repose essentiellement sur un recueil précis de l'histoire de la maladie et l'examen clinique neurologique. Les principaux critères sur lesquels repose le diagnostic sont les suivants :

- syndrome parkinsonien ;
- d'aggravation progressive ;
- en l'absence d'une autre cause (pas de prise de neuroleptiques ni de signe en faveur d'une atteinte plus étendue du système nerveux) ;
- amélioration significative et prolongée par le traitement antiparkinsonien à une dose suffisante.

Des examens complémentaires (par exemple, résonance magnétique cérébrale, bilan du cuivre) peuvent être utilisés au cas par cas, le plus souvent pour éliminer d'autres causes de syndrome parkinsonien, mais aucun examen complémentaire ne permet d'établir le diagnostic avec certitude en dehors de l'examen anatomopathologique du cerveau.

Le diagnostic de maladie de Parkinson est donc principalement clinique et il existe un risque d'erreur diagnostique. Les principales erreurs diagnostiques sont les autres syndromes parkinsoniens neurodégénératifs (paralysie supranucléaire progressive, atrophie multisystématisée, dégénérescence corticobasale) ou liés à la prise de neuroleptiques et le tremblement essentiel.

Le recours à un avis spécialisé permet de diminuer le risque d'erreur [5;6]. Dans une série clinico-pathologique récente, 10 % des sujets ayant un diagnostic de maladie de Parkinson avaient un autre diagnostic après examen anatomopathologique du cerveau [7]. Toutefois, cette étude a été réalisée dans un centre spécialisé dans la prise en charge des patients parkinsoniens et il est probable que ses résultats ne puissent pas être étendus directement à la population générale (patients sélectionnés ; médecins spécialisés).

3.3 Traitement antiparkinsonien

Les symptômes moteurs de la maladie de Parkinson sont considérablement améliorés par le traitement antiparkinsonien ; il s'agit d'ailleurs de l'un des critères pris en compte pour établir le diagnostic.

3.3.1 Traitement médical symptomatique

Le traitement de la maladie de Parkinson est avant tout médical et symptomatique ; il vise à améliorer les symptômes de la maladie mais ne permet pas de ralentir son évolution. Des médicaments appartenant à différentes classes thérapeutiques peuvent être utilisés, souvent en association (tableau 1).

| Tableau 1 |

Médicaments antiparkinsoniens disponibles en France en 2007

Familles de médicaments antiparkinsoniens	Médicaments antiparkinsoniens	Equivalent de dose de lévodopa (EDL) en mg pour 100 mg lévodopa ^a
Agonistes dopaminergiques de type 1 ^b	Pramipexole (forme sel)	1
	Ropinirole	5
	Pergolide	1
Agonistes dopaminergiques de type 2 ^c	Apomorphine	10
	Bromocriptine	10
	Lisuride	1
Amantadine	Amantadine	100
Anticholinergiques	Trihexyphenidyl	ND
	Bipéridène	ND
	Tropatépine	ND
Inhibiteurs de la COMT	Entacapone	ND
	Tolcapone	ND
Lévodopa	Lévodopa + carbidopa/bensérazide	100 ; 75 ^d ; 133 ^e
Piribédil	Piribédil	100
Selegiline	Sélégiline	10

Abréviations : COMT : catéchol-O-méthyltransférase ; ND : non disponible.

^aLes équivalences de doses sont définies à partir d'une synthèse de 56 études [40].

^bLe pramipexole, ropinirole et pergolide sont souvent prescrits comme traitement de la maladie de Parkinson.

^cLe lisuride, la bromocriptine et l'apomorphine sont rarement prescrits comme traitement de la maladie de Parkinson.

^dDans le cas où des inhibiteurs de la COMT sont prescrits en même temps que la lévodopa. Il s'agit exclusivement d'entacapone (aucun utilisateur de tolcapone n'a été identifié dans cette étude).

^eLibération prolongée.

Aucun de ces médicaments n'est spécifique de la maladie de Parkinson et ils peuvent être utilisés pour traiter d'autres pathologies ; dans ce cas, la dose utilisée est généralement inférieure et les traitements sont prescrits moins régulièrement [8;9]. Bien sûr, ils peuvent également être utilisés chez des patients ayant reçu un diagnostic de maladie de Parkinson par erreur.

- Lévodopa : il s'agit du traitement de référence car le plus efficace. Son utilisation prolongée s'accompagne d'un risque de complications motrices (dyskinésies, fluctuations) qui peuvent être très gênantes. La lévodopa peut, plus rarement, être utilisée pour traiter d'autres maladies, comme certaines formes rares de dystonie sensible à la levodopa. De plus, les autres syndromes parkinsoniens neurodégénératifs peuvent, parfois et pendant une période de temps limitée, être améliorés partiellement par la levodopa, ce qui explique que ce médicament est très souvent utilisé pour traiter ces pathologies puisqu'aucun autre traitement n'est disponible.

- Inhibiteurs de la catéchol-O-méthyltransférase (COMT) : il s'agit d'un médicament qui est uniquement prescrit en association avec la lévodopa afin de prolonger la durée de son effet en inhibant sa transformation périphérique.

- Agonistes dopaminergiques : il s'agit d'une alternative intéressante à l'utilisation de la lévodopa. Ils ne sont pas indiqués en priorité après 70 ans car leurs effets secondaires sont plus fréquents après cet âge. En revanche, il existe un consensus pour utiliser ce type de traitement en priorité chez les patients les plus jeunes (avant 60 ans) afin de retarder la mise sous lévodopa et la survenue des complications motrices ; si son effet est insuffisant, on peut y associer de la lévodopa, mais à une dose plus faible que si elle avait été prescrite seule [10]. La dysrégulation dopaminergique est un effet secondaire relativement fréquent de cette classe de médicaments se traduisant par des troubles du comportement (addiction, hypersexualité, etc.) et nécessite d'être dépistée afin de diminuer la dose ou de changer de médicament.

Plusieurs agonistes existent. Certains, comme le ropinirole ou le pramipexole, sont souvent utilisés dans la maladie de Parkinson. Ces deux médicaments sont également utilisés pour traiter une autre pathologie, d'ailleurs plus fréquente que la maladie de Parkinson, le syndrome des jambes sans repos (SJSR). Dans ce cas, la dose utilisée est plus faible que dans la maladie de Parkinson et prescrite généralement en une seule prise avant le coucher. D'autres agonistes sont plus rarement utilisés dans la maladie de Parkinson tandis qu'ils sont souvent utilisés dans d'autres indications : c'est par exemple le cas de la bromocriptine, principalement utilisée pour inhiber la lactation après la grossesse ou comme traitement de certaines tumeurs cérébrales associées à une hyperprolactinémie ; de même, le lisuride et la cabergoline sont rarement utilisées dans la maladie de Parkinson. L'apomorphine est un agoniste dopaminergique avec une courte demi-vie d'élimination qui est généralement injecté par voie sous-cutanée, en particulier chez des patients présentant des formes évoluées avec des complications motrices.

- Piribédil : il s'agit d'un médicament faisant partie des agonistes dopaminergiques mais dont les modalités de prescription sont très différentes. Il a des indications variées : "vieillesse cérébrale", artériopathies chroniques des membres inférieurs, pathologies oto-rhino-laryngologiques (ORL), manifestations ischémiques en ophtalmologie. Il peut être utilisé dans la maladie de Parkinson mais à des doses plus élevées que dans les autres indications.

- Anticholinergiques : ils peuvent être utilisés dans la maladie de Parkinson, en particulier dans les formes avec un tremblement important, mais leur usage est limité en raison d'effets secondaires importants, en particulier cognitifs. Ils sont en revanche souvent utilisés pour corriger les effets secondaires des neuroleptiques chez les patients psychotiques.

- Amantadine : ce traitement est exceptionnellement utilisé de nos jours pour traiter la maladie de Parkinson, sauf dans des cas particuliers (dyskinésies). Il est également prescrit dans d'autres indications comme la prophylaxie de la grippe ou le traitement de la fatigue dans la sclérose en plaques.

3.3.2 Traitement chirurgical

La stimulation cérébrale profonde est un traitement chirurgical de développement récent qui est principalement proposé à des patients jeunes chez qui le traitement médical est associé à des complications motrices invalidantes et sans contre-indications à ce type de traitement. Elle est pratiquée dans un nombre limité de centres spécialisés et reste un traitement cher dont les indications sont encore limitées. Lorsqu'elle est pratiquée, le traitement symptomatique médicamenteux reste indispensable, quoi qu'à des doses généralement inférieures. Elle peut être pratiquée pour d'autres indications, comme la dystonie généralisée ou le tremblement essentiel.

3.3.3 Traitement neuroprotecteur

Deux médicaments sont utilisés dans cette indication et leur usage est limité à la maladie de Parkinson. La sélégiline a été le premier d'entre eux et son efficacité n'a pas été démontrée mais elle est toujours commercialisée. La rasagiline a récemment été commercialisée en France suite à un essai thérapeutique suggérant qu'à la dose d'un milligramme par jour, ce traitement permettrait de ralentir l'évolution de la maladie ; elle n'était pas disponible au moment de l'étude.

3.4 Données épidémiologiques sur la maladie de Parkinson

3.4.1 Identification des patients parkinsoniens dans les études épidémiologiques

Dans le cadre des études épidémiologiques sur la maladie de Parkinson, les études en porte-à-porte restent la méthode de référence [11-13]. Cette approche consiste à dépister les patients présentant un syndrome parkinsonien dans une population définie grâce à des questionnaires et/ou des épreuves cliniques simples et à examiner ensuite les personnes dépistées positives afin d'établir un diagnostic ; cette approche est toutefois coûteuse et difficile à mettre en œuvre. De plus, compte tenu de la fréquence de la maladie, elle permet habituellement d'identifier un petit nombre de personnes atteintes. Elles ont l'avantage d'identifier les patients parkinsoniens qui ne sont pas diagnostiqués et traités ; toutes les autres méthodes décrites ci-dessous permettent d'identifier uniquement les patients qui ont reçu un diagnostic. La proportion des cas non diagnostiqués est variable en fonction du contexte, de l'accès aux soins, ou des caractéristiques socio-économiques [14] ; elle dépend fortement de l'âge et dans une étude de prévalence réalisée en France dans l'étude Paquid, la grande majorité des patients non diagnostiqués étaient âgés de plus de 80 ans [15].

Les certificats de décès ne sont pas considérés comme étant une source fiable car plusieurs études ont montré que la maladie de Parkinson est souvent négligée sur les certificats de décès [16;17] : aux Etats-Unis, dans un cas sur deux, la maladie de Parkinson n'apparaissait pas sur le certificat de décès de patients parkinsoniens.

Les bases incluant un codage des motifs d'hospitalisation comme le Programme de médicalisation des systèmes d'information (PMSI) ne sont pas une source fiable ; d'une part, la maladie de Parkinson n'est pas un motif fréquent d'hospitalisation, en particulier dans les formes non évoluées, et, d'autre part, la qualité du codage des motifs d'hospitalisation dépend en grande partie du codeur [18].

Quelques études ont identifié des patients atteints de maladie de Parkinson à partir de données d'utilisation de médicaments antiparkinsoniens [19-25], mais la fiabilité de cette approche reste inconnue puisque, dans la plupart des ces études, cette source n'a pas été validée en comparaison d'un diagnostic de référence. La plupart des études ont utilisé la levodopa comme médicament "traceur", tandis qu'un plus petit nombre d'études ont également inclus d'autres types de médicaments.

3.4.2 Fréquence

La maladie de Parkinson est la maladie neurodégénérative la plus fréquente après la maladie d'Alzheimer et la première cause de syndrome parkinsonien. Son incidence est habituellement comprise entre 10 et 20 pour 100 000 personnes-années ; la maladie est exceptionnelle avant 50 ans et sa fréquence augmente après 60 ans avec une incidence un peu plus élevée chez les hommes que chez les femmes à tous les âges [26;27]. Le risque vie entière de maladie de Parkinson a été estimé comme étant approximativement de 2,0 % chez les hommes et 1,3 % chez les femmes, après prise en compte de la différence d'espérance de vie entre les deux sexes [28]. Dans le Minnesota entre 1976 et 2000, la mortalité des patients parkinsoniens était environ 1,6 fois plus élevée que celles de sujets non atteints de même âge et sexe [16].

La prévalence de la maladie de Parkinson est généralement comprise entre 100 et 200 pour 100 000 habitants [27]. Elle augmente avec l'âge et elle est approximativement égale à 2 % après 65 ans [29].

A titre de comparaison, l'incidence de la paralysie supranucleaire progressive et de l'atrophie multisystématisée est d'environ 1,0 pour 100 000 personnes-années [30], avec une prévalence de l'ordre de 1 à 10 pour 100 000 personnes [31].

3.4.3 Agriculture et maladie de Parkinson

Il y a plus de 20 ans, l'identification de syndromes parkinsoniens suite à l'exposition au 1 - méthyle 4 - phényl 1,2,3,6-tétrahydro pyridine (MPTP) a été à l'origine d'études sur la relation entre l'exposition aux pesticides et la maladie de Parkinson [32]. Le MPTP pénètre dans les neurones dopaminergiques où il est métabolisé en 1-méthyl-4-phenylpyridinium (MPP+) qui est un puissant inhibiteur du complexe I mitochondrial et a une structure chimique proche de celle du paraquat, un herbicide qui a été largement utilisé dans de nombreux pays.

Plusieurs études cas-témoins et un plus petit nombre d'études de cohorte ont montré une association entre la maladie de Parkinson et le métier d'agriculteur ou l'exposition professionnelle aux pesticides. Une méta-

analyse d'études cas-témoins a estimé un odds ratio de 1,9 pour l'utilisation professionnelle de pesticides et de 1,4 pour le métier d'agriculteur [33;34].

L'une des principales difficultés rencontrées par ces études est liée à l'évaluation rétrospective de l'exposition aux pesticides et la plupart d'entre elles ont utilisé des méthodes peu précises. Des études plus récentes ont utilisé des méthodes d'évaluation de l'exposition plus détaillées pour tenter d'identifier des familles de produits mais leurs résultats sont encore peu cohérents [35-38]. Il existe donc une association générique entre l'exposition aux pesticides et la maladie de Parkinson mais il existe peu de données pour l'instant sur les pesticides responsables de cette association, en grande partie en raison de difficultés pour l'évaluation de l'exposition à ces produits.

Les caractéristiques de l'utilisation des pesticides (fréquence, type de produits, modalités d'épandage) dépendent étroitement du type de culture [39], or celui-ci est considérablement plus facile à évaluer que l'exposition aux pesticides, soit par interrogatoire soit par des données disponibles comme le recensement agricole. Une approche permettant de contribuer à mieux caractériser le type d'exposition aux pesticides associée avec la maladie de Parkinson pourrait donc consister à étudier la relation entre la maladie de Parkinson et le type d'agriculture [1] ; c'est l'approche que nous avons choisi de suivre pour ce rapport.

4. Méthodes

Cette étude a été menée auprès des affiliés de la MSA, âgés de 18 ans ou plus, dans cinq départements français (Charente-Maritime, Côte-d'Or, Gironde, Haute-Vienne, Mayenne). La MSA assure la protection sociale du monde agricole et rural et elle est à ce titre responsable du remboursement des frais de santé des travailleurs dans l'agriculture et des professions connexes (exploitants ; ouvriers agricoles ; travailleurs dans les silos, magasins de semences ou coopératives agricoles ; jardiniers professionnels ; employés de la MSA, d'une compagnie d'assurance et d'une banque). Ces personnes bénéficient de l'Assurance maladie pendant qu'elles sont en activité ou retraitées. La MSA couvre leurs ayants-droit (conjoint non actifs et enfants). En 2007, la MSA constituait le deuxième régime de protection sociale en France, avec environ 3 800 000 personnes protégées en maladie.

Le protocole de l'étude a reçu un avis favorable du Comité de protection des personnes de l'Hôpital de la Pitié-Salpêtrière et a fait l'objet d'une déclaration à la Commission nationale de l'informatique et des libertés (Cnil).

4.1 Identification des patients atteints de maladie de Parkinson à partir des bases de remboursement de médicaments de l'Assurance maladie

En France, les traitements antiparkinsoniens ne peuvent être délivrés sans prescription médicale. Leur délivrance est enregistrée dans les bases de remboursement de médicaments de l'Assurance maladie.

A partir des bases de remboursement de médicaments de la MSA dans cinq départements nous avons identifié les personnes qui avaient acheté en 2007 au moins un médicament potentiellement antiparkinsonien (défini comme tout médicament pouvant être utilisé pour traiter la maladie de Parkinson et identifié à partir des codes N04, G02CB01 et G02CB02 de la classification anatomique, thérapeutique et chimique des médicaments, ATC ; tableau 1), et qui vérifiaient les critères d'inclusion suivants :

- âge au 1^{er} janvier 2007 inférieur ou égal à 80 ans ;
- durée de la maladie inférieure ou égale à 15 ans. Ce critère était uniquement appliqué d'emblée pour les personnes qui étaient en affection de longue durée (ALD) pour maladie de Parkinson (ALD 16) et pour qui la date de première demande d'ALD était connue ;
- pas d'ALD pour démence (ALD 15) ou maladie psychiatrique (ALD 23).

Toutes les personnes ayant acheté en 2007 au moins une fois un médicament à base de lévodopa, entacapone, tolcapone, ropinirole, pramipexole, apomorphine, bromocriptine ou sélégiline ont été directement contactées par téléphone et invitées à être examinées par un neurologue afin de confirmer la maladie de Parkinson sauf :

- si elles déclaraient prendre de petites doses d'agonistes dopaminergiques pour un SJSR ;
- si le traitement a été définitivement interrompu après moins d'un mois de traitement ;
- s'il existait une histoire bien documentée de syndrome parkinsonien iatrogène ;
- si la durée d'évolution de la maladie était supérieure à 15 ans.

Afin d'établir le diagnostic de maladie de Parkinson, nous avons utilisé les critères décrits par Bower et collègues [2] :

- syndrome parkinsonien défini par la présence d'au moins deux signes cardinaux sur quatre ;
- d'aggravation progressive ;
- en l'absence d'une autre cause (pas de prise de neuroleptiques ni de signe en faveur d'une atteinte plus étendue du système nerveux) ;
- amélioration significative et prolongée par le traitement antiparkinsonien à une dose suffisante.

Les personnes ayant uniquement consommé du piribédil, de l'amantadine ou des anticholinergiques ont d'abord été contactées par courrier afin de connaître le motif de prescription car ces médicaments sont rarement utilisés pour traiter la maladie de Parkinson ; les personnes qui ont répondu que ces médicaments étaient prescrits pour une maladie de Parkinson ou un syndrome parkinsonien ou qui ne connaissaient pas le motif de prescription ont été invitées à être examinées par le neurologue de l'étude.

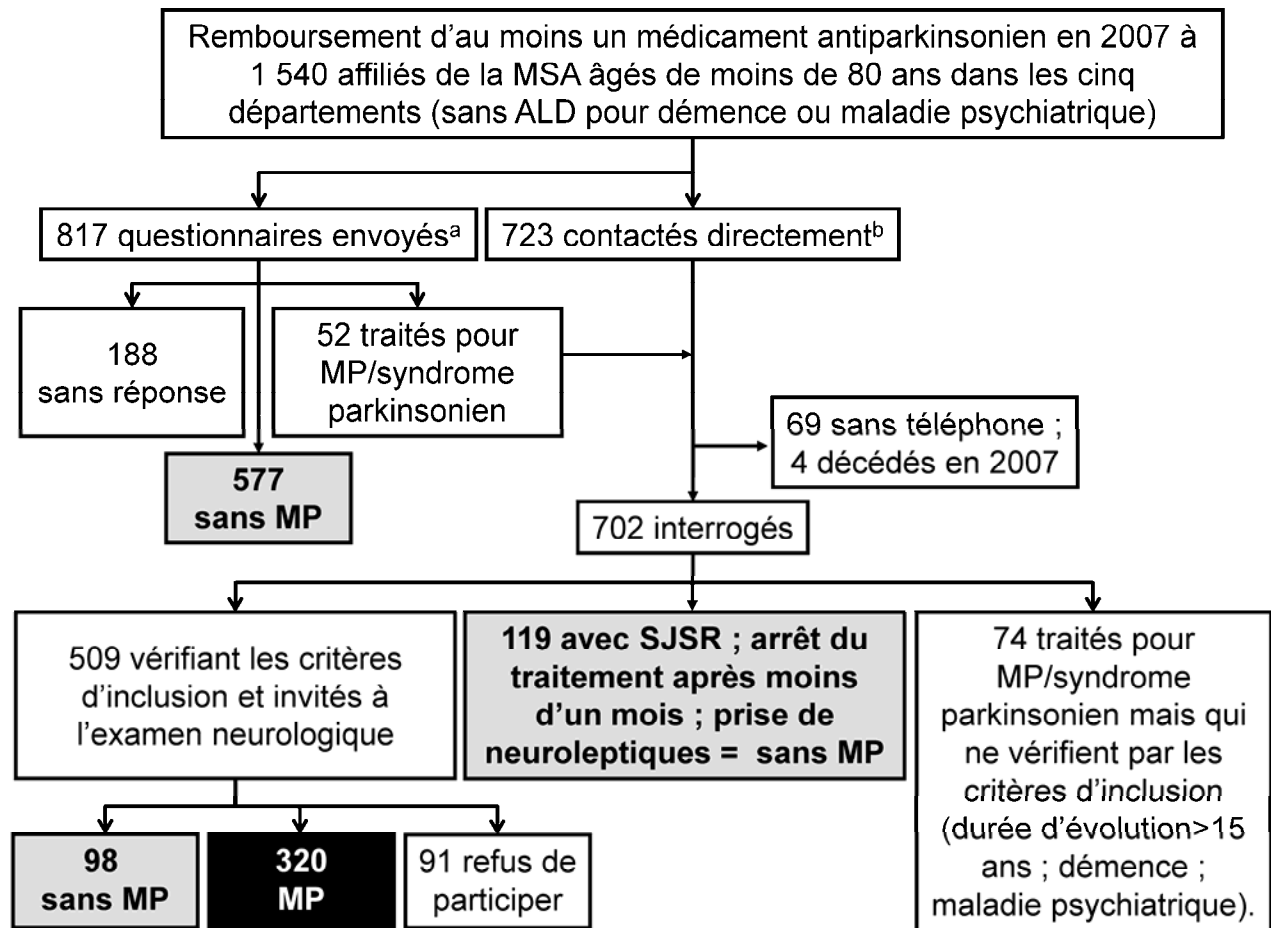
Nous n'avons pas contacté les femmes âgées de 50 ans ou moins utilisant de petites doses de bromocriptine pendant de courtes périodes (inhibition de lactation) et les personnes utilisant des anticholinergiques en association avec des neuroleptiques (syndrome parkinsonien d'origine iatrogène).

Nous n'avons pas pu identifier les sujets institutionnalisés dans des maisons de retraite avec des pharmacies à usage interne car les médicaments qui leur sont délivrés ne sont pas inclus dans les bases de remboursement de médicaments sous leur identité.

L'inclusion des participants dans l'étude est résumée dans la figure 1.

Figure 1

Inclusion des participants dans l'étude



Abréviations : ALD : affection de longue durée ; MP : maladie de Parkinson ; MSA : Mutualité sociale agricole ; SJSR : syndromes des jambes sans repos.

Les cases grisées correspondent à 794 (577+119+98) sujets considérés comme n'ayant pas la maladie de Parkinson. La boîte noire inclus 320 sujets pour qui le neurologue de l'étude a conclu à un diagnostic de maladie de Parkinson après examen neurologique. Par conséquent, 794 sujets sans maladie de Parkinson et 320 sujets atteints de maladie de Parkinson ont été inclus dans cette étude.

^a*Sujets avec au moins une prescription de piribédil, amantadine ou anticholinergiques en 2007 et qui n'ont reçu aucun autre médicament antiparkinsonien.*

^b*Sujets avec au moins une prescription de lévodopa, entacapone, tolcapone, ropinirole, pramipexole, apomorphine, bromocriptine ou sélégiline en 2007 ; parmi ces sujets, 374 sujets étaient en ALD pour la maladie de Parkinson.*

4.2 Définition des prédicteurs

A partir des bases de remboursement de médicaments, nous avons sélectionné les variables suivantes pour chaque médicament délivré durant l'année 2007 aux personnes identifiées précédemment :

- code identifiant de présentation (CIP) et de la classification ATC ;
- dosage par comprimé de la substance active ;

- nombre de comprimés par boîte ;
- date de la prescription et de l'achat ;
- spécialité du prescripteur.

Les informations suivantes ont également été obtenues à partir des bases de la MSA :

- sexe ;
- âge au 1^{er} janvier 2007 ;
- canton de résidence ;
- statut vital au cours de 2007 et date de décès le cas échéant ;
- ALD pour maladie de Parkinson (ALD 16) et date de mise en ALD 16 le cas échéant.

A partir de ces variables, les variables suivantes ont été définies :

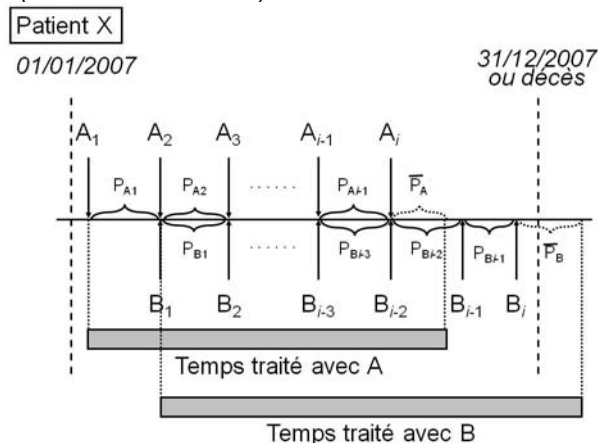
- nombre de consultations avec un neurologue par an. Le codage de la spécialité des prescripteurs ne permet pas d'identifier les neurologues hospitaliers ;
- nombre de consultations par an avec un médecin hospitalier, quelle que soit sa spécialité ;
- nombre de consultations par an avec un médecin généraliste ;
- nous avons combiné les antiparkinsoniens appartenant aux mêmes classes thérapeutiques afin de réduire le nombre de prédicteurs et définir des variables avec assez de sujets exposés (tableau 1). Pour chaque classe d'antiparkinsonien, nous avons défini des variables binaires (oui/non) et calculé différentes variables quantitatives (figure 2) :
 - nombre de prescriptions par an,
 - nombre de boîtes achetées par an,
 - dose cumulée annuelle (nombre de boîtes achetées par an multiplié par le nombre de comprimés par boîte et par le dosage des comprimés),
 - délai entre le début du traitement et la fin du traitement,
 - dose moyenne journalière (dose cumulée annuelle divisée par le délai entre le début du traitement et la fin du traitement),
 - afin de faciliter les regroupements en classes plus larges, nous avons exprimé la dose de lévodopa, piribédil, pramipexole, ropinirole, bromocriptine, apomorphine, et pergolide en équivalents de lévodopa (tableau 1) [40]. Pour les classes d'antiparkinsoniens, les doses cumulées annuelles et les doses moyennes journalière ont été calculées en équivalents de lévodopa lorsque tous les antiparkinsoniens inclus dans la classe pouvaient être exprimés en équivalents de lévodopa (lévodopa ; piribédil ; agonistes dopaminergiques de type 1),
 - afin d'évaluer la régularité de la prise d'antiparkinsoniens nous avons estimé la proportion de temps pendant laquelle chaque sujet a été traité (délai entre le début traitement et la fin du traitement ou la fin du suivi défini comme le 31 décembre 2007 ou le décès).

I Figure 2 I

Définitions des doses moyennes (A) pour chaque médicament antiparkinsonien et de la proportion du temps traité (B)

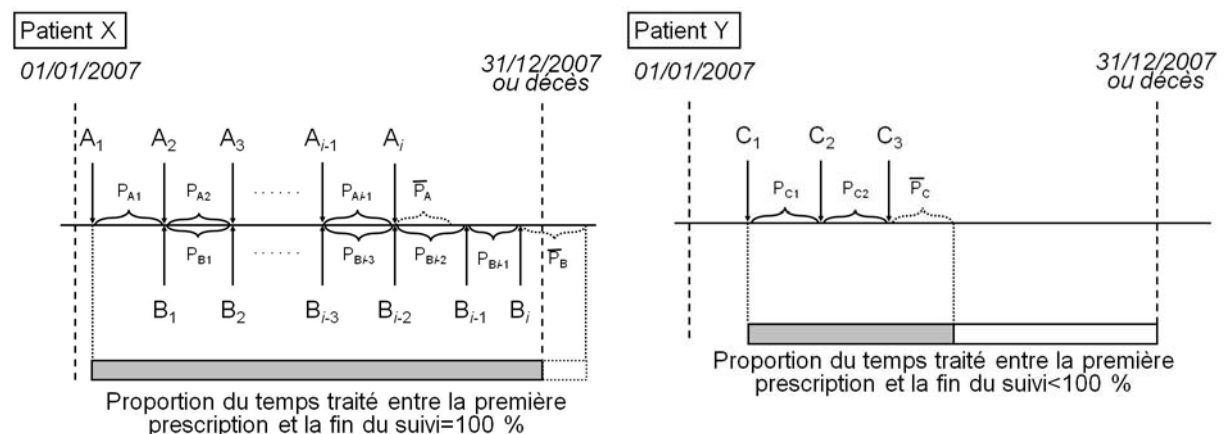
A- Dose moyenne journalière : le patient X a d'abord été traité avec l'antiparkinsonien A et il a par la suite reçu l'antiparkinsonien B. L'antiparkinsonien A a été arrêté en 2007, mais le traitement avec B a été poursuivi. Pour les deux antiparkinsoniens, une date de fin de traitement a été calculée en ajoutant à la date de la dernière prescription de 2007, le délai moyen entre deux prescriptions pour le patient X. Pour les patients n'ayant reçu qu'une fois un antiparkinsonien en 2007, la date de fin de traitement a été calculée en ajoutant à la date de la première prescription, le délai moyen entre deux prescriptions du même médicament calculé parmi tous les participants ayant reçu ce médicament.

Pour les deux antiparkinsoniens, la dose moyenne quotidienne a été calculée en divisant la dose cumulée prescrite en 2007 (nombre de boîtes achetées durant l'année multiplié par le nombre de comprimés par boîte multiplié par le dosage des comprimés) par le délai entre la date de première prescription et la date de fin de traitement calculée (durée du traitement).



B- Proportion du temps traité sur un an : le patient X a d'abord été traité avec l'antiparkinsonien A et il a par la suite reçu l'antiparkinsonien B. L'antiparkinsonien A a été arrêté en 2007, mais le traitement par B a été poursuivi. En conséquence, la proportion du temps traité entre la première prescription de l'antiparkinsonien A et le 31 décembre 2007 ou la mort, est de 100 %.

L'antiparkinsonien C a été prescrit au patient Y trois fois en 2007 ; la date de fin de traitement calculée est antérieure à la date de décès ou au 31 décembre 2007. Par conséquent, la proportion du temps traité entre la première prescription de l'antiparkinsonien C et la fin du suivi est inférieure à 100 %.



Par exemple, un participant a reçu 12 prescriptions (8 pour 1 boîte, 4 pour 2 boîtes) de lévodopa en 2007. La première a eu lieu le 9 janvier et la dernière le 17 décembre ; le délai moyen entre deux prescriptions est de 25 jours et la date de fin de traitement calculée est le 11 Janvier 2008. Les boîtes contiennent 100 comprimés (100 mg/comprimé). En conséquence, pour ce participant :

- le nombre de prescriptions par an est de 12 ;
- le nombre de boîtes par an est de 16 (1 x 8 + 2 x 4) ;
- la dose cumulée par an est 160 000 mg (1 x 8 x 100 x 100 + 2 x 4 x 100 x 100) ;
- la dose moyenne journalière est de 436 mg (160 000 / 367) ;
- la proportion du temps traité est de 100 %.

4.3 Développement du modèle prédictif et validation interne

Nous avons utilisé la régression logistique afin de développer un modèle prédictif du statut parkinsonien (oui/non) en suivant une approche pas à pas structurée [41] ; une description détaillée de cette approche et des mesures utilisées pour évaluer les performances du modèle sont disponibles en annexe 1.

Nous avons choisi de présenter comme principal modèle celui qui a été développé à partir de la dose cumulée des médicaments antiparkinsoniens prescrits pendant un an (2007) ; des modèles développés à partir d'autres mesures quantitatives de la consommation de médicaments seront plus brièvement présentés.

Nous avons utilisé les polynômes fractionnaires pour transformer les variables continues (âge, proportion de temps traité ; pour les classes d'antiparkinsoniens : doses cumulées et moyennes quotidiennes, nombre d'ordonnances et de boîtes achetées) à l'aide des commandes `fracpoly` et `mfp` du logiciel Stata® (version 10, StataCorp LP, collège Station, Texas). Cette approche permet d'obtenir une transformation flexible des variables dans un modèle multivarié à l'aide d'un algorithme itératif [42;43]. Pour les variables caractérisant l'utilisation des classes d'antiparkinsoniens, cette approche a été utilisée uniquement pour ceux qui étaient utilisés par au moins 55 personnes ; pour les autres classes, nous avons utilisé des variables binaires (oui/non). Les variables correspondantes au nombre de consultations avec un neurologue ou un généraliste ont été définies comme des variables à trois ou quatre classes respectivement.

Nous avons tout d'abord inclus dans le modèle l'ensemble des prédicteurs ($n=18$). Nous avons ensuite utilisé une procédure de sélection descendante pas-à-pas ($p \leq 0,20$ pour retenir les variables dans le modèle). Pour chaque variable retenue dans le modèle multivarié, nous avons calculé le rapport entre le coefficient de régression et son erreur standard (*Z-ratio*) ; la comparaison des *Z-ratio* associés à différentes variables permet de comparer la force de l'association entre l'évènement d'intérêt et ces variables. Puisque le nombre d'observations par prédicteur était élevé ($n=62$), le surajustement aux données (overfitting) n'était pas un problème et des méthodes de pénalisation ou de rétrécissement (shrinkage) des coefficients n'ont pas été nécessaires [44].

Plusieurs mesures ont été utilisées pour évaluer différents aspects des performances du modèle (annexe 1) : le R^2 de Nagelkerke et le score de Brier évaluent les performances globales du modèle ; l'aire sous la courbe (statistique *c*), la sensibilité, la spécificité, la valeur prédictive positive et négative et la pente de discrimination permettent d'évaluer la capacité de discrimination du modèle ; le test de le Cessie et Van Houwelingen évalue l'ajustement du modèle aux données.

Pour la validation interne du modèle final, nous avons utilisé la méthode du bootstrap pour estimer l'optimisme du modèle, qui a ensuite été utilisé pour corriger les performances des modèles (R^2 de Nagelkerke, statistique *c*) [45]. Deux mesures de surajustement aux données ont été estimées : la calibration globale (calibration in the large) et la pente de calibration [46].

Les performances du modèle ont été estimées parmi l'ensemble des personnes ayant reçu un antiparkinsonien en 2007. Dans certains cas, il est intéressant de calculer la spécificité et la valeur prédictive négative pour l'ensemble de la population ; pour cela, nous avons considéré que les affiliés de la MSA des cinq départements qui n'ont reçu aucun antiparkinsonien en 2007 et vérifiaient les critères d'inclusion étaient de vrais négatifs.

Dans certains cas, des informations permettant d'estimer des aspects quantitatifs de l'utilisation des médicaments (comme la dose cumulée annuelle ou le nombre de prescriptions ou de boîtes) ne sont pas disponibles. Afin d'évaluer si l'inclusion d'informations quantitatives sur l'utilisation des médicaments antiparkinsoniens, la proportion de temps traité et le nombre de consultations par an avec un neurologue ou un généraliste permettent d'améliorer les performances du modèle par rapport à un modèle plus simple, nous avons développé un autre modèle prédictif en suivant les mêmes étapes et en définissant l'ensemble de ces prédicteurs comme des variables binaires ; l'âge est généralement disponible et a donc été conservé dans le modèle comme une variable continue et codée à l'aide d'un polynôme fractionnaire.

Dans le cadre d'analyses de sensibilité, nous avons utilisé d'autres variables quantitatives (nombre de prescriptions par an, nombre de boîtes achetées par an, dose moyenne journalière) et évalué l'impact de l'utilisation de courtes périodes de temps pour définir la consommation médicamenteuses (3, 6 ou 9 mois).

Nous avons réalisé des tests bilatéraux qui ont été considérés significatifs au seuil de signification de 0,05.

Ces analyses ont été effectuées à l'aide du package `DiagnosisMed` et de la fonction `val.prob.ci` implémentées dans le logiciel R (version 2.11.0, R Foundation for Statistical Computing, Vienna, Austria, Vienne, Autriche).

4.4 Prévalence de la maladie de Parkinson

Afin d'estimer la prévalence de la maladie de Parkinson au 31 décembre 2007 parmi les affiliés à la MSA des cinq départements, nous avons utilisé le modèle prédictif développé à partir de la dose cumulée des médicaments antiparkinsoniens.

Le modèle permet de prédire la probabilité qu'une personne a d'être traitée pour la maladie de Parkinson, et non pour une autre raison, à partir de certaines de ses caractéristiques et de son profil de consommation d'antiparkinsoniens sur un an. Ce modèle a été appliqué aux affiliés âgés de 18 ans et plus vivants au 31 décembre 2007 qui ont consommé au moins un antiparkinsonien en 2007 (après exclusion des femmes ayant uniquement consommé de faibles doses de bromocriptine et des personnes ayant uniquement consommé des anticholinergiques en association avec des neuroleptiques).

Pour une personne donnée, le logit de la probabilité de la maladie de Parkinson est calculé en additionnant la constante du modèle et les coefficients de régression estimés par le modèle logistique multipliés par les valeurs des variables correspondantes pour cette personne. La probabilité d'être traité pour la maladie de Parkinson pour cette personne est l'inverse de $(1 + e^{-\text{logit}})$. Les personnes qui ont une probabilité prédite égale ou supérieure au seuil de probabilité qui maximise l'indice de Youden sont considérées comme étant traitées pour maladie de Parkinson tandis que les personnes qui ont une probabilité inférieure sont considérées comme étant traitées pour une autre raison ; le seuil qui maximise l'indice de Youden est celui qui maximise le nombre d'individus bien classés [47].

La prévalence a été calculée en divisant le nombre de personnes traitées pour maladie de Parkinson d'après le modèle prédictif par le nombre d'affiliés vivants au 31 décembre 2007 qui a été fourni par chacune des caisses de MSA ; nous disposons de l'âge, du sexe et du canton de résidence des affiliés inclus dans le dénominateur. Nous avons calculé la prévalence de la maladie de Parkinson chez les sujets âgés de 18 ans et plus, dans l'ensemble, selon le sexe et par groupe d'âge de 10 ans.

En faisant l'hypothèse que le nombre de cas de maladie de Parkinson est négligeable avant 18 ans, nous avons pu estimer la prévalence standardisée sur l'âge et le sexe (standardisation directe) pour l'ensemble de la population, à partir de la distribution par âge et sexe de la population française en 2007 [48].

Pour évaluer l'impact des erreurs de classement induites par le modèle prédictif, nous avons calculé un nombre corrigé de cas traités pour la maladie de Parkinson et la prévalence correspondante, grâce aux estimations de la sensibilité et de la spécificité du modèle prédictif [49].

Dans le cadre d'analyses de sensibilité, différents seuils ont été utilisés pour appliquer le modèle prédictif ; pour chacun des seuils, le nombre de personnes identifiées comme étant atteintes de maladie de Parkinson ainsi que le rapport du nombre d'hommes sur le nombre de femmes et l'âge moyen de ces personnes ont été calculés.

4.5 Relation entre la prévalence de la maladie de Parkinson et les caractéristiques agricoles

4.5.1 Définition des variables agricoles

Les recensements agricoles sont des enquêtes statistiques réalisées en 1970, 1979, 1988 et 2000 par le ministère chargé de l'Agriculture et de la Pêche. Leur objectif est de décrire l'état de l'agriculture française et de suivre son évolution, de produire des résultats à petite échelle géographique (cantons, communes) et d'aider aux choix stratégiques et politiques des décideurs [50]. Les recensements agricoles sont des enquêtes exhaustives dont l'unité d'observation est l'exploitation agricole. Les exploitations sont identifiées à partir de fichiers préexistants dans les services statistiques du ministère et de listes fournies par les mairies. De plus, les enquêteurs recrutés pour les recensements agricoles prennent soin de s'assurer de l'exhaustivité des exploitations agricoles en se déplaçant dans les communes. Lors du recensement, le chef d'exploitation est interrogé et de nombreuses informations sont recueillies (population agricole, utilisation des sols, effectifs animaux, moyens de production...). Cependant, aucune information sur l'utilisation des produits phytosanitaires n'est renseignée dans les questionnaires.

Le type d'agriculture a été défini à partir de l'orientation technico-économique (Otex) des exploitations [51]. Il s'agit d'une définition européenne qui permet de classer les exploitations en une Otex parmi 16 (tableau 2) en fonction de l'importance relative des différentes productions de l'exploitation agricole. Cette importance relative est évaluée quantitativement à partir de la part de chaque production dans la marge brute standard (MBS) totale de l'exploitation. Les MBS de chaque production sont calculées à partir des quantités produites et de valeurs de référence qui correspondent à la valeur de la production par hectare

ou tête de bétail moins le coût des facteurs de production. Les Otex sont exclusives et chaque exploitation est classée en une seule orientation. Un exemple de classification d'une exploitation agricole est présenté en annexe 2.

Les Otex sont définies depuis 1985 et sont électroniquement disponibles à partir du recensement agricole de 1988 au niveau cantonal [50]. Pour chaque canton, le nombre d'exploitations agricoles de chaque orientation est renseigné et la densité en chaque Otex (nombre d'exploitations par 100 km²) a été calculée en divisant le nombre d'exploitations d'une Otex par la superficie du canton. Dans les cinq départements participant à l'étude, les cantons avaient une superficie moyenne de 176,20 km², avec, en moyenne, 1 150 affiliés par canton.

| Tableau 2 |

Fréquence des 16 orientations technico-économiques en 1988 dans les cinq départements de l'étude

Orientations technico-économiques des exploitations (Otex)	Ensemble		Charente-Maritime		Côte-d'Or		Gironde		Haute-Vienne		Mayenne	
	n	%	n	%	n	%	n	%	n	%	n	%
Céréales et oléoprotéagineux	2 790	4,0	955	6,1	743	9,0	697	3,8	56	0,5	339	1,9
Cultures générales	5 261	7,5	2 813	17,8	1 614	19,5	310	1,7	55	0,5	469	2,7
Maraîchage	965	1,4	320	2,0	184	2,2	392	2,1	46	0,4	23	0,1
Fleurs et horticulture diverse	455	0,6	124	0,8	56	0,7	180	1,0	46	0,4	49	0,3
Viticulture (appellation et autre)	15 080	21,5	3 071	19,5	1 529	18,5	10 479	57,1	1	0,1	0	0,0
Fruits et autres cultures permanentes	695	1,0	153	1,0	128	1,5	299	1,6	60	0,6	55	0,3
Bovins lait	9 068	12,9	412	2,6	157	1,9	436	2,4	573	5,2	7 490	42,9
Bovins viande	8 192	11,7	239	1,5	870	10,5	492	2,7	3 975	38,1	2 616	15,0
Bovins lait-viande	2 063	2,9	123	0,8	192	2,3	92	0,5	222	2,1	1 434	8,2
Ovins-caprins et autres herbivores	8 346	11,9	595	3,8	588	7,1	1 069	5,8	3 998	38,4	2 096	12,0
Granivores	506	0,7	85	0,5	40	0,5	99	0,5	38	0,4	244	1,4
Polyculture	6 319	9,0	4 027	25,5	349	4,2	1 720	9,4	91	0,9	132	0,8
Polyélevage à orientation herbivores	3 112	4,4	605	3,8	324	3,9	673	3,7	788	7,6	722	4,1
Polyélevage à orientation granivores	751	1,1	23	0,1	27	0,3	39	0,2	56	0,5	606	3,5
Grandes cultures et herbivores	4 088	5,8	1 249	7,9	1 236	15,0	413	2,3	260	2,5	930	5,3
Autres combinaisons cultures-élevage	2 555	3,6	972	6,2	230	2,8	958	5,2	153	1,5	242	1,4
Total	70 246	100	15 766	100	8 267	100	18 348	100	10 418	100	17 447	100

4.5.2 Analyse statistique

La prévalence de la maladie de Parkinson a été estimée pour chaque canton grâce au modèle prédictif et en utilisant le seuil optimal défini par l'index de Youden. La relation entre la prévalence de la maladie de Parkinson et la densité des Otex au niveau cantonal a été étudiée grâce à la régression logistique [52]. Dans un premier temps, nous avons utilisé un modèle logistique mixte avec un effet aléatoire correspondant aux cantons [53]. Ce modèle ne représentait pas une amélioration significative par rapport à un modèle à effets fixes plus simple ($p=0,063$), et ce dernier type de modèle a été utilisé pour la suite des analyses.

Au niveau individuel, les variables d'ajustement étaient l'âge et le sexe ; après comparaison de différents codages, l'âge a été pris en compte grâce à un terme linéaire et quadratique. Les modèles ont également été ajustés sur un indicateur du niveau socio-économique (revenu médian des ménages en 2007) défini au niveau cantonal par l'Insee [54] car des variations de la prévalence de la maladie de Parkinson suivant le niveau socio-économique ont été observées [55]. Enfin, nous avons également ajusté sur le département afin de prendre en compte des différences entre départements non capturées par les variables précédentes.

Chaque Otex a d'abord été considérée comme une variable explicative dans un modèle indépendant. Des variables à cinq catégories ont été définies à partir des quintiles de la distribution de la densité de chacune des Otex parmi les sujets n'ayant pas la maladie de Parkinson. Pour étudier les relations dose-effets, nous avons défini des variables ordinales à partir de la médiane des cinq catégories précédemment définies [56]. Le revenu médian des ménages a également été catégorisé en quintiles ; il n'existait pas de différence de prévalence entre les quatre catégories les plus élevées et nous avons défini une variable binaire afin de comparer les cantons dans le quintile inférieur aux cantons dans les quatre quintiles supérieurs. Les interactions entre les variables ont été recherchées en incluant un terme multiplicatif dans le modèle.

Dans un deuxième temps et afin de prendre en compte la corrélation entre les variables agricoles, nous avons implémenté une approche multivariée à l'aide d'un modèle logistique semi-bayésien [57;58]. La présentation des résultats de ce modèle, qui sont décrits ailleurs [1], dépasse les objectifs de ce rapport.

Les analyses ont été conduites dans l'ensemble puis séparément chez les hommes et les femmes car la prévalence de la maladie de Parkinson est plus élevée chez les hommes et l'exposition professionnelle aux pesticides est considérablement plus fréquente chez les hommes.

Dans le cadre d'analyses de sensibilité, nous avons étudié l'impact sur les résultats de ces analyses de l'utilisation de seuils différents du modèle prédictif pour estimer la prévalence de la maladie de Parkinson.

Nous avons réalisé des tests bilatéraux qui ont été considérés significatifs au seuil de signification de 0,05.

Les analyses ont été réalisées avec le logiciel SAS[®] version 9.2 (SAS Institute, Inc., Cary, North Carolina) et le logiciel Stata[®] version 10 (StataCorp LP, College Station, Texas).

5. Résultats

5.1 Inclusion des participants

Parmi 202 087 personnes vérifiant les critères d'inclusion, 1 540 ont consommé au moins un médicament antiparkinsonien en 2007 (figure 1). Nous avons directement contacté 723 personnes car elles recevaient un traitement souvent utilisé pour la maladie de Parkinson, tandis que 817 personnes ont été dans un premier temps contactées par courrier pour obtenir des renseignements supplémentaires. Parmi ces dernières, 52 personnes ont déclaré utiliser un traitement antiparkinsonien pour un syndrome parkinsonien ou ne savaient pas pourquoi il leur était prescrit et 188 (23 %) n'ont pas répondu au courrier : ces 188 personnes étaient plus jeunes (64 ans, écart-type, ET=15) que les personnes ayant répondu (68 ans, ET=11 ; $p < 0,001$), mais elles utilisaient le même type d'antiparkinsoniens ($p = 0,579$).

Un nombre total de 775 (723 + 52) personnes ont donc été contactées pour être interrogées ; 4 personnes sont décédées avant le début de l'étude et 69 personnes n'ont pas pu être contactées car nous n'avons pas pu obtenir leur numéro de téléphone. Parmi les 702 personnes restantes, 74 ne vérifiaient pas les critères d'inclusion dans l'étude et 119 étaient traitées pour un SJSR ou un syndrome parkinsonien d'origine iatrogène ou ont interrompu leur traitement avant un mois et ont été considérées comme non traitées pour maladie de Parkinson. Au total, 509 personnes ont été invitées à être examinées par le neurologue de l'étude. Quatre vingt-onze (18 %) d'entre elles ont refusé cet examen ; elles étaient plus âgées (75 ans, ET=4) que les personnes ayant accepté (72 ans, ET=6 ; $p < 0,001$) et consommaient des agonistes dopaminergiques de type 1 moins fréquemment (15 % vs 29 %, $p = 0,006$). Parmi les 418 personnes restantes, 320 ont reçu un diagnostic de maladie de Parkinson après examen neurologique, tandis qu'un autre diagnostic a été retenu pour 98 personnes.

Nos analyses reposent donc sur 1 114 participants : 320 traités pour maladie de Parkinson et 794 (98 + 119 + 577) traités par antiparkinsoniens pour une autre raison (figure 1) ; en conséquence, une proportion élevée (71 %) des personnes qui ont reçu au moins une prescription d'antiparkinsonien en 2007 ne sont pas traités pour la maladie de Parkinson.

5.2 Caractéristiques des participants

Les caractéristiques des participants retenus pour les analyses sont présentées dans le tableau 3. Dans des analyses univariées, les patients parkinsoniens étaient plus âgés et plus souvent des hommes que les patients non parkinsoniens. Ils ont également vu plus souvent un neurologue avec un plus grand nombre de consultations par an, tandis qu'une proportion similaire de patients avec et sans maladie de Parkinson ont vu un médecin généraliste au moins une fois par an ; toutefois, les patients parkinsoniens avaient consulté plus fréquemment le généraliste que les autres participants. La proportion de temps traité était plus élevée chez les patients parkinsoniens que chez les autres patients. Les médicaments consommés étaient très différents entre les deux groupes. Les patients atteints de maladie de Parkinson ont été plus souvent traités par lévodopa, amantadine, sélégiline, agonistes dopaminergiques et inhibiteurs de la COMT que les patients non parkinsoniens ; les patients parkinsoniens ont également reçu une dose cumulée plus élevée pour l'ensemble de ces médicaments. A l'inverse, les anticholinergiques et le piribédil étaient plus souvent prescrits aux patients non parkinsoniens, mais la dose cumulée de piribédil était plus élevée chez les patients parkinsoniens. L'annexe 3 présente d'autres mesures quantitatives de l'utilisation des médicaments antiparkinsoniens.

L'annexe 4 présente les performances de chaque antiparkinsonien parmi l'ensemble des personnes ayant reçu des médicaments antiparkinsoniens. La lévodopa a la meilleure combinaison de sensibilité (86,6 %) et de spécificité (82,9 %). Les autres médicaments antiparkinsoniens sont caractérisés par une faible sensibilité (comprise entre 4,4 % et 34,4 %).

I Tableau 3 I

Caractéristiques des participants avec ou sans maladie de Parkinson

Caractéristiques		Avec MP N=320	Sans MP N=794	OR ^a	IC 95 % ^a
Age moyen (ET), années		71,7 (6,7)	68,7 (11,3)	1,4	1,2-1,7
Nombre d'hommes, n (%)		185 (57,8)	353 (44,5)	1,7	1,3-2,2
Consultations avec un neurologue	Au moins une consultation, n (%)	131 (40,9)	36 (4,5)	14,6	9,8-21,8
	Moyenne (ET) du nombre de consultations	2,4 (1,4)	1,8 (1,0)	1,8	1,1-2,9
Consultations avec un généraliste	Au moins une consultation, n (%)	275 (85,9)	714 (89,9)	0,7	0,5-1,0
	Moyenne (ET) du nombre de consultations	5,9 (3,1)	4,8 (3,2)	1,4	1,2-1,6
Proportion du temps traité	Egale à 100 %, n (%)	294 (91,9)	483 (60,8)	7,3	4,8-11,1
	Moyenne (ET)	98,8 (7,4)	80,7 (31,7)	5,2	3,1-8,8
Agonistes dopaminergiques type 1	Au moins un remboursement, n (%)	110 (34,4)	69 (8,7)	5,5	3,9-7,7
	Moyenne (ET) des doses cumulées ^d	81 002,1 (51 696,4)	19 467,3 (36 934,4)	6,9	3,6-13,0
Agonistes dopaminergiques type 2 ^b	Au moins un remboursement, n (%)	15 (4,7)	9 (1,1)	4,3	1,9-9,9
Amantadine ^b	Au moins un remboursement, n (%)	19 (5,9)	8 (1,0)	6,2	2,7-14,3
Anticholinergiques ^c	Au moins un remboursement, n (%)	14 (4,4)	83 (10,5)	0,4	0,2-0,7
Inhibiteurs de la COMT	Au moins un remboursement, n (%)	84 (26,3)	11 (1,4)	25,3	13,3-48,3
Lévodopa	Au moins un remboursement, n (%)	277 (86,6)	136 (17,1)	31,2	21,5-45,1
	Moyenne (ET) des doses cumulées ^d	162 461,8 (115 900,7)	74 796,3 (68 197,3)	3,8	2,7-5,5
Piribédil	Au moins un remboursement, n (%)	80 (25,0)	531 (66,9)	0,2	0,1-0,2
	Moyenne (ET) des doses cumulées ^d	40 683,8 (21 793,0)	14 666,7 (11 844,5)	4,0	3,1-5,2
Sélégiline ^b	Au moins un remboursement, n (%)	40 (12,5)	11 (1,4)	10,2	5,1-20,1

Abréviations : COMT : catéchol-O-méthyltransférase ; ET : écart-type ; EDL : équivalent de dose de lévodopa ; IC : intervalle de confiance ; MP : maladie de Parkinson ; OR : odds ratio.

^aLes OR et les IC sont calculés en utilisant la régression logistique. Pour les variables continues, les OR sont exprimés pour l'augmentation d'un écart-type.

^bLes doses cumulées pendant un an (en EDL) ne sont pas calculées car moins de 55 personnes ont reçu le traitement.

^cAucune EDL n'est disponible pour ces familles de médicaments antiparkinsoniens.

^dLa dose cumulée pendant un an est calculée parmi les personnes ayant reçu ce traitement et exprimée en milligrammes d'EDL.

5.3 Développement du modèle prédictif

Le tableau 4 présente les estimations des paramètres du modèle prédictif multivarié développé à partir des doses cumulées d'antiparkinsoniens. L'âge et l'amantadine n'ont pas été retenus dans le modèle final. Les variables les plus fortement associées avec la maladie de Parkinson sont (par *Z-ratio* décroissant) : lévodopa, agonistes dopaminergiques de type 1, piribédil, la proportion de temps traité, nombre de consultations avec un neurologue, sélégiline.

Les performances du modèle sont présentées dans le tableau 5. La proportion de variance expliquée (R^2) est égale à 71,6 %. La figure 3 présente la courbe ROC (receiver operating characteristics) et la distribution des probabilités prédites dans chaque groupe. Le modèle est caractérisé par de bonnes performances en termes de discrimination (statistique $c=0,953$; pente de discrimination= $0,625$) et représente une amélioration considérable par rapport aux modèles qui comprenaient un seul antiparkinsonien (annexe 4). Pour le seuil optimal, la sensibilité est de 92,5 %, tandis que sa spécificité est un peu inférieure (86,4 %). L'annexe 5 présente les performances en termes de discrimination pour d'autres seuils de probabilité ; lorsque le seuil augmente, la spécificité augmente tandis que la sensibilité diminue. La courbe de calibration a montré une bonne corrélation entre les probabilités observées et prédites (figure 4) ; d'après le test de la Cessie et Van Houwelingen, le modèle est caractérisé par un bon ajustement aux données ($p=0,794$).

Afin de calculer la spécificité et la valeur prédictive négative parmi l'ensemble des affiliés, nous avons considéré l'ensemble des sujets qui n'ont pas utilisé d'antiparkinsonien en 2007 et qui vérifiaient les critères d'inclusion comme de vrais négatifs ($n=200\ 547$) ; en utilisant le même seuil que précédemment, la spécificité est égale à 99,95 % et la valeur prédictive négative à 99,99 %.

Nous avons vérifié si des personnes présentant certains diagnostics ont été plus fréquemment identifiées à tort comme ayant une maladie de Parkinson (tableau 6). Parmi les vrais négatifs, les troubles de la mémoire, les pathologies ORL (vertiges, acouphènes, hypoacousie), le tremblement essentiel, le SJSR et les syndromes parkinsoniens iatrogènes représentaient plus de 80 % des diagnostics, tandis que chez les faux positifs, le tremblement essentiel et les syndromes parkinsoniens (à l'exception de ceux d'origine iatrogène) représentaient plus de 70 % des diagnostics. La plupart des patients avec un syndrome parkinsonien (à l'exception de ceux d'origine iatrogène), environ 25 % des patients atteints de tremblement essentiel et 10 % des patients atteints de SJSR ou de syndrome parkinsonien iatrogène ont été classés comme de faux positifs. La majorité des patients atteints de troubles de la mémoire, de pathologies ORL, d'une pathologie vasculaire, ou d'autres pathologies ont été classés correctement.

Nous avons également comparé, parmi les personnes atteintes de maladie de Parkinson ($n=320$), les caractéristiques de celles identifiées par le modèle prédictif (vrais positifs, $n=296$) et de celles qui n'ont pas été non identifiées comme étant atteintes (faux négatifs, $n=24$). Par rapport aux vrais positifs, les faux négatifs étaient moins fréquemment traités par lévodopa (25 % contre 92 %, $p<0,001$), inhibiteurs de la COMT (0 % contre 28 %, $p=0,001$) et agonistes dopaminergiques de type 1 (13 % contre 36 %, $p=0,024$). En revanche, les faux négatifs étaient plus souvent traités par trivastal (58 % contre 22 %, $p<0,001$). De plus, les doses cumulées d'antiparkinsoniens reçues par les faux négatifs étaient plus faibles pour tous les médicaments (p compris entre $<0,001$ et $0,012$). Seulement 8 % des faux négatifs avaient vu un neurologue pendant l'année contre 44 % parmi les vrais positifs ($p<0,001$). Enfin, bien que les deux groupes avaient un âge moyen comparable ($p=0,191$), les faux négatifs avaient une durée d'évolution plus courte (médiane=4 ans) que les vrais positifs (médiane=6 ans) ($p=0,027$) et donc une maladie moins évoluée.

L'optimisme du R^2 de Nagelkerke et de la statistique c estimé après correction par la méthode du bootstrap était très faible (<3 % ; tableau 5). Les performances du modèle étaient peu modifiées après correction pour l'optimisme. En moyenne, les valeurs prédites et observées après bootstrap étaient très proches (calibration-in-the-large= $-3,8$ %). Le surajustement était limité (pente de calibration= $92,2$ %).

I Tableau 4 I

Modèle prédictif de la maladie de Parkinson développé à partir des doses cumulées de médicaments antiparkinsoniens

Caractéristiques	Codage	Coefficients (ES)	Z-ratio ^a	OR ^b	IC 95 % ^b	p ^b
Constante	--	-5,332 (1,03)	-5,2	--	--	--
Âge en années	PF : x	NR	--	--	--	--
Sexe	Homme vs femme	0,293 (0,22)	1,3	1,3	0,9-2,1	0,179
Consultations avec un neurologue	1 ou 2 vs aucune	1,396 (0,35)	4,0	4,0	2,0-8,0	<0,001
	>2 vs aucune	0,171 (0,50)	0,3	1,2	0,4-3,2	0,340
Consultations avec un généraliste	1 ou 2 vs aucune	-0,502 (0,48)	-1,0	0,6	0,2-1,6	0,297
	3-6 vs aucune	-0,759 (0,44)	-1,8	0,5	0,2-1,1	0,081
	>6 vs aucune	-1,245 (0,46)	-2,7	0,3	0,1-0,7	0,007
Proportion du temps traité	PF : x	1,182 (0,29)	4,1	3,3	1,9-5,7	<0,001
Agonistes dopaminergiques type 1	PF ^c : x + x ²	1,879 (0,39)	4,8	6,5	3,0-14,1	<0,001
		-0,303 (0,08)	-3,9	0,7	0,6-0,9	<0,001
Agonistes dopaminergiques type 2	Oui vs non	1,048 (0,76)	1,4	2,9	0,6-13,0	0,166
Amantadine	Oui vs non	NR	--	--	--	--
Anticholinergiques	Oui vs non	-0,897 (0,49)	-1,9	0,4	0,2-1,0	0,064
Inhibiteurs de la COMT	Oui vs non	0,752 (0,41)	1,9	2,1	1,0-4,7	0,064
Lévodopa	PF ^c : log(x)	0,448 (0,04)	10,2	1,6	1,4-1,7	<0,001
Piribédil	PF ^c : x + log(x)	1,220 (0,26)	4,7	3,4	2,0-5,6	<0,001
		-0,643 (0,20)	-3,2	0,5	0,4-0,8	0,001
Sélégiline	Oui vs non	2,104 (0,53)	4,0	8,2	2,9-23,1	<0,001

Abréviations : COMT : catéchol-O-méthyltransférase ; ES : erreur standard ; IC : intervalle de confiance ; NR : non retenu dans le modèle ; OR : odds ratio ; PF : polynômes fractionnaires ; x : terme du polynôme fractionnaire (pour les médicaments antiparkinsoniens, il représente la dose cumulée).

Le logit de la probabilité qu'une personne soit traitée pour la maladie Parkinson est calculé en sommant la constante du modèle et les estimations des coefficients du modèle multipliées par la valeur de chaque variable pour cette personne. La probabilité est ensuite calculée en prenant l'inverse de $(1 + e^{-\text{logit}})$.

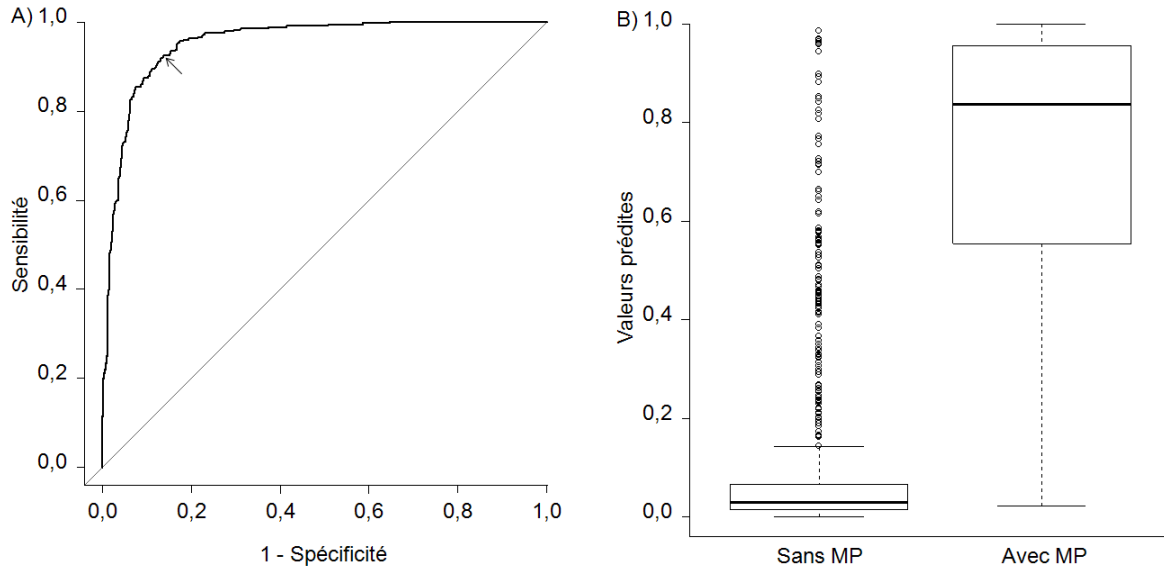
^aLes Z-ratios sont calculés en divisant les coefficients de régression par leur erreur standard.

^bLes OR, les IC et les tests sont calculés en utilisant la régression logistique. Pour les variables continues, les OR sont exprimés pour l'augmentation d'un écart-type (proportion du temps traité, ET=28,2 ; lévodopa, ET=110 563,2 ; piribédil, ET=16 141,6 ; agonistes dopaminergiques type 1, ET=55 314,4). Une procédure de sélection descendante, avec une valeur de p égale à 0,20 pour qu'une variable reste dans le modèle, a été utilisée.

^cLes médicaments antiparkinsoniens codés avec des polynômes fractionnaires sont exprimés en milligrammes d'équivalent de dose de lévodopa.

I Figure 3 I

Courbe ROC (Receiver Operating Characteristics) (A) et boîtes à moustaches des probabilités prédites (B) par le modèle prédictif développé à partir des doses cumulées de médicaments antiparkinsoniens



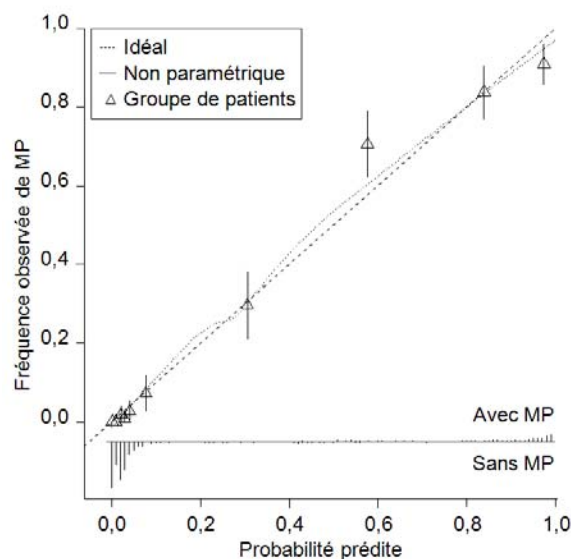
Abréviation : MP : maladie de Parkinson.

Dans la partie A, la statistique c (aire sous la courbe) est égale à 0,953. La sensibilité (92,5 %), la spécificité (86,4 %), la valeur prédictive positive (73,3 %) et la valeur prédictive négative (96,6 %) sont calculées pour le seuil optimal ($p=0,255$) qui maximise le nombre d'individus correctement classés parmi les personnes ayant reçu au moins un médicament antiparkinsonien en 2007 (indiqué par une flèche).

Dans la partie B, la pente de discrimination (différence absolue entre les moyennes des probabilités prédites parmi les personnes avec et sans maladie de Parkinson) est égale à 0,625.

I Figure 4 I

Courbe de calibration du modèle prédictif de maladie de Parkinson développé à partir des doses cumulées de médicaments antiparkinsoniens



Abréviation : MP : maladie de Parkinson.

I Tableau 5 I

Performances du modèle prédictif développé à partir des doses cumulées de médicaments antiparkinsoniens

Performances	Modèle doses cumulées
Performances globales	
R ² de Nagelkerke	71,4 %
Score de Brier standardisé	62,6 %
Discrimination	
Statistique c (IC 95 %)	0,953 (0,941-0,965)
Seuil optimal	0,255
Sensibilité (IC 95 %) ^a	92,5 % (89,1-94,9)
Spécificité (IC 95 %) ^a	86,4 % (83,8-88,6)
Valeur prédictive positive (IC 95 %) ^a	73,3 % (46,0-94,9)
Valeur prédictive négative (IC 95 %) ^a	96,6 % (95,0-97,7)
Pente de discrimination	0,625
Calibration	
Test de Le Cessie-Van Houwelingen	p=0,794
Validation interne^b	
Optimisme du R ² de Nagelkerke	2,7 %
R ² de Nagelkerke corrigé	68,7 %
Optimisme de la statistique c	1,3 %
Statistique c corrigée	0,947
Calibration globale ^c	-3,8 %
Pente de calibration	92,2 %

Abréviation : IC : intervalle de confiance.

^aCalculée pour le seuil optimal qui maximise le nombre d'individus correctement classés parmi les personnes ayant reçu au moins un médicament antiparkinsonien en 2007.

^bValidation interne réalisée par la méthode du bootstrap incluant l'étape de sélection.

^cAppelé calibration-in-the-large dans la littérature anglophone.

Diagnostiques des vrais négatifs et des faux positifs définis par le modèle prédictif développé à partir des doses cumulées de médicaments antiparkinsoniens

Diagnostics	Nombre de participants sans MP	Vrais négatifs ^a n (%)	Faux positifs ^a n (%)
	N=794	N=686	N=108
Syndrome parkinsonien neurodégénératif ^b	18	1 (5,6)	17 (94,4)
Syndrome parkinsonien avec démence ^c	20	2 (10,0)	18 (90,0)
Syndrome parkinsonien vasculaire	17	5 (29,4)	12 (70,6)
Syndrome parkinsonien iatrogène (neuroleptiques)	78	72 (92,3)	6 (7,7)
Autre syndrome parkinsonien	11	1 (9,1)	10 (90,9)
Tremblement essentiel	103	76 (73,8)	27 (26,2)
Syndrome des jambes sans repos	95	87 (91,6)	8 (8,4)
Troubles de la mémoire	181	181 (100)	--
Vertiges/acouphènes/hypoacousie	147	147 (100)	--
Maladies vasculaires	37	37 (100)	--
Autres pathologies	74	69 (93,2)	5 (6,8)
Examen normal	9	6 (66,7)	3 (33,3)
Inconnu	4	3 (75,0)	1 (25,0)

Abréviation : MP : maladie de Parkinson.

^aDéfinis à partir du seuil optimal maximisant le nombre d'individus correctement classés parmi les personnes ayant reçu au moins un médicament antiparkinsonien en 2007.

^bParalysie supranucléaire progressive, atrophie multisystématisée, atrophie cortico-basale.

^cSyndrome parkinsonien secondaire à un syndrome démentiel, maladie des corps de Lewy.

5.4 Analyses de sensibilité

Dans le modèle construit à partir de variables binaires (annexe 6), avoir vu au moins une fois un généraliste, l'amantadine, les agonistes dopaminergiques de type 2 et le piribédil n'ont pas été retenus dans le modèle final. Les caractéristiques associées à la maladie de Parkinson sont (par *Z-ratio* décroissant) : la lévodopa, la proportion de temps traité, avoir consulté un neurologue au moins une fois, les agonistes dopaminergiques de type 1, la sélégiline et les inhibiteurs de la COMT. Les performances de ce modèle (annexe 7) étaient légèrement inférieures à celle du modèle construit à partir des doses cumulées mais demeuraient bonnes. L'optimisme et le surajustement étaient limités.

Lorsque d'autres variables quantitatives ont été utilisées pour construire le modèle prédictif (annexe 7), les modèles construits à partir du nombre de boîtes d'antiparkinsoniens achetées par an et les doses moyennes journalières avaient des performances qui étaient très similaires à celles du modèle construit à partir des doses cumulées, tandis que le modèle construit à partir du nombre de prescriptions par an avait de moins bonnes performances en termes de discrimination.

L'utilisation de périodes plus courtes pour évaluer la consommation de médicaments était associée à de moins bonnes performances (annexe 8).

5.5 Prévalence de la maladie de Parkinson

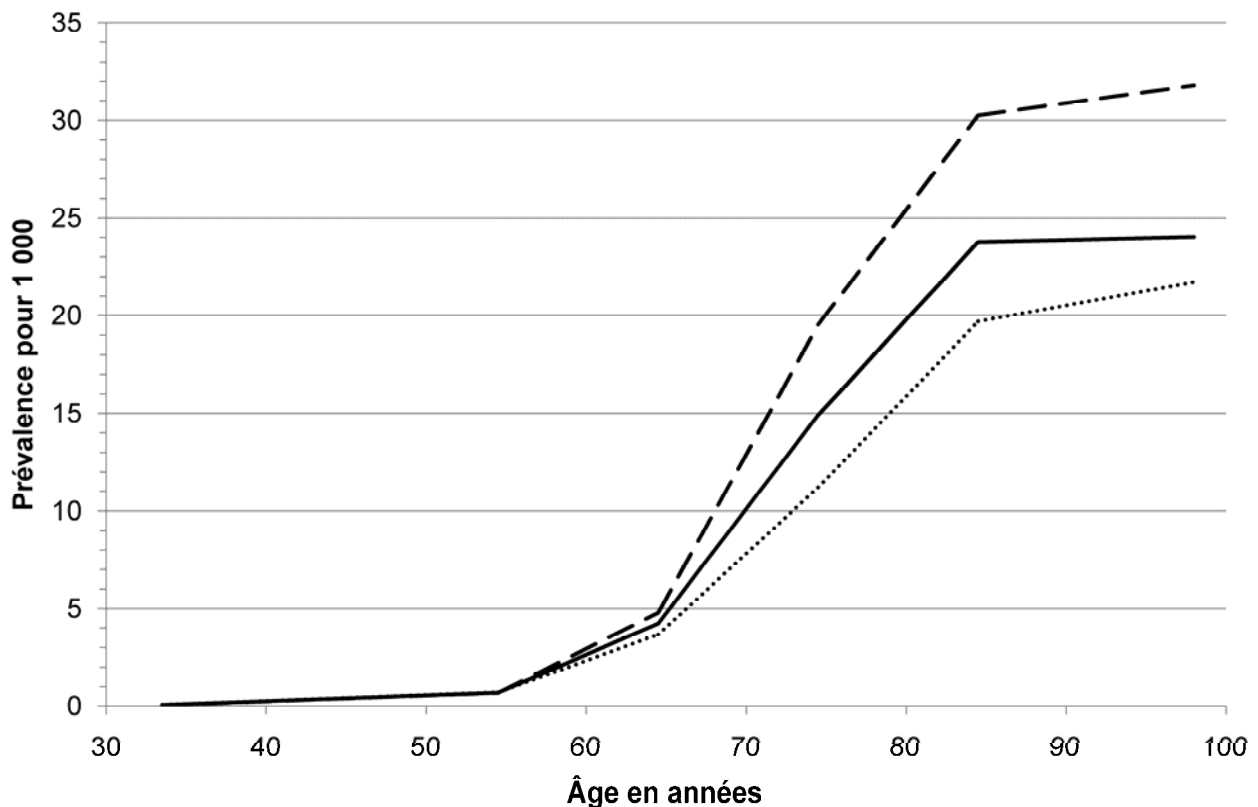
Parmi 239 123 affiliés âgés de 18 ans ou plus et vivants au 31 décembre 2007, nous avons identifié 3 337 affiliés qui ont acheté au moins un médicament antiparkinsonien en 2007 ; d'après le modèle

prédictif développé à partir des doses cumulées et le seuil optimal, 1 593 d'entre eux étaient traités pour la maladie de Parkinson.

La prévalence brute de la maladie de Parkinson prédite par le modèle après 18 ans était de 6,66/1 000 ; en tenant compte des estimations de sensibilité et spécificité du modèle prédictif, nous avons calculé une prévalence corrigée qui était légèrement inférieure (6,04/1 000). La prévalence augmentait avec l'âge et était plus élevée chez les hommes que les femmes à tous les âges (figure 5).

I Figure 5 I

Prévalence de la maladie de Parkinson au 31 décembre 2007 estimée par le modèle prédictif développé à partir des doses cumulées de médicaments antiparkinsoniens parmi les affiliés de la Mutualité sociale agricole âgés de plus de 18 ans dans les cinq départements



La ligne pleine correspond à la prévalence totale, tandis que la ligne en pointillés courts indique la prévalence chez les femmes et la ligne en pointillés longs correspond à la prévalence chez les hommes.

En faisant l'hypothèse que la prévalence de la maladie de Parkinson est négligeable avant 18 ans, la prévalence standardisée sur l'âge et le sexe est de 2,93/1 000 ; la diminution marquée de la prévalence après standardisation est liée au fait que la population de la MSA est plus âgée que la population générale française (annexe 9).

Lors de l'utilisation de seuils différents pour appliquer le modèle prédictif, le nombre de personnes identifiées comme ayant la maladie de Parkinson diminuait de 3 337 pour le seuil le plus faible (le moins spécifique) à 149 pour le seuil le plus élevé (le plus spécifique) (annexes 10 et 11). Les caractéristiques des patients identifiés comme étant parkinsoniens changent avec le seuil : les personnes identifiées sont plus souvent des femmes pour des seuils inférieurs à 0,115 tandis que la proportion d'hommes augmente pour des seuils plus élevés (annexe 10). L'âge moyen est de 77,2 ans pour le seuil le plus faible ; il augmente jusqu'à 79,1 ans pour un seuil égal à 0,375, puis diminue régulièrement jusqu'à 73,4 ans pour le seuil le plus élevé (annexe 11).

5.6 Caractéristiques agricoles et maladie de Parkinson

Dans les cinq départements de l'étude, 70 246 exploitations agricoles ont été recensées en 1988 (tableau 2). Les orientations les plus fréquentes étaient la viticulture (21,5 %), puis l'élevage de bovins-lait (12,9 %) et d'ovins, caprins et autres herbivores (11,9 %).

Les annexes 12 à 16 représentent, dans chaque département, la densité par canton des 16 Otex en 1988. Il existait une importante variabilité pour la plupart des Otex entre les départements et entre les cantons (tableau 2).

Les résultats des analyses portant sur la relation entre la prévalence de la maladie de Parkinson et la densité des Otex sont présentés dans le tableau 7.

Après prise en compte du sexe, de l'âge, du département et du revenu médian cantonal, la prévalence de la maladie de Parkinson augmentait avec la densité en exploitations spécialisées dans les fruits et autres cultures permanentes ($p_{\text{tendance}}=0,005$) ; la prévalence était environ 1,3 fois plus élevée dans les cantons ayant plus de 4 exploitations pour 100 km² (C5) par rapport aux cantons n'ayant pas ce type d'exploitations (C1, groupe de référence ; OR=1,27, IC 95 %=1,07 - 1,51). Bien que la densité de cette orientation varie entre les cinq départements, la relation entre la prévalence de la maladie de Parkinson et la densité en exploitations spécialisées dans les fruits et autres cultures permanentes était similaire entre les départements ($p_{\text{interaction}}=0,848$).

La prévalence de la maladie de Parkinson était également plus élevée pour certaines catégories des autres types d'orientations (viticulture et élevage bovin) mais sans aucune tendance.

Parmi les hommes (annexe 17), la prévalence de la maladie de Parkinson augmentait également avec la densité en exploitations spécialisées dans les fruits et autres cultures permanentes (OR_{C5 vs C1}=1,40, IC 95 %=1,10 - 1,79 ; $p_{\text{tendance}}=0,009$). Cette relation était moins nette parmi les femmes (annexe 18), mais l'interaction entre l'exposition et le sexe n'était pas significative ($p_{\text{interaction}}=0,222$). Aucune tendance avec les autres orientations n'est observée dans les analyses conduites séparément chez les hommes et les femmes.

La relation entre la densité des exploitations spécialisées dans les fruits et autres cultures permanentes et la prévalence de la maladie de Parkinson persistait lorsqu'on faisait varier le seuil du modèle prédictif permettant de définir les cas de maladie de Parkinson (tableau 8). Toutefois, l'association était moins forte pour un seuil faible et associé à une spécificité du modèle plus faible (seuil=0,005 ; OR_{C5 vs C1}=1,19, IC 95 %=1,05 - 1,35) tandis que la relation augmentait pour un seuil plus élevé associé à un modèle moins sensible mais plus spécifique (seuil=0,705 ; OR_{C5 vs C1}=1,47, IC 95 %=1,10 - 1,96).

I Tableau 7 I

Relation entre la prévalence de la maladie de Parkinson en 2007 et la densité des orientations technico-économiques dans les cinq départements de l'étude (Charente-Maritime, Côte-d'Or, Gironde, Haute-Vienne, Mayenne)

Orientations technico-économiques des exploitations (Otex)	C	Densité (/100 km ²)	Cas parkinsoniens N (%)	Sujets non parkinsoniens N (%)	OR (IC 95 %) ^a	p ^b
Céréales et oléoprotéagineux	1	0-1	338 (21,22)	46 539 (19,59)	Référence	
	2	2-4	303 (19,02)	47 984 (20,20)	1,01 (0,85-1,20)	
	3	4-8	275 (17,26)	47 913 (20,17)	1,08 (0,89-1,31)	
	4	8-14	326 (20,46)	46 566 (19,60)	1,11 (0,92-1,34)	
	5	14-81	351 (22,03)	48 525 (20,43)	1,05 (0,86-1,27)	0,779
Cultures générales	1	0-0	235 (14,75)	47 356 (19,94)	Référence	
	2	0-2	340 (21,34)	47 053 (19,81)	1,00 (0,84-1,19)	
	3	2-8	313 (19,65)	47 875 (20,16)	0,87 (0,72-1,05)	
	4	9-26	316 (19,84)	46 813 (19,71)	0,93 (0,75-1,13)	
	5	26-106	389 (24,42)	48 430 (20,39)	1,03 (0,83-1,30)	0,288
Maraîchage	1	0-0	515 (32,33)	68 217 (28,72)	Référence	
	2	0-1	295 (18,52)	42 057 (17,71)	0,97 (0,83-1,12)	
	3	1-2	298 (18,71)	42 146 (17,74)	1,13 (0,96-1,34)	
	4	2-6	259 (16,26)	42 181 (17,76)	1,13 (0,94-1,36)	
	5	6-349	226 (14,19)	42 926 (18,07)	0,99 (0,84-1,18)	0,669
Fleurs et horticulture diverse	1	0-0	355 (22,28)	46 252 (19,47)	Référence	
	2	0-1	349 (21,91)	47 766 (20,11)	1,05 (0,90-1,22)	
	3	1-2	349 (21,91)	48 218 (20,30)	1,02 (0,87-1,20)	
	4	2-3	267 (16,76)	47 089 (19,82)	1,03 (0,87-1,23)	
	5	3-47	273 (17,14)	48 202 (20,29)	1,08 (0,91-1,27)	0,431
Viticulture (appellation et autre)	1	0-0	719 (45,13)	86 958 (36,61)	Référence	
	2	0-24	211 (13,25)	36 142 (15,22)	1,00 (0,78-1,29)	
	3	24-76	267 (16,76)	37 926 (15,97)	1,30 (1,00-1,70)	
	4	78-203	233 (14,63)	36 773 (15,48)	1,17 (0,92-1,49)	
	5	213-565	163 (10,23)	39 728 (16,73)	0,99 (0,76-1,29)	0,533
Fruits et autres cultures permanentes	1	0-0	291 (18,27)	47 111 (19,83)	Référence	
	2	0-1	308 (19,33)	47 583 (20,03)	1,12 (0,95-1,32)	
	3	1-2	363 (22,79)	47 218 (19,88)	1,02 (0,87-1,19)	
	4	2-4	331 (20,78)	47 431 (19,97)	1,20 (1,02-1,41)	
	5	4-31	300 (18,83)	48 184 (20,29)	1,27 (1,07-1,51)	0,005

I Tableau 7 (suite) I

Orientations technico-économiques des exploitations (Otex)	C	Densité (/100 km ²)	Cas parkinsoniens N (%)	Sujets non parkinsoniens N (%)	OR (IC 95 %) ^a	p ^b
Bovins lait	1	0-1	225 (14,12)	46 698 (19,66)	Référence	
	2	1-3	326 (20,46)	45 853 (19,30)	1,32 (1,11-1,57)	
	3	3-6	271 (17,01)	49 406 (20,80)	1,06 (0,88-1,27)	
	4	6-19	364 (22,85)	47 000 (19,79)	1,19 (1,00-1,43)	
	5	19-331	407 (25,55)	48 570 (20,45)	1,02 (0,79-1,32)	0,419
Bovins viande	1	0-1	310 (19,46)	47 373 (19,94)	Référence,	
	2	1-4	265 (16,64)	47 503 (20,00)	0,92 (0,78-1,09)	
	3	4-7	249 (15,63)	47 649 (20,06)	0,86 (0,72-1,02)	
	4	8-39	294 (18,46)	44 531 (18,75)	0,91 (0,76-1,09)	
	5	40-129	475 (29,82)	50 471 (21,25)	1,00 (0,79-1,27)	0,581
Bovins lait-viande	1	0-0	378 (23,73)	69 507 (29,26)	Référence	
	2	0-1	238 (14,94)	41 628 (17,53)	1,10 (0,93-1,31)	
	3	1-2	276 (17,33)	42 038 (17,70)	1,01 (0,86-1,18)	
	4	2-9	333 (20,90)	41 204 (17,35)	1,08 (0,91-1,28)	
	5	9-58	368 (23,10)	43 150 (18,17)	0,79 (0,59-1,07)	0,151
Ovins-caprins et autres herbivores	1	0-4	247 (15,51)	46 926 (19,76)	Référence	
	2	4-8	316 (19,84)	47 524 (20,01)	1,09 (0,91-1,30)	
	3	8-14	243 (15,25)	46 852 (19,72)	0,98 (0,81-1,19)	
	4	14-34	332 (20,84)	46 707 (19,66)	1,06 (0,88-1,28)	
	5	34-210	455 (28,56)	49 518 (20,85)	1,08 (0,85-1,38)	0,696
Granivores	1	0-0	371 (23,29)	65 449 (27,55)	Référence	
	2	0-1	261 (16,38)	43 013 (18,11)	0,96 (0,82-1,13)	
	3	1-1	271 (17,01)	42 955 (18,08)	0,98 (0,83-1,15)	
	4	1-2	327 (20,53)	43 000 (18,10)	1,10 (0,94-1,29)	
	5	2-12	363 (22,79)	43 110 (18,15)	1,11 (0,93-1,32)	0,158
Polyculture	1	0-1	346 (21,72)	47 342 (19,93)	Référence	
	2	1-5	365 (22,91)	46 951 (19,77)	1,07 (0,92-1,24)	
	3	5-18	257 (16,13)	47 256 (19,90)	1,19 (0,98-1,43)	
	4	19-39	277 (17,39)	47 962 (20,19)	1,08 (0,87-1,35)	
	5	42-133	348 (21,85)	48 016 (20,21)	1,16 (0,92-1,47)	0,421

I Tableau 7 (suite) I

Orientations technico-économiques des exploitations (Otex)	C	Densité (/100 km ²)	Cas parkinsoniens N (%)	Sujets non parkinsoniens N (%)	OR (IC 95 %) ^a	p ^b
Polyélevage à orientation herbivores	1	0-3	244 (15,32)	47 084 (19,82)	Référence	
	2	3-5	276 (17,33)	47 539 (20,01)	1,13 (0,94-1,35)	
	3	5-9	299 (18,77)	47 010 (19,79)	0,95 (0,79-1,14)	
	4	10-15	369 (23,16)	47 558 (20,02)	1,04 (0,87-1,25)	
	5	15-61	405 (25,42)	48 336 (20,35)	1,03 (0,85-1,23)	0,932
Polyélevage à orientation granivores	1	0-0	596 (37,41)	101 335 (42,66)	Référence	
	2	0-1	225 (14,12)	33 694 (14,19)	1,02 (0,87-1,19)	
	3	1-1	207 (12,99)	33 981 (14,31)	1,01 (0,86-1,19)	
	4	1-4	250 (15,69)	34 434 (14,50)	0,99 (0,84-1,17)	
	5	4-27	315 (19,77)	34 083 (14,35)	1,16 (0,82-1,63)	0,376
Grandes cultures et herbivores	1	0-2	199 (12,49)	47 045 (19,81)	Référence	
	2	2-4	288 (18,08)	47 682 (20,07)	1,08 (0,89-1,30)	
	3	4-8	379 (23,79)	47 827 (20,14)	1,04 (0,86-1,27)	
	4	8-16	357 (22,41)	47 261 (19,90)	1,06 (0,87-1,29)	
	5	16-53	370 (23,23)	47 712 (20,09)	1,04 (0,85-1,28)	0,950
Autres combinaisons cultures-élevage	1	0-3	356 (22,35)	46 668 (19,65)	Référence	
	2	3-5	339 (21,28)	46 994 (19,78)	0,95 (0,81-1,12)	
	3	5-10	297 (18,64)	48 481 (20,41)	1,18 (0,99-1,41)	
	4	10-16	304 (19,08)	46 981 (19,78)	1,05 (0,86-1,27)	
	5	16-44	297 (18,64)	48 403 (20,38)	1,07 (0,88-1,30)	0,728

Abréviations : IC : intervalle de confiance ; OR : odds ratio ; C : catégorie.

Les catégories d'exposition sont définies à partir des quintiles de la distribution des densités sauf pour les Otex suivantes: maraîchage, viticulture, bovin lait-viande, granivore, polyélevage à orientation granivore. Pour ces dernières, la catégorie de référence correspond aux cantons sans ce type d'exploitation et les catégories suivantes sont définies à partir des quartiles de la distribution des densités.

^aOR et IC sont ajustés sur le sexe, l'âge (terme linéaire et quadratique), le département et le revenu médian des ménages du canton.

^bTest de tendance des OR.

| Tableau 8 |

Relation entre la prévalence de la maladie de Parkinson en 2007 et la densité des exploitations spécialisées en fruits et autres cultures permanentes en fonction du seuil d'application du modèle prédictif développé à partir des doses cumulées de médicaments antiparkinsoniens

	Seuil du modèle prédictif utilisé pour classer les consommateurs de médicaments antiparkinsoniens en atteint ou non atteint						
	0,005	0,055	0,105	0,255 ^a	0,405	0,555	0,705
Cas identifiés (N)^b	2 945	2 083	1 834	1 593	1 192	873	548
Sensibilité (%)	100,0	97,8	96,3	92,5	87,8	83,1	60,9
Spécificité (%)	16,6	70,9	80,6	86,4	89,3	93,5	96,5
OR (IC 95 %)^c							
Quintile 1	Référence	Référence	Référence	Référence	Référence	Référence	Référence
Quintile 2	1,01 (0,89-1,14)	1,03 (0,89-1,19)	1,05 (0,90-1,22)	1,12 (0,95-1,32)	1,07 (0,89-1,30)	1,00 (0,80-1,24)	1,04 (0,79-1,39)
Quintile 3	0,98 (0,88-1,11)	1,01 (0,88-1,16)	1,03 (0,89-1,19)	1,02 (0,87-1,19)	1,00 (0,83-1,21)	0,92 (0,74-1,15)	0,95 (0,72-1,25)
Quintile 4	1,11 (0,98-1,25)	1,16 (1,00-1,33)	1,19 (1,02-1,38)	1,20 (1,02-1,41)	1,23 (1,02-1,48)	1,19 (0,96-1,47)	1,34 (1,02-1,76)
Quintile 5	1,19 (1,05-1,35)	1,23 (1,06-1,43)	1,23 (1,05-1,44)	1,27 (1,07-1,51)	1,39 (1,14-1,70)	1,34 (1,07-1,69)	1,47 (1,10-1,96)
p^d	0,001	0,001	0,004	0,005	0,001	0,001	0,001

Abréviations : IC : intervalle de confiance ; OR : odds ratio.

^aSeuil optimal maximisant le nombre d'individus correctement classés parmi les personnes ayant reçu au moins un médicament antiparkinsonien en 2007.

^bNombre de sujets classés comme parkinsoniens parmi 3 337 sujets ayant consommé un médicament antiparkinsonien en 2007.

^cOR ajustés sur le sexe, l'âge (terme linéaire et quadratique), le département et le revenu médian des ménages du canton ; ils comparent la prévalence de la maladie de Parkinson en fonction des quintiles de densité des cantons en exploitations spécialisées en fruits et autres cultures permanentes.

^dTest de tendance des OR.

6. Discussion

6.1 Résumé des principaux résultats

Cette étude, conduite en 2007 parmi les affiliés à la MSA de cinq départements français, montre que les bases de remboursement de médicaments représentent un outil intéressant pour estimer la probabilité qu'une personne a d'être traitée pour la maladie de Parkinson. Un modèle prédictif a été développé à partir des profils de consommation de médicaments pouvant être utilisés pour traiter la maladie de Parkinson ; il est caractérisé par des performances satisfaisantes.

Le modèle prédictif développé à partir d'informations quantitatives sur la consommation de médicaments (doses cumulées reçues pendant un an) et sur la régularité des prescriptions présente de meilleures performances en termes de prédiction qu'un modèle développé à partir de variables binaires. De plus, les performances du modèle prédictif sont d'autant meilleures que la durée d'évaluation de la consommation médicamenteuse est longue.

Ce modèle prédictif a été utilisé pour estimer la prévalence de la maladie de Parkinson après 18 ans (6,66/1 000) ; la prévalence augmentait avec l'âge et était plus élevée chez les hommes que chez les femmes à tous les âges. La correction de cette estimation grâce à la sensibilité et la spécificité du modèle a conduit à une estimation proche (6,04/1 000). En faisant l'hypothèse raisonnable que la fréquence de la maladie de Parkinson est négligeable avant 18 ans, nous avons obtenu une estimation de la prévalence de la maladie de Parkinson standardisée sur l'âge et le sexe (avec la population française de 2007 comme référence) de 2,93/1 000 ; cette estimation est un peu plus élevée que la prévalence habituelle de la maladie de Parkinson [27;59].

Comme attendu, les profils de consommation médicamenteuse des patients parkinsoniens étaient très différents de ceux des personnes ayant consommé des médicaments antiparkinsoniens pour une autre raison. Il existait des différences importantes quant aux médicaments consommés par les deux groupes, tant en termes de fréquence de consommation que de dose. Le modèle prédictif incluant plusieurs médicaments antiparkinsoniens présentait de meilleures performances en termes de prédiction que les modèles incluant un seul médicament antiparkinsonien à la fois, en particulier la lévodopa. La prise en compte d'informations sur la dose consommée permettait d'améliorer les performances du modèle, en particulier la calibration. La régularité du traitement, estimée par la proportion de temps traité, était également une variable importante : une proportion de temps traité plus élevée est un indicateur d'une utilisation plus régulière et les utilisateurs réguliers avaient une probabilité plus élevée d'avoir la maladie de Parkinson que les personnes utilisant des médicaments antiparkinsoniens pendant de courtes durées.

Peu d'études ont estimé la prévalence de la maladie de Parkinson à partir des profils de consommation médicamenteuse. Certaines études reposent sur un seul médicament antiparkinsonien (lévodopa) [20], tandis que d'autres considèrent plusieurs médicaments antiparkinsoniens [19;21;23;25;60], ou une combinaison de médicaments antiparkinsoniens et de codes de la classification internationale des maladies [22;24]. Aucune de ces études n'a inclus d'information sur la dose ou à la régularité de la consommation médicamenteuse. De plus, le diagnostic n'a pas été vérifié dans la plupart des études ; lorsqu'il a été vérifié, différentes méthodes ont été utilisées : examen neurologique dans deux études [20;25] ; dossiers médicaux [23;24] ; interrogatoire des participants [22].

Un seul modèle prédictif de maladie de Parkinson a été développé antérieurement dans le cadre de l'étude de cohorte de Rotterdam à partir d'un petit nombre d'utilisateurs de médicaments antiparkinsoniens (n=63) [61]. Les performances en termes de discrimination ont été calculées parmi l'ensemble des participants avec une statistique c égale à 0,93. Aucune information quantitative sur la consommation médicamenteuse n'a été prise en compte dans le modèle et la calibration et la validité interne n'ont pas été évaluées.

6.2 Erreurs de classement

Parmi les faux positifs prédits par le modèle, certains diagnostics sont plus fréquents. Les patients atteints de syndrome parkinsonien neurodégénératif (paralysie supranucléaire progressive, atrophie multisystématique, dégénérescence cortico-basale, syndrome parkinsonien secondaire à un syndrome démentiel, démence à corps de Lewy) ont été incorrectement classés comme ayant la maladie de Parkinson dans plus de 75 % des cas. En l'absence de tout autre traitement, ces patients reçoivent de la lévodopa aux mêmes doses que les patients parkinsoniens. Dans d'autres cas, ces patients sont considérés comme ayant la maladie de Parkinson par leur médecin traitant (erreur de diagnostic) et ils sont

donc traités comme tels. De même, certains patients atteints de tremblement essentiel sont incorrectement traités comme étant parkinsoniens (erreur de diagnostic) et ils sont donc identifiés par le modèle comme ayant la maladie de Parkinson.

En conséquence, le modèle prédictif entraîne des erreurs de classement (faux positifs) plus particulièrement (i) pour des personnes qui ont un diagnostic erroné de maladie de Parkinson, ou (ii) pour des personnes présentant des pathologies qui sont traitées comme la maladie de Parkinson, essentiellement les autres syndromes parkinsoniens neurodégénératifs ; ces pathologies sont toutefois considérablement plus rares que la maladie de Parkinson.

Les faux négatifs recevaient des doses moins élevées de médicaments, étaient moins fréquemment suivis par un neurologue et avaient une durée d'évolution de la maladie plus courte. Le modèle détecte donc moins bien les patients avec des formes moins évoluées et qui nécessitent des doses moins élevées de médicaments antiparkinsoniens.

6.3 Facteurs influençant les performances du modèle prédictif

Plusieurs facteurs sont susceptibles d'influencer les performances du modèle prédictif. Tout d'abord, par définition, seuls les patients traités sont présents dans les bases de remboursement de médicaments. Cette approche nécessite donc de faire l'hypothèse que la plupart des patients sont traités. En France, cette hypothèse est raisonnable pour la maladie de Parkinson, sauf, peut-être, chez les sujets les plus âgés. Dans une étude en population générale conduite en Gironde (Paquid, 1988-89), 11 % des patients parkinsoniens n'avaient pas été préalablement diagnostiqués et ont été identifiés au moment de l'étude alors qu'ils n'étaient pas traités ; ils étaient, pour la plupart, âgés de plus de 80 ans [11]. La difficulté à identifier les patients parkinsoniens chez les sujets les plus âgés résulte de différents facteurs (incertitude diagnostique, comorbidités, diagnostic plus tardif et difficile, etc.) [62]. Les modèles prédictifs établis à partir de profils de consommation médicamenteuse pourraient donc être moins fiables chez les personnes les plus âgées, en particulier en raison d'une moindre sensibilité.

Deuxièmement, le modèle repose uniquement sur les données enregistrées dans les bases de remboursement. Les maladies traitées par des médicaments en vente libre ou non remboursés ne peuvent donc pas être étudiées grâce à cette approche. De plus, les prescriptions peuvent ne pas être incluses dans les bases de remboursement de médicaments pour certains sous-groupes de la population. Dans cette étude, les personnes institutionnalisées dans des maisons de retraite avec des pharmacies à usage interne n'ont pas été identifiées. Toutefois, d'après une étude précédente, nous estimons qu'une minorité (<2 %) des utilisateurs de médicaments antiparkinsoniens se trouvent dans cette situation avant 80 ans [63].

Troisièmement, les médicaments utilisés pour développer le modèle prédictif doivent être relativement spécifiques de la maladie étudiée. Néanmoins, même si les mêmes médicaments sont utilisés pour d'autres pathologies que la maladie d'intérêt, un modèle prédictif peut présenter de bonnes performances si le mode d'utilisation des médicaments (par exemple en termes de dose et de fréquence de prescription) pour la maladie d'intérêt se distingue de celui des autres pathologies. Ainsi, nos résultats montrent que l'inclusion dans le modèle d'informations sur la dose de médicaments et la régularité des prescriptions améliore les performances du modèle en termes de discrimination et de calibration.

Il serait possible d'inclure dans le modèle une information sur les ALD pour maladie de Parkinson, mais nous ne l'avons volontairement pas fait pour les raisons suivantes : (i) tous les patients atteints de maladie de Parkinson ne sont pas en ALD. Dans cette étude, 68 % des patients parkinsoniens bénéficiaient de l'ALD : en dehors des raisons administratives décrites dans les points suivants, différentes explications peuvent être envisagées ; par exemple, oubli ou méconnaissance de la part des médecins traitants, choix volontaire des patients pour ne pas être pénalisés lors de la souscription d'assurances, etc. ; (ii) jusqu'en 2011, l'attribution de l'ALD pour maladie de Parkinson n'était pas systématique et dépendait du degré d'invalidité estimé par le médecin conseil ce qui explique probablement une certaine hétérogénéité dans la reconnaissance en ALD en fonction des caisses de l'Assurance maladie. Depuis le début de l'année 2011, l'attribution de l'ALD pour maladie de Parkinson a été modifiée par un décret qui stipule que toute affection comportant un syndrome parkinsonien non réversible (maladie de Parkinson et syndromes parkinsoniens atypiques) nécessitant l'administration d'un traitement antiparkinsonien pendant au moins six mois relève de l'exonération du ticket modérateur. Il est encore trop tôt pour savoir quel sera l'impact de cette modification ; (iii) le modèle prédictif pourrait être utilisé pour estimer la fréquence de la maladie de Parkinson dans différents régimes de l'Assurance maladie ; pour inclure l'ALD parmi les prédicteurs, il serait nécessaire de faire l'hypothèse que l'ALD est demandée ou attribuée de manière similaire pour les

affiliés aux différents régimes, or nous n'avons pas d'éléments permettant d'évaluer si cette hypothèse est raisonnable ou pas.

6.4 Possibilités d'utilisations des modèles prédictifs

Des bases de remboursement de médicaments sont maintenant disponibles dans de nombreux pays et permettent d'étudier des questions épidémiologiques bien qu'elles n'aient pas été conçues pour cet objectif.

Les modèles prédictifs développés à partir de ces bases présentent un certain nombre d'avantages pour estimer la fréquence d'une maladie. Pour des maladies comme la maladie de Parkinson qui sont sous-déclarées dans les certificats de décès [17], les modèles prédictifs représentent une nette amélioration par rapport aux études de mortalité [64]. Ils sont également utiles pour des maladies qui ne peuvent être identifiées par les résultats d'examens biologiques. Les enquêtes en deux phases (porte-à-porte) sont considérées comme la méthode de référence pour la maladie de Parkinson [11-13]. Cependant, cette approche est coûteuse et difficile à mettre en œuvre. De plus, en raison du caractère relativement rare de la maladie de Parkinson, un petit nombre de cas sont généralement identifiés. En revanche, les bases de remboursement de médicaments sont généralement disponibles à une grande échelle et permettent d'identifier un plus grand nombre de patients, ce qui permet d'obtenir des estimations de fréquence plus stables et une plus grande puissance pour étudier des facteurs de risque.

Un autre avantage de ces modèles est qu'une fois développés et évalués, ils peuvent être utilisés de façon répétée au cours du temps si aucun changement majeur n'intervient dans le traitement de la maladie d'intérêt. En revanche, si de nouveaux médicaments sont disponibles ou si les stratégies thérapeutiques évoluent, leurs performances doivent être réévaluées.

Cette approche semble particulièrement intéressante pour la surveillance épidémiologique des maladies car elle permet d'étudier des tendances temporelles. Si le lieu de résidence est disponible dans les bases de remboursement de médicaments, la distribution spatiale des maladies et leur association avec des facteurs environnementaux (comme les caractéristiques du milieu agricole dans notre exemple) peuvent être étudiées. Si d'autres caractéristiques individuelles, comme le sexe ou la profession, étaient disponibles dans les bases de données, les différences de fréquence de la maladie en fonction de ces caractéristiques pourraient être analysées et surveillées.

Ces modèles prédictifs pourraient être utilisés de différentes manières dans le cadre d'études étiologiques. Premièrement, des patients atteints d'une maladie d'intérêt peuvent être directement identifiés à l'aide d'un seuil de probabilité donné et la relation entre l'exposition et la maladie peut être étudiée en utilisant différents plans d'étude (étude cas-témoins, rapports de prévalence). Dans le cas de la maladie de Parkinson, le seuil maximisant la sensibilité (92,5 %) et la spécificité (86,4 %) parmi les utilisateurs d'antiparkinsoniens était de 0,255, tandis que la spécificité était supérieure à 99 % pour l'ensemble de la population. Lorsque le nombre de faux positifs pour l'évènement d'intérêt est négligeable, comme dans cet exemple, les estimations de l'association entre une exposition et la maladie d'intérêt ne sont pas biaisées sauf si l'erreur de classement dépend de l'exposition [65]. Si la spécificité du modèle (pour le seuil qui maximise le nombre de sujets bien classés) est plus faible, il est possible d'utiliser des seuils associés à une spécificité plus élevée et un nombre de faux positifs plus faible. Enfin, si la sensibilité et la spécificité du modèle sont connues, il est également possible de corriger les mesures d'association en utilisant des formules simples [65].

Une approche sensiblement différente consiste à utiliser le modèle prédictif comme un outil de dépistage dans le cadre d'une étude en deux temps. Le modèle permet alors, dans un premier temps, d'identifier des patients avec une forte probabilité d'avoir la maladie d'intérêt et pour lesquels des informations complémentaires (par exemple, des données cliniques ou des résultats d'examens complémentaires) sont recueillies dans un second temps afin de confirmer le diagnostic. Cette approche est envisageable dans des situations où les bases de remboursement de médicaments ne sont pas anonymes. Elle est préférable pour des études où des phénotypes purs sont nécessaires, comme les études d'association génétique, ou lorsque l'effet attendu des expositions est faible. On pourrait également, suivant une approche similaire, envisager d'utiliser les modèles prédictifs pour recruter des participants dans des essais thérapeutiques. Les investigateurs peuvent alors sélectionner le seuil de probabilité lors de la première étape en fonction de la représentativité qu'ils souhaitent et du nombre de faux positifs acceptables en termes de coûts et de temps. Le modèle prédictif de maladie de Parkinson a une spécificité au niveau de la population générale bien plus élevée que les autres outils de dépistage de la maladie de Parkinson [66-69] et son utilisation peut donc représenter un gain considérable de temps et d'argent en raison du faible nombre de faux positifs à examiner dans un deuxième temps.

6.5 Association avec les variables agricoles

Lorsque nous avons utilisé le modèle prédictif pour étudier la relation entre la prévalence de la maladie de Parkinson et les caractéristiques agricoles, nous avons observé une augmentation de 27 % de la prévalence parmi les affiliés à la MSA vivant dans des cantons caractérisés par les densités en exploitations spécialisées en fruits et autres cultures permanentes les plus élevées. Cette association est d'autant plus forte que la spécificité associée au seuil utilisé pour définir les cas parkinsonien est élevée ; cette observation indique que, dans cet exemple, des erreurs de classement (faux positifs) ont tendance à biaiser l'estimation de l'association vers la valeur nulle.

Les fruits de vergers (pompes, poires, cerises, abricots, prunes et pêches), les citrons, les kiwis, les fruits à coques (noix, noisettes...), les petits fruits (cassis, framboises) et les arbres de pépinières représentent les principales productions des exploitations spécialisées en fruits et autres cultures permanentes. En 1989, ces exploitations étaient la seconde orientation en termes de coûts alloués à la protection des cultures après les exploitations spécialisées dans l'horticulture et le maraîchage [70]. En 1992, les exploitations spécialisées en fruits et autres cultures permanentes (à l'exception des petits fruits et des pépinières) utilisaient trois fois plus d'herbicides, cinq fois plus de fongicides et neuf fois plus d'insecticides par hectare que les autres exploitations et elles étaient les exploitations qui utilisaient le plus d'herbicides et d'insecticides par hectare [71]. En 1998, tandis que les exploitations spécialisées en fruits et autres cultures permanentes (sans les pépinières) représentaient 1 % de la superficie agricole utilisée française, elles représentaient 21 % du marché total des insecticides [72]. En conséquence, les personnes qui appliquent des pesticides dans ce type d'exploitation sont probablement plus fortement exposées aux pesticides, en particulier aux insecticides, que dans d'autres types d'exploitations.

Cette association est cohérente avec une étude antérieure qui retrouve un risque augmenté de maladie de Parkinson parmi les travailleurs des vergers [73]. Plusieurs études observent également une association entre le risque de maladie de Parkinson et l'exposition aux insecticides [37;74-78]. Une discussion plus détaillée de ces résultats est disponible ailleurs [1].

6.6 Limites

Notre étude présente des limites. Tout d'abord, le modèle prédictif a été développé chez des personnes qui vérifient un certain nombre de critères d'inclusion. Les cas parkinsoniens avaient une durée d'évolution de la maladie de 15 ans ou moins. Toutefois, la sensibilité du modèle est similaire parmi les cas ayant une durée d'évolution longue (6-15 ans) ou plus courte (<6 ans). D'autre part, le traitement ne change pas de manière considérable avant ou après 15 ans d'évolution en termes de classe de médicaments utilisés. Il est donc peu probable que les performances du modèle changent considérablement pour les personnes dont la durée d'évolution est supérieure à 15 ans. Les sujets bénéficiant d'une ALD pour démence ou psychose n'ont pas été inclus dans l'échantillon qui a permis de développer le modèle. Toutefois, la fréquence de la démence augmente avec la durée d'évolution de la maladie et, comme indiqué ci-dessus, nos résultats montrent que les performances du modèle ne diminuent pas avec la durée de la maladie. Enfin, tous les sujets sont âgés de 80 ans ou moins. Parmi les sujets âgés de plus de 80 ans, la lévodopa (56 %) et le piribédil (46 %) représentent les principaux médicaments antiparkinsoniens. Pour ces deux médicaments, il n'y a pas d'arguments en faveur d'une interaction avec l'âge ; il est donc peu probable que les performances du modèle soient très différentes chez les sujets âgés traités. En revanche, une partie des sujets parkinsoniens institutionnalisés ne sont pas identifiés par le modèle prédictif, ce qui conduit probablement à sous-estimer la prévalence chez les personnes les plus âgées.

Deuxièmement, nous n'avons pas réalisé de validation externe du modèle, ce qui nécessiterait que des données similaires soient recueillies dans une autre population. En France, toutes les personnes qui bénéficient de l'Assurance maladie ont potentiellement le même accès aux soins médicaux et les affiliés des différents systèmes de l'Assurance maladie ont accès aux mêmes médecins, y compris des neurologues. Dans cette étude, la distribution des médicaments utilisés pour traiter les patients parkinsoniens et les doses journalières estimées sont classiques et similaires à celle observées dans des consultations spécialisées. En 2000, une conférence de consensus sur le traitement et le diagnostic de la maladie de Parkinson a été publiée par la Haute autorité de santé et elle a été depuis très largement diffusée aussi bien auprès des neurologues que des généralistes [10;79]. Parmi les cas atteints de maladie de Parkinson, 92 % avaient vu au moins une fois un neurologue et 78 % d'entre eux étaient suivis par un neurologue ; nous ne disposons pas d'estimations similaires parmi des affiliés à d'autres régimes de l'Assurance maladie, mais ces chiffres indiquent un suivi satisfaisant. De plus, le nombre de prédicteurs inclus dans les modèles est faible par rapport au nombre de sujets et il n'existait pas d'arguments statistiques en faveur d'un surapprentissage (overfitting). Ces observations suggèrent qu'il est peu vraisemblable qu'une étude réalisée dans une autre population française donnerait des résultats très

différents. En revanche, il n'est pas raisonnable d'envisager qu'un modèle prédictif développé dans un pays puisse être appliqué sans modification dans un autre contexte. En effet, il existe souvent des différences quant à l'accès aux soins médicaux et à la disponibilité des médicaments entre les pays. Dans le cas de la maladie de Parkinson, il existe des différences d'utilisation des médicaments antiparkinsoniens en Europe [80;81]. Néanmoins, un modèle développé dans un pays peut aider des chercheurs à développer un modèle similaire dans un autre pays.

Troisièmement, nous n'avons pas inclus des interactions entre les médicaments antiparkinsoniens lors du développement du modèle. Il est possible que la prescription simultanée de deux médicaments soit plus fortement associée à la maladie de Parkinson que chacun des médicaments considérés séparément. Des modèles plus complexes et des échantillons de plus grande taille seraient nécessaires pour réaliser les analyses correspondantes.

6.7 Conclusions

Les modèles prédictifs développés à partir des bases de remboursement de médicaments représentent une approche pertinente pour identifier des patients atteints de certaines maladies dont il serait ainsi possible d'estimer la fréquence à partir des seules bases de données médico-administratives. Cette approche a été illustrée dans le cadre d'une étude de prévalence de la maladie de Parkinson, mais une stratégie similaire pourrait être suivie pour estimer l'incidence. Ainsi, pour certaines pathologies, un système de surveillance sans recueil d'informations supplémentaires auprès des sujets pourrait être mis en place au niveau national. La possibilité d'études répétées représente un avantage important pour la surveillance.

Comme le montre l'exemple que nous présentons sur la relation entre la prévalence de la maladie de Parkinson et des caractéristiques agricoles, l'utilisation de cette approche dans le cadre d'études étiologiques peut également être envisagée si des données sur la fréquence d'une exposition, par exemple environnementale, sont disponibles au niveau géographique. Toutefois, une difficulté importante est liée au fait que les bases de remboursement comprennent peu d'informations individuelles ; dans le domaine "Santé-travail", par exemple, elles n'incluent pas de données sur la profession ou le secteur d'activité ni sur le statut professionnel (en activité, retraité, chômage).

Plusieurs des caractéristiques du modèle prédictif décrites dans ce rapport sont susceptibles d'être pertinentes pour d'autres pathologies. Le développement des modèles prédictifs doit cependant tenir compte des caractéristiques de la maladie d'intérêt ; par exemple, l'examen clinique a été utilisé dans cette étude sur la maladie de Parkinson pour établir le diagnostic de référence mais, pour d'autres maladies, des questionnaires, des échantillons biologiques, ou des bases de données administratives peuvent être utilisés ; de même, des codes de la classification internationale des maladies pourraient être inclus dans le modèle afin d'en améliorer les performances [82].

Références bibliographiques

- [1] Moisan F, Spinosi J, Dupupet JL, Delabre L, Mazurie JL, Goldberg M *et al.* The relation between type of farming and prevalence of Parkinson's disease among agricultural workers in five French districts. *Mov Disord* 2011;26(2):271-9.
- [2] Bower JH, Maraganore DM, McDonnell SK, Rocca WA. Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976-1990. *Neurology* 1999;52(6):1214-20.
- [3] Chaudhuri KR, Healy DG, Schapira AH. Non-motor symptoms of Parkinson's disease: diagnosis and management. *Lancet Neurol* 2006;5(3):235-45.
- [4] Savica R, Rocca WA, Ahlskog JE. When does Parkinson disease start? *Arch Neurol* 2010;67(7):798-801.
- [5] Hughes AJ, Daniel SE, BenShlomo Y, Lees AJ. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain* 2002;125(Pt 4):861-70.
- [6] Newman EJ, Breen K, Patterson J, Hadley DM, Grosset KA, Grosset DG. Accuracy of Parkinson's disease diagnosis in 610 general practice patients in the West of Scotland. *Mov Disord* 2009;24(16):2379-85.
- [7] Hughes AJ, Daniel SE, Lees AJ. Improved accuracy of clinical diagnosis of Lewy body Parkinson's disease. *Neurology* 2001;57(8):1497-9.
- [8] Ferini-Strambi L, Manconi M. Treatment of restless legs syndrome. *Parkinsonism Relat Disord* 2009;15 Suppl 4:S65-S70.
- [9] De Azevedo AA, Langguth B, de Oliveira PM, Rodrigues FR. Tinnitus treatment with piribedil guided by electrocochleography and acoustic otoemissions. *Otol Neurotol* 2009;30(5):676-80.
- [10] HAS. Haute autorité de santé. Conférence de consensus, la maladie de Parkinson : critères diagnostiques et thérapeutiques; 2000. 29 p. [consulté le 29/11/2011]. Disponible à partir de l'URL : http://www.has-sante.fr/portail/jcms/c_272069/la-maladie-de-parkinson-criteres-diagnostiques-et-therapeutiques.
- [11] De Rijk MC, Tzourio C, Breteler MMB, Dartigues JF, Amaducci L, Lopez-Pousa S *et al.* Prevalence of parkinsonism and Parkinson's disease in Europe: The EUROPARKINSON collaborative study. *J Neurol Neurosurg Psychiatry* 1997;62(1):10-5.
- [12] De Rijk MC, Breteler MM, Graveland GA, Ott A, Grobbee DE, van der Meche FG *et al.* Prevalence of Parkinson's disease in the elderly: the Rotterdam Study. *Neurology* 1995;45(12):2143-6.
- [13] Zhang ZX, Roman GC, Hong Z, Wu CB, Qu QM, Huang JB *et al.* Parkinson's disease in China: prevalence in Beijing, Xian, and Shanghai. *Lancet* 2005;365(9459):595-7.
- [14] Anderson DW, Schoenberg BS, Haerer AF. Prevalence surveys of neurologic disorders: methodologic implications of the Copiah County Study. *J Clin Epidemiol* 1988;41(4):339-45.

- [15] Tison F, Dartigues JF, Dubes L, Zuber M, Alperovitch A, Henry P. Prevalence of Parkinson's disease in the elderly: a population study in Gironde, France. *Acta Neurol Scand* 1994;90(2):111-5.
- [16] Elbaz A, Bower JH, Peterson BJ, Maraganore D, McDonnell SK, Ahlskog JE *et al.* Survival study of Parkinson Disease in Olmsted County, Minnesota. *Arch Neurol* 2003;60(1):91-6.
- [17] Phillips NJ, Reay J, Martyn CN. Validity of mortality data for Parkinson's disease. *J Epidemiol Community Health* 1999;53(9):587-8.
- [18] White D, Moore S, Waring S, Cook K, Lai E. Identifying incident cases of parkinsonism among veterans using a tertiary medical center. *Mov Disord* 2007;22(7):915-23.
- [19] Menniti-Ippolito F, Spila-Alegiani S, Vanacore N, Bonifati V, Diana G, Meco G *et al.* Estimate of parkinsonism prevalence through drug prescription histories in the Province of Rome, Italy. *Acta Neurol Scand* 1995;92(1):49-54.
- [20] Chio A, Magnani C, Schiffer D. Prevalence of Parkinson's disease in Northwestern Italy: Comparison of tracer methodology and clinical ascertainment of cases. *Mov Disord* 1998;13(3):400-5.
- [21] Van de Vijver DA, Roos RA, Jansen PA, Porsius AJ, de Boer A. Estimation of incidence and prevalence of Parkinson's disease in the elderly using pharmacy records. *Pharmacoepidemiol Drug Saf* 2001;10(6):549-54.
- [22] Noyes K, Liu H, Holloway R, Dick AW. Accuracy of Medicare claims data in identifying Parkinsonism cases: comparison with the Medicare current beneficiary survey. *Mov Disord* 2007;22(4):509-14.
- [23] Newman EJ, Grosset KA, Grosset DG. Geographical difference in Parkinson's disease prevalence within West Scotland. *Mov Disord* 2009;24(3):401-6.
- [24] Szumski NR, Cheng EM. Optimizing algorithms to identify Parkinson's disease cases within an administrative database. *Mov Disord* 2009;24(1):51-6.
- [25] Masalha R, Kordysh E, Alpert G, Hallak M, Morad M, Mahajnah M *et al.* The prevalence of Parkinson's disease in an Arab population, Wadi Ara, Israel. *Isr Med Assoc J* 2010;12(1):32-5.
- [26] Twelves D, Perkins KSM, Counsell C. Systematic review of incidence studies of Parkinson's disease. *Mov Disord* 2003;18(1):19-31.
- [27] Von Campenhausen S, Bornschein B, Wick R, Bötzel K, Sampaio C, Poewe W *et al.* Prevalence and incidence of Parkinson's disease in Europe. *Eur Neuropsychopharmacol* 2005;15(4):473-90.
- [28] Elbaz A, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE *et al.* Risk tables for parkinsonism and Parkinson's disease. *J Clin Epidemiol* 2002;55(1):25-31.
- [29] De Rijk MC, Launer LJ, Berger K, Breteler MM, Dartigues JF, Baldereschi M *et al.* Prevalence of Parkinson's disease in Europe: a collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology* 2000;54(11 Suppl 5):S21-S23.

- [30] Bower JH, Maraganore DM, McDonnell SK, Rocca WA. Incidence of progressive supranuclear palsy and multiple system atrophy in Olmsted County, Minnesota, 1976 to 1990. *Neurology* 1997;49(5):1284-8.
- [31] Nath U, Ben-Shlomo Y, Thomson RG, Morris HR, Wood NW, Lees AJ *et al.* The prevalence of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome) in the UK. *Brain* 2001;124(Pt 7):1438-49.
- [32] Langston JW. The etiology of Parkinson's disease with emphasis on the MPTP story. *Neurology* 1996;47(6):S153-S160.
- [33] Priyadarshi A, Khuder SA, Schaub EA, Shrivastava S. A meta-analysis of Parkinson's disease and exposure to pesticides. *Neurotoxicology* 2000;21(4):435-40.
- [34] Priyadarshi A, Khuder SA, Schaub EA, Priyadarshi SS. Environmental risk factors and Parkinson's disease: a metaanalysis. *Environ Res* 2001;86(2):122-7.
- [35] Kamel F, Hoppin JA. Association of pesticide exposure with neurologic dysfunction and disease. *Environ Health Perspect* 2004;112(9):950-8.
- [36] Dick FD, De Palma G, Ahmadi A, Scott NW, Prescott GJ, Bennett J *et al.* Environmental risk factors for Parkinson's disease and parkinsonism: the Geoparkinson study. *Occup Environ Med* 2007;64(10):666-72.
- [37] Elbaz A, Clavel J, Rathouz PJ, Moisan F, Galanaud JP, Delemotte B *et al.* Professional exposure to pesticides and Parkinson disease. *Ann Neurol* 2009;66(4):494-504.
- [38] Tanner CM, Kamel F, Ross GW, Hoppin JA, Goldman SM, Korell M *et al.* Rotenone, paraquat and Parkinson's disease. *Environ Health Perspect* 2011;119(6):866-72.
- [39] Spinosi J, Févotte J. Le programme MATPHYTO - Matrices cultures-expositions aux produits phytosanitaires. Saint-Maurice : Institut de veille sanitaire; 2008. 19 p. [consulté le 29/11/2011]. Disponible à partir de l'URL : <http://www.invs.sante.fr>.
- [40] Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov Disord* 2010;25(15):2649-53.
- [41] Steyerberg EW. Developing valid prediction models. Dans: Steyerberg EW, éditeur. *Clinical prediction models: a practical approach to development, validation, and updating*. New York: Springer; 2009: 113-331.
- [42] Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J R Stat Soc Ser C Appl Stat* 1994;43(3):429-67.
- [43] Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology--with an emphasis on fractional polynomials. *Methods Inf Med* 2005;44(4):561-71.
- [44] Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49(12):1373-9.
- [45] Harrell FE. Resampling, validating, describing and simplifying the model. Dans: Harrell FE, éditeur. *Regression modeling strategies with applications to linear models, logistic regression and survival analysis*. New York: Springer; 2001: 87-103.

- [46] Harrell FE. Binary logistic regression. Dans: Harrell FE, éditeur. Regression modeling strategies with applications to linear models, logistic regression and survival analysis. New York: Springer; 2001: 215-67.
- [47] Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biom J* 2008;50(4):457-79.
- [48] Insee. Institut national de la statistique et des études économiques. Pyramides des âges 2007. Paris, France; 2007. [consulté le 29/11/2011]. Disponible à partir de l'URL : <http://www.insee.fr>.
- [49] Couris CM, Colin C, Rabilloud M, Schott AM, Ecochard R. Method of correction to assess the number of hospitalized incident breast cancer cases based on claims databases. *J Clin Epidemiol* 2002;55(4):386-91.
- [50] Ministère de l'alimentation, de l'agriculture et de la pêche. Enquêtes - Recensement agricole - Pourquoi un recensement ? 2010. [consulté le 29/11/2011]. Disponible à partir de l'URL : <http://www.agreste.agriculture.gouv.fr>.
- [51] Réseau d'information comptable agricole. Commission européenne 2010. [consulté le 29/11/2011]. Disponible à partir de l'URL : http://ec.europa.eu/agriculture/rica/index_fr.cfm.
- [52] Zocchetti C, Consonni D, Bertazzi PA. Relationship between prevalence rate ratios and odds ratios in cross-sectional studies. *Int J Epidemiol* 1997;26(1):220-3.
- [53] Diez-Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* 2000;21:171-92.
- [54] Insee. Institut national de la statistique et des études économiques. Statistiques locales : revenus fiscaux localisés des ménages niveau supracommunal - année 2007-2009. [consulté le 29/11/2011]. Disponible à partir de l'URL : <http://www.statistiques-locales.insee.fr>.
- [55] Lix LM, Hobson DE, Azimaee M, Leslie WD, Burchill C, Hobson S. Socioeconomic variations in the prevalence and incidence of Parkinson's disease: a population-based analysis. *J Epidemiol Community Health* 2010;64(4):335-40.
- [56] Greenland S. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology* 1995;6(4):450-4.
- [57] Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. *Arch Environ Health* 1994;49(1):9-16.
- [58] Greenland S. A semi-Bayes approach to the analysis of correlated multiple associations, with an application to an occupational cancer-mortality study. *Stat Med* 1992;11(2):219-30.
- [59] Checkoway H, Nelson LM. Epidemiologic approaches to the study of Parkinson's disease etiology. *Epidemiology* 1999;10(3):327-36.
- [60] Lai BC, Schulzer M, Marion S, Teschke K, Tsui JK. The prevalence of Parkinson's disease in British Columbia, Canada, estimated by using drug tracer methodology. *Parkinsonism Relat Disord* 2003;9(4):233-8.

- [61] Van de Vijver DA, Stricker BH, Breteler MM, Roos RA, Porsius AJ, de Boer A. Evaluation of antiparkinsonian drugs in pharmacy records as a marker for Parkinson's disease. *Pharm World Sci* 2001;23(4):148-52.
- [62] Van Den Eeden SK, Tanner CM, Bernstein AL, Fross RD, Leimpeter A, Bloch DA *et al.* Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. *Am J Epidemiol* 2003;157(11):1015-22.
- [63] Elbaz A, Leleu H, Houssinot J, Mazurie JL, Dupupet JL, Goldberg M *et al.* Increased use of potentially inappropriate medications in patients with Parkinson's disease [Abstract]. *Neurology* 2010;74(9):A53
- [64] Mylne AQ, Griffiths C, Rooney C, Doyle P. Trends in Parkinson's disease related mortality in England and Wales, 1993-2006. *Eur J Neurol* 2009;16(9):1010-6.
- [65] Greenland S, Lash TL. Bias Analysis. Dans : Rothman KJ, Greenland S, Lash TL, éditeurs. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins.2008: 345-80.
- [66] Tanner CM, Gilley DW, Goetz CG. A brief screening questionnaire for parkinsonism [Abstract]. *Ann Neurology* 1990;28:267-8.
- [67] Rocca WA, Maraganore DM, McDonnell SK, Schaid DJ. Validation of a telephone questionnaire for Parkinson's disease. *J Clin Epidemiol* 1998;51(6):517-23.
- [68] Pramstaller PP, Falk M, Schoenhuber R, Poewe W. Validation of a mail questionnaire for parkinsonism in two languages (German and Italian). *J Neurol* 1999;246(2):79-86.
- [69] Kim JH, Cheong HK, Lee CS, Yi SE, Park KW. The validity and reliability of a screening questionnaire for Parkinson's disease in a community. *J Prev Med Public Health* 2010;43(1):9-17.
- [70] Réseau d'information comptable agricole. Commission européenne. *Méthodologie* 2009. [consulté le 29/11/2011]. Disponible à partir de l'URL : http://ec.europa.eu/agriculture/rica/index_fr.cfm.
- [71] European Crop Protection Association. *Agriculture and Environment : Consumption of Pesticides*.2009. [consulté le 29/11/2011]. Disponible à partir de l'URL : <http://nui.epp.eurostat.ec.europa.eu>.
- [72] Inra. Institut national de recherche agronomique. *Pesticides, agriculture et environnement – Réduire l'utilisation des pesticides et en limiter les impacts environnementaux. Expertise scientifique collective*. 2005.
- [73] Hertzman C, Wiens M, Bowering D, Snow B, Calne D. Parkinson's disease: a case-control study of occupational and environmental risk factors. *Am J Ind Med* 1990;17(3):349-55.
- [74] Gorell JM, Johnson CC, Rybicki BA, Peterson EL, Richardson RJ. The risk of Parkinson's disease with exposure to pesticides, farming, well water, and rural living. *Neurology* 1998;50(5):1346-50.
- [75] Kamel F, Tanner C, Umbach D, Hoppin J, Alavanja M, Blair A *et al.* Pesticide exposure and self-reported Parkinson's disease in the agricultural health study. *Am J Epidemiol* 2007;165(4):364-74.

- [76] Hancock DB, Martin ER, Mayhew GM, Stajich JM, Jewett R, Stacy MA *et al.* Pesticide exposure and risk of Parkinson's disease: A family-based case-control study. *BMC Neurol* 2008;8:6.
- [77] Brown TP, Rumsby PC, Capleton AC, Rushton L, Levy LS. Pesticides and Parkinson's disease--is there a link? *Environ Health Perspect* 2006;114(2):156-64.
- [78] Waller LA, Gotway CA. Analyzing public health data. Dans: Waller LA, Gotway CA, éditeurs. *Applied spatial statistics for public health data*. Hoboken: Wiley-Interscience; 2004: 7-37.
- [79] Fayard C, Bonaventure A, Benatru I, Roze E, Dumurgier J, Moisan F *et al.* Impact of recommendations on the initial therapy of Parkinson's disease: a population-based study in France. *Parkinsonism Relat Disord* 2011;17(7):543-6.
- [80] de Pedro-Cuesta J, Petersen IJ, Vassilopoulos D, Micheli F, Garcia-Inesta A. Epidemiological assessment of levodopa use by populations. *Acta Neurol Scand* 1991;83(5):328-35.
- [81] Rosa MM, Ferreira JJ, Coelho M, Freire R, Sampaio C. Prescribing patterns of antiparkinsonian agents in Europe. *Mov Disord* 2010;25(8):1053-60.
- [82] Henderson T, Shephard J, Sundararajan V. Quality of diagnosis and procedure coding in ICD-10 administrative data. *Med Care* 2006;44(11):1011-9.

Annexes

ANNEXE 1 - Développement d'un modèle prédictif et mesure des performances

A- Développement du modèle prédictif

Nous avons développé un modèle prédictif en suivant une approche séquentielle qui s'inspire en grande partie des travaux antérieurs de FE Harrell et EW Steyerberg [1-4]. Les principales étapes de la construction du modèle prédictif sont brièvement décrites ci-dessous.

Inspection des données

La première étape vise à caractériser la distribution des prédicteurs. Combien de personnes sont exposées à chaque prédicteur ? Y a-t-il un nombre suffisant de sujets dans chaque niveau des variables catégorielles ou certaines catégories doivent-elles être regroupées ? La répartition des prédicteurs continus est-elle normale ? Si non, une transformation est-elle nécessaire ? Le traitement des valeurs manquantes est un autre aspect important mais ne concernait pas directement les analyses que nous avons réalisées.

Codage des prédicteurs continus

Les variables continues peuvent être transformées en variables catégorielles pour des analyses exploratoires, mais on considère généralement que cette stratégie présente des inconvénients pour la construction des modèles prédictifs car elle conduit à une perte d'information et à des probabilités prédites discontinues. Une solution consiste à utiliser des termes linéaires, mais cela suppose que la relation entre l'évènement d'intérêt et la variable continue soit linéaire sur l'échelle du modèle de régression.

Un certain nombre de solutions plus souples sont disponibles, comme les polynômes fractionnaires, les splines ou les modèles additifs généralisés. Lorsque plus de 55 personnes avaient utilisé un médicament donné, nous avons utilisé les polynômes fractionnaires pour identifier les transformations des variables continues les plus parcimonieuses et associées à la meilleure adéquation. Nous avons utilisé pour cela une stratégie de sélection simultanée de polynômes fractionnaires pour plusieurs variables continues (multivariable fractional polynomial modeling, MFP) [5].

La spécification du modèle

Restrictions sur le nombre de prédicteurs : la construction d'un modèle prédictif consiste à trouver un compromis entre l'inclusion de prédicteurs importants tout en évitant le surajustement. Il est important de prendre en compte à la fois le nombre de prédicteurs potentiels et la taille de l'échantillon. Le nombre de prédicteurs potentiels est souvent grand et une manière d'améliorer la validité d'un modèle prédictif est de restreindre le nombre de prédicteurs qui seront inclus dans le modèle ; de plus, des modèles plus simples sont généralement plus faciles à interpréter.

Une combinaison de différentes approches est habituellement nécessaire pour restreindre le nombre de prédicteurs. Des connaissances antérieures et avis d'experts peuvent être pris recueillis. Il est important de prendre en compte la distribution des prédicteurs. Par exemple, il peut être déraisonnable d'inclure des prédicteurs avec beaucoup de valeurs manquantes ou avec une distribution étroite et peu de variabilité. Enfin, certaines variables peuvent être regroupées sur la base d'avis d'experts ou en utilisant des méthodes statistiques telles que l'analyse en composantes principales. Nous avons par exemple combiné des antiparkinsoniens appartenant à la même classe thérapeutique et qui sont utilisés de manière similaire (tableau 1).

Sélection des prédicteurs : lorsqu'une liste limitée de prédicteurs candidats a été définie, leur nombre peut être encore réduit au cours de la phase de modélisation. La colinéarité peut parfois conduire à des estimations instables des coefficients de régression et des variables fortement corrélés peuvent être combinées ou l'une d'entre elles peut être supprimée. La décision d'inclure ou non une variable dans le modèle final peut aussi s'appuyer sur des approches statistiques, telles que la sélection pas-à-pas. Nous avons utilisé une procédure de sélection descendante ($p \leq 0,20$ pour qu'une variable soit retenue dans le modèle). Une discussion des avantages et inconvénients des procédures de sélection pas-à-pas dépasse les objectifs de ce rapport et de nombreuses références portant sur cette question sont disponibles par ailleurs [1;2;6]. Des méthodes plus sophistiquées sont disponibles (par exemple, rééchantillonnage, méthodes bayésiennes, rétrécissement ou shrinkage, Lasso) [6;7], mais elles sont généralement plus

difficiles à mettre en œuvre à l'aide des logiciels statistiques habituels et demandent plus de temps de calcul.

Estimation du modèle

Dans certains cas, après estimation du modèle de régression, il peut être nécessaire de réaliser un rétrécissement (shrinkage) des coefficients de régression (il s'agit de ramener les coefficients de régression estimés par le modèle à des valeurs moins extrêmes) ou d'utiliser des méthodes de maximum de vraisemblance pénalisée afin d'améliorer les prédictions du modèle, en particulier lorsqu'il existe un surajustement.

Ces méthodes sont particulièrement utiles pour des échantillons de petite taille ou lorsque le nombre de prédicteurs est grand. Nous n'avons pas rencontré cette difficulté dans notre étude, car le nombre d'observations par prédicteur était important et nous n'avons pas observé d'arguments en faveur d'un surajustement aux données.

B- Performances des modèles prédictifs

Nous présentons une brève description des principales mesures utilisées dans le rapport pour caractériser différents aspects des performances des modèles prédictifs. Une description plus complète de ces mesures et d'autres mesures non décrites ici est disponible ailleurs [2;3;8].

Mesures de performance globale

La performance globale du modèle est quantifiée à partir de la différence entre les résultats observés et prédits ; pour des événements binaires, l'évènement prédit est défini par la probabilité prédite par un modèle logistique. Les meilleurs modèles sont caractérisés par des différences entre les observations et prédictions faibles.

- *R² de Nagelkerke* : il s'agit d'une version adaptée à la régression logistique du R² des modèles de régression linéaire. Il s'agit d'une estimation de la proportion de variabilité de l'évènement qui est prédite par le modèle ; elle est comprise entre 0 % et 100 %. Cette mesure repose sur la comparaison des observations et du logarithme des prédictions. Elle peut être utilisée pour comparer l'impact de différents codages des variables [9].

- *Le score de Brier* (scaled Brier score) varie aussi entre 0 % et 100 % et repose sur la comparaison des observations et du carré des prédictions. Un avantage du score de Brier par rapport au R² de Nagelkerke est qu'il pénalise moins les prédictions proches de 0 % ou 100 % qui sont fausses.

Mesures de discrimination

Les mesures de discrimination ont pour objectif de quantifier si les personnes sont bien classées par le modèle comme ayant ou pas l'évènement d'intérêt.

- *La statistique c* est la mesure de discrimination la plus couramment utilisée et représente la probabilité de concordance entre les prédictions et les observations. Pour des événements binaires, elle correspond à l'aire sous la courbe. Une valeur de 0,5 correspond à des prédictions aléatoires tandis qu'une valeur de 1,0 indique une prédiction parfaite.

- La *sensibilité* est calculée en divisant le nombre de vrais positifs par le nombre total de patients malades. La *spécificité* est calculée en divisant le nombre de vrais négatifs par le nombre de patients non malades. La *valeur prédictive positive* est calculée en divisant le nombre de vrais positifs par le nombre de patients classés comme étant atteints. La *valeur prédictive négative* est calculée en divisant le nombre de vrais négatifs par le nombre de patients classés comme n'étant pas atteints. La sensibilité, spécificité, valeur prédictive positive et valeur prédictive négative peuvent être estimées pour différents seuils de probabilité du modèle prédictif. En particulier, elles peuvent être calculées pour le seuil de probabilité optimal qui maximise l'indice de Youden et qui correspond à la probabilité qui maximise le nombre d'individus correctement classés [8].

- *La pente la discrimination* est une autre mesure simple. Elle est calculée comme la différence entre les moyennes des prédictions des personnes avec et sans l'évènement d'intérêt ; elle mesure la séparation des prédictions entre les deux groupes.

Calibration

La calibration mesure la concordance entre les observations et les prédictions.

- *Les graphiques de calibration* présentent les probabilités prédites sur l'axe X et les probabilités observées sur l'axe Y. La relation entre les probabilités observées et les quantiles (par exemple, déciles) des probabilités prédites est représentée graphiquement. Une concordance parfaite est représentée par une ligne oblique à 45°.

- *Le test de Le Cessie et Van Houwelingen* est un test d'adéquation des données au modèle. Il repose sur la somme des carrés des résidus lissés [10] et on considère qu'il a de meilleures propriétés statistiques qu'un autre test courant (test de Hosmer et Lemeshow [11;12]).

Validation interne, surajustement, et optimisme

- La validation du modèle a pour objectif d'évaluer si un modèle est susceptible de prédire avec précision un événement parmi des sujets qui n'ont pas été utilisés pour construire le modèle [13]. La validité interne (ou reproductibilité) d'un modèle prédictif est estimée grâce à l'échantillon qui a servi à construire le modèle ; plusieurs méthodes sont possibles, y compris la cross-validation, le jackknife ou le bootstrap [14,15].

- *Surajustement* : lorsqu'un modèle prédictif a de bonnes performances dans l'échantillon qui a été utilisé pour le développer, mais qu'il ne prédit pas bien l'évènement dans un nouvel échantillon, on parle de surajustement : en d'autres termes, le surajustement conduit à être optimiste sur les performances du modèle. Le rééchantillonnage par bootstrap est une technique statistique qui permet de tirer au sort des échantillons de taille n avec remplacement à partir d'un échantillon initial de taille n . Une application importante du bootstrap est la quantification de l'optimisme d'un modèle prédictif.

- *L'optimisme* est la différence entre les mesures des performances du modèle prédictif lorsqu'il est appliqué à l'échantillon initial et aux échantillons issus du bootstrap ; il peut être estimé pour n'importe quelle mesure de performance (statistique c , R^2 , ou mesures de calibration). Très sommairement, le modèle prédictif est construit à partir d'un échantillon obtenu par bootstrap (en suivant toutes les étapes décrites ci-dessus, y compris la sélection des variables) et les performances du modèle résultant sont évaluées dans l'échantillon initial : l'optimisme est la différence entre les performances dans l'échantillon obtenu par bootstrap et initial ; d'autres approches existent [14;15]. Nous avons calculé l'optimisme à partir de 200 échantillons obtenus par bootstrap afin d'obtenir une estimation stable.

- *La statistique c corrigée* pour l'optimisme est calculée en soustrayant l'estimation de l'optimisme de la statistique c originale. L'estimation corrigée pour l'optimisme est utilisée pour estimer la performance attendue dans un échantillon indépendant. Une approche similaire peut être utilisée pour le R^2 de Nagelkerke.

- *La calibration globale* est la différence entre la moyenne des valeurs observées et la moyenne des valeurs prédites par le modèle dans les échantillons obtenus par bootstrap. Elle mesure si les prédictions sont systématiquement trop basses ou trop élevées [3].

- *La pente de calibration* est une mesure du surajustement. En l'absence de surajustement, elle est égale à 1 ; elle est autrement inférieure à 1 et peut être utilisée comme un facteur de shrinkage afin de diminuer l'effet des prédicteurs avant d'appliquer le modèle prédictif dans un nouvel échantillon.

Références complémentaires

- [1] Harrell FE. Regression modeling strategies with applications to linear models, logistic regression and survival analysis. New York: Springer; 2001.
- [2] Steyerberg EW. Clinical prediction models: A practical approach to development, validation, and updating. New York: Springer; 2009.
- [3] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21(1):128-38.
- [4] Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15(4):361-87.
- [5] Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society Series A-Statistics in Society* 1999;162(1):71-94.

- [6] Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 2000;19(8):1059-79.
- [7] Austin PC. Bootstrap model selection had similar performance for selecting authentic and noise variables compared to backward variable elimination: a simulation study. *J Clin Epidemiol* 2008;61(10):1009-17.
- [8] Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biom J* 2008;50(4):457-79.
- [9] Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika* 1991;78(3):691-2.
- [10] Le Cessie S., Van Houwelingen HC. Testing the fit of a regression model via score tests in random effects models. *Biometrics* 1995;51(2):600-14.
- [11] Marcin JP, Romano PS. Size matters to a model's fit. *Crit Care Med* 2007;35(9):2212-3.
- [12] Hosmer DW, Hosmer T, le CS, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997;16(9):965-80.
- [13] Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19(4):453-73.
- [14] Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005;21(15):3301-7.
- [15] Efron B, Tibshirani RJ. *An introduction to the bootstrap*. Boca Raton: Chapman & Hall/CRC; 1998.

ANNEXE 2 - Exemple de classification d'une exploitation agricole selon les orientations technico-économiques des exploitations

Cet exemple est présenté par le Réseau d'information comptable agricole (RICA) sur leur site internet (http://ec.europa.eu/agriculture/rica/methodology2_fr.cfm).

Soit une exploitation agricole ayant deux types de production, vaches laitières (n=50) et truies reproductrices (n=10). D'après les productions de cette exploitation et la liste des orientations présentées dans le tableau 2, deux orientations peuvent convenir : « Exploitations bovines spécialisées - orientation lait » ou « Exploitations de polyélevage à orientation herbivores ». Pour classer correctement cette exploitation dans l'une des deux orientations, il est nécessaire de calculer la marge brute standard (MBS) de chaque production ainsi que la MBS totale de l'exploitation. Ces calculs sont présentés dans le tableau ci-dessous à partir des MBS de référence fournies par l'Union Européenne.

Production	Quantité	MBS de référence ^a	MBS de la production	% par rapport à la MBS de l'exploitation
Vaches laitières	50	700	50 x 700=35 000	92 %
Truies reproductrices	10	300	10 x 300=3 000	8 %
Total exploitation			38 000	100 %

Abréviation : MBS : marge brute standard.

^a *D'après l'Union européenne.*

La MBS totale de l'exploitation résulte donc pour 92 % du cheptel laitier et pour 8 % de la production de truies reproductrices. A partir ces importances relatives, il est possible de classer correctement l'exploitation en se reportant aux définitions des orientations technico-économiques présentées dans le tableau ci-dessous.

Orientations technico-économiques des exploitations	Définition
Exploitations bovines spécialisées - orientation lait	Le cheptel laitier représente plus de 2/3 de la MBS totale.
Exploitations de polyélevage à orientation herbivores	Le cheptel herbivore représente entre 1/3 et 2/3 de la MBS totale.

Le cheptel laitier représentant plus des deux tiers de la MBS totale, cette exploitation agricole est donc classée dans l'orientation technico-économique à titre principal « Exploitations bovines spécialisées - orientation lait ».

ANNEXE 3 - Autres variables quantitatives caractérisant la consommation de médicaments antiparkinsoniens chez les participants avec ou sans maladie de Parkinson

Médicaments antiparkinsoniens ^a	Avec MP N=320	Sans MP N=794	OR ^b	IC 95 % ^b
Moyenne du nombre de prescriptions par an (ET)				
Agonistes dopaminergiques de type 1	13,5 (6,6)	6,4 (6,3)	4,5	2,7-7,6
Agonistes dopaminergiques de type 2	9,4 (5,2)	8,8 (4,3)	1,2	0,5-2,7
Anticholinergiques	9,1 (3,3)	7,8 (4,7)	1,3	0,8-2,3
Inhibiteurs de la COMT	9,0 (4,0)	4,6 (4,0)	3,5	1,6-7,9
Lévodopa	13,3 (7,4)	8,4 (5,5)	2,5	1,9-3,4
Piribédil	9,6 (3,4)	7,3 (4,2)	1,9	1,4-2,5
Moyenne du nombre de boîtes achetées par an (ET)				
Agonistes dopaminergiques de type 1	49,4 (36,3)	18,8 (27,2)	3,9	2,3-6,7
Agonistes dopaminergiques de type 2	24,0 (15,9)	16,2 (10,9)	1,9	0,7-5,3
Anticholinergiques	15,0 (8,7)	13,7 (12,0)	1,1	0,7-1,9
Inhibiteurs de la COMT	11,3 (6,5)	4,7 (4,3)	5,8	1,9-17,9
Lévodopa	22,3 (15,8)	12,9 (10,5)	2,7	2,0-3,9
Piribédil	27,7 (14,2)	10,1 (8,2)	4,0	3,1-5,3
Dose moyenne journalière (ET)^c				
Agonistes dopaminergiques de type 1	232,5 (138,4)	88,8 (123,5)	4,1	2,5-6,7
Lévodopa	466,5 (301,0)	278,4 (183,1)	2,8	2,0-4,0
Piribédil	124,2 (51,4)	58,4 (42,7)	3,9	2,9-5,1

Abréviations : COMT : catéchol-O-méthyltransférase ; ET : écart-type ; IC : intervalle de confiance ; MP : maladie de Parkinson ; OR : odds ratio.

^aLa moyenne du nombre de prescriptions par an, la moyenne du nombre de boîtes par an et la dose moyenne journalière ont été calculé parmi les personnes ayant reçu le traitement. Les variables quantitatives ont été définies pour les médicaments antiparkinsoniens avec plus de 55 personnes prenant le traitement.

^bLes OR, les IC et les tests sont calculés en utilisant la régression logistique. Pour les variables continues, les OR sont exprimés pour l'augmentation d'un écart-type.

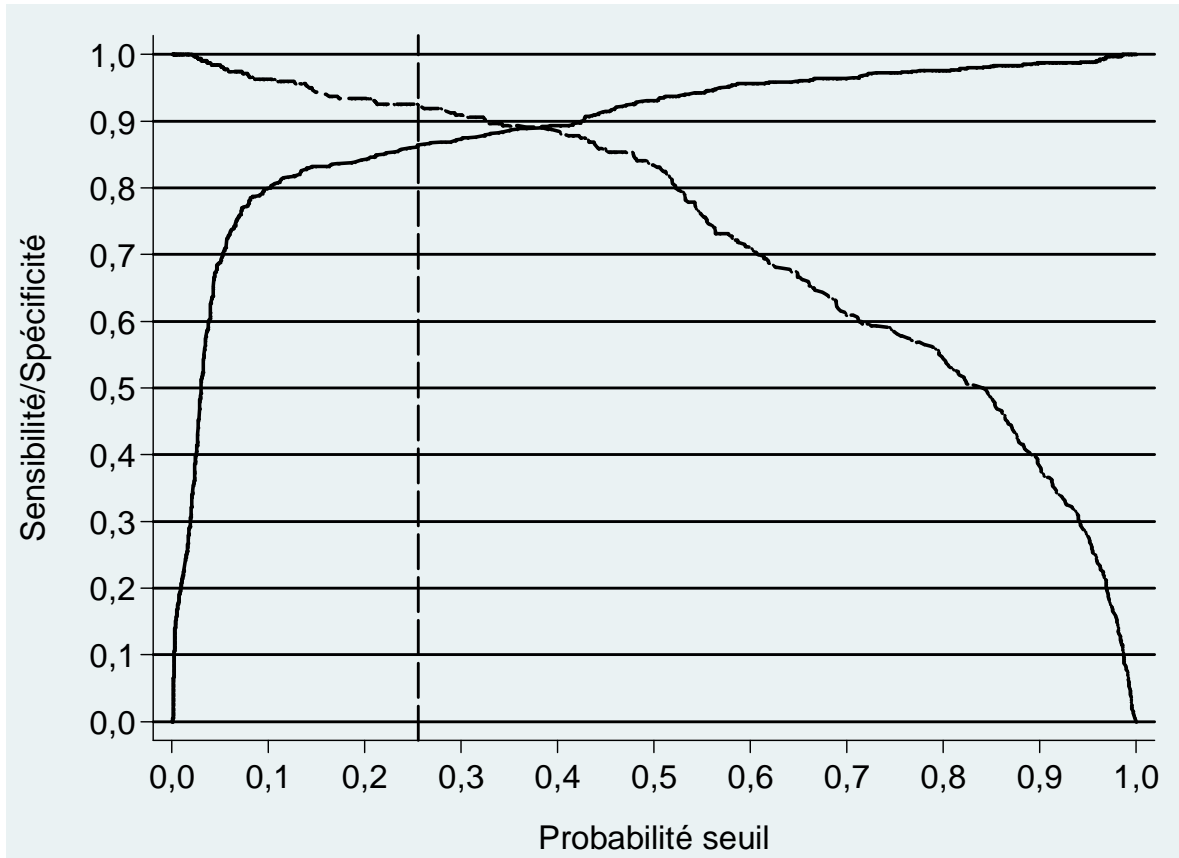
^cLa dose moyenne journalière est exprimée en milligrammes d'équivalent dose de lévodopa (EDL) par jour. Elle n'a pas été calculée si moins de 55 personnes ont reçu le médicament ou si aucune équivalence de dose de lévodopa est disponible.

ANNEXE 4 - Capacité de discrimination de chaque médicament antiparkinsonien (variable binaire) parmi les personnes ayant reçu au moins un médicament antiparkinsonien en 2007

Médicaments antiparkinsoniens	Sensibilité (%)	Spécificité (%)	Valeur prédictive positive (%)	Valeur prédictive négative (%)	statistique c
Agonistes dopaminergiques de type 1	34,4	91,3	61,5	77,5	0,628
Agonistes dopaminergiques de type 2	4,7	98,9	62,5	72,0	0,518
Amantadine	5,9	99,0	70,4	72,3	0,525
Anticholinergiques	4,4	89,5	14,4	69,9	0,530
Inhibiteurs de la COMT	26,3	98,6	88,4	76,8	0,624
Lévodopa	86,6	82,9	67,1	93,9	0,847
Piribédil	25,0	33,1	13,1	52,3	0,709
Sélégiline	12,5	98,6	78,4	73,7	0,556

Abréviation : COMT : catéchol-O-méthyltransférase.

ANNEXE 5 - Sensibilité et spécificité du modèle prédictif développé à partir des doses cumulées de médicaments antiparkinsoniens suivant le seuil utilisé pour appliquer le modèle parmi les personnes ayant reçu au moins un médicament antiparkinsonien en 2007



La ligne en pointillé correspond à la sensibilité, la ligne pleine à la spécificité. La ligne verticale en pointillé long représente le seuil qui maximise le nombre d'individus correctement classés ($p=0,255$). Ce seuil est associé à une sensibilité de 92,5 % et une spécificité de 86,4 %.

ANNEXE 6 - Modèle prédictif de la maladie de Parkinson développé à partir des médicaments antiparkinsoniens codés comme des variables binaires

Caractéristiques	Codage	Coefficient (ES)	Z-ratio ^a	OR ^b	IC 95 % ^b	p ^b
Constante	--	-17,710 (5,48)	-3,2	--	--	--
Âge en années	PF : $x + x^2$	4,301 (1,72)	2,5	8,1	1,6-41,9	0,012
		-0,034 (0,01)	-2,6	0,7	0,5-0,9	0,009
Sexe	Homme vs femme	0,315 (0,21)	1,5	1,4	0,9-2,1	0,126
Consultations avec un neurologue	≥1 vs aucune	1,377 (0,27)	5,0	4,0	2,3-6,8	<0,001
Consultations avec un généraliste	≥1 vs aucune	NR	--	--	--	--
Proportion du temps traité	100 % vs <100 %	1,885 (0,29)	6,6	6,6	3,8-11,5	<0,001
Agonistes dopaminergiques de type 1	Oui vs non	1,341 (0,29)	4,7	3,8	2,2-6,7	<0,001
Agonistes dopaminergiques de type 2	Oui vs non	NR	--	--	--	--
Amantadine	Oui vs non	NR	--	--	--	--
Anticholinergiques	Oui vs non	-0,626 (0,47)	-1,3	0,5	0,2-1,4	0,198
Inhibiteurs de la COMT	Oui vs non	1,176 (0,38)	3,1	3,2	1,5-6,9	0,002
Lévodopa	Oui vs non	3,139 (0,25)	12,8	23,1	14,3-37,3	<0,001
Piribédil	Oui vs non	NR	--	--	--	--
Sélégiline	Oui vs non	2,151 (0,53)	4,1	8,6	3,0-24,2	<0,001

Abréviations : COMT : catéchol-O-méthyltransférase ; ES : erreur standard ; IC : Intervalle de confiance ; NR : non retenu dans le modèle ; OR : odds ratio ; PF : polynômes fractionnaires ; x : terme du polynôme fractionnaire.

Le logit de la probabilité qu'une personne ait la maladie Parkinson est calculé en sommant la constante du modèle et les estimations des coefficients du modèle multipliées par la valeur de chaque variable pour cette personne. La probabilité est ensuite calculée en prenant l'inverse de $(1 + e^{-\text{logit}})$.

^aLes Z-ratios sont calculés en divisant les coefficients de régression par leur erreur standard.

^bLes OR, les IC et les tests sont calculés en utilisant la régression logistique. Pour l'âge, l'OR est exprimé pour l'augmentation d'un écart-type (ET=10,3). Une procédure de sélection descendante, avec une valeur de p égale à 0,20 pour qu'une variable reste dans le modèle, a été utilisée.

ANNEXE 7 - Analyses de sensibilité, performances des modèles développés à partir de variables binaires ou d'autres variables quantitatives pour caractériser la consommation de médicaments antiparkinsoniens

Performances	Variables binaires	Nombre de prescriptions par an	Nombre de boîtes par an	Dose moyenne journalière
Performances globales				
R ² de Nagelkerke	67,2 %	68,3 %	71,1 %	71,7 %
Score de Brier standardisé	57,9 %	59,1 %	62,4 %	62,1 %
Discrimination				
Statistique c (IC 95 %)	0,939 (0,924-0,953)	0,938 (0,923-0,954)	0,952 (0,939-0,964)	0,953 (0,942-0,965)
Seuil optimal	0,256	0,143	0,272	0,271
Sensibilité (IC 95 %) ^a	89,7 % (85,9-92,6)	93,8 % (90,5-95,9)	91,9 % (88,4-94,4)	92,5 % (89,1-94,9)
Spécificité (IC 95 %) ^a	86,9 % (84,4-89,1)	83,5 % (80,8-85,9)	87,3 % (84,8-89,2)	86,0 % (83,4-88,3)
Valeur prédictive positive (IC 95 %) ^a	73,4 % (45,9-92,6)	69,6 % (46,0-95,9)	74,4 % (46,0-94,4)	72,7 % (45,9-94,9)
Valeur prédictive négative (IC 95 %) ^a	95,4 % (93,7-96,7)	97,1 % (95,5-98,1)	96,4 % (94,8-97,5)	96,6 % (95,0-97,7)
Pente de discrimination	0,580	0,592	0,623	0,622
Calibration				
Test de Le Cessie-Van Houwelingen	p=0,591	p=0,645	p=0,654	p=0,671
Validation interne^b				
Optimisme du R ² de Nagelkerke	1,8 %	2,2 %	2,7 %	2,4 %
R ² de Nagelkerke corrigé	65,4 %	65,9 %	68,4 %	69,3 %
Optimisme de la statistique c	0,9 %	1,2 %	1,3 %	1,2 %
Statistique c corrigée	0,934	0,932	0,945	0,947
Calibration globale ^c	-2,5 %	-2,8 %	-4,4 %	0,5 %
Pente de calibration	95,2 %	93,9 %	92,1 %	92,8 %

Abréviation : IC : intervalle de confiance.

^aCalculée pour le seuil optimal qui maximise le nombre d'individus correctement classés parmi les personnes ayant reçu au moins un médicament antiparkinsonien en 2007.

^bValidation interne réalisée par la méthode du bootstrap incluant l'étape de sélection.

^cCalibration-in-the-large dans la littérature anglophone.

ANNEXE 8 - Analyses de sensibilité : performances de différents modèles prédictifs développés à partir des doses cumulées de médicaments antiparkinsoniens pour différentes durées de suivi pour définir les consommations de médicaments (3, 6, 9 mois)

Performances	Durée du suivi		
	3 mois ^a	6 mois ^b	9 mois ^c
Performances globales			
R ² de Nagelkerke	66,9 %	68,9 %	70,8 %
Score de Brier standardisé	58,4 %	60,5 %	62,7 %
Discrimination			
Statistique c (IC 95 %)	0,936 (0,920-0,952)	0,943 (0,928-0,958)	0,950 (0,937-0,963)
Seuil optimal	0,264	0,355	0,427
Sensibilité (IC 95 %) ^a	92,1 % (88,6-94,6)	90,9 % (87,3-93,6)	87,8 % (83,7-90,9)
Spécificité (IC 95 %) ^a	82,0 % (78,7-84,9)	85,8 % (82,9-88,2)	91,1 % (88,8-92,9)
Valeur prédictive positive (IC 95 %) ^d	73,6 % (46,0-94,6)	75,3 % (45,9-93,6)	80,9 % (45,9-90,9)
Valeur prédictive négative (IC 95 %) ^d	95,0 % (92,7-96,6)	95,2 % (93,2-96,6)	94,5 % (92,6-96,0)
Pente de discrimination	0,580	0,601	0,624
Calibration			
Test de Le Cessie-Van Houwelingen	p=0,132	p=0,047	p=0,192
Validation interne^e			
Optimisme du R ² de Nagelkerke	2,5 %	2,5 %	2,8 %
R ² de Nagelkerke corrigé	64,4 %	66,5 %	68,0 %
Optimisme de la statistique c	1,4 %	1,3 %	1,4 %
Statistique c corrigée	0,929	0,936	0,943
Calibration globale ^f	-1,8 %	-1,6 %	-3,0 %
Pente de calibration	93,6 %	93,4 %	92,7 %

Abréviations : IC : intervalle de confiance ; MP : maladie de Parkinson.

^aModèle développé à partir de 900 consommateurs de médicaments antiparkinsoniens (317 patients parkinsoniens et 583 patients non parkinsoniens).

^bModèle développé à partir de 986 consommateurs de médicaments antiparkinsoniens (319 patients parkinsoniens et 667 patients non parkinsoniens).

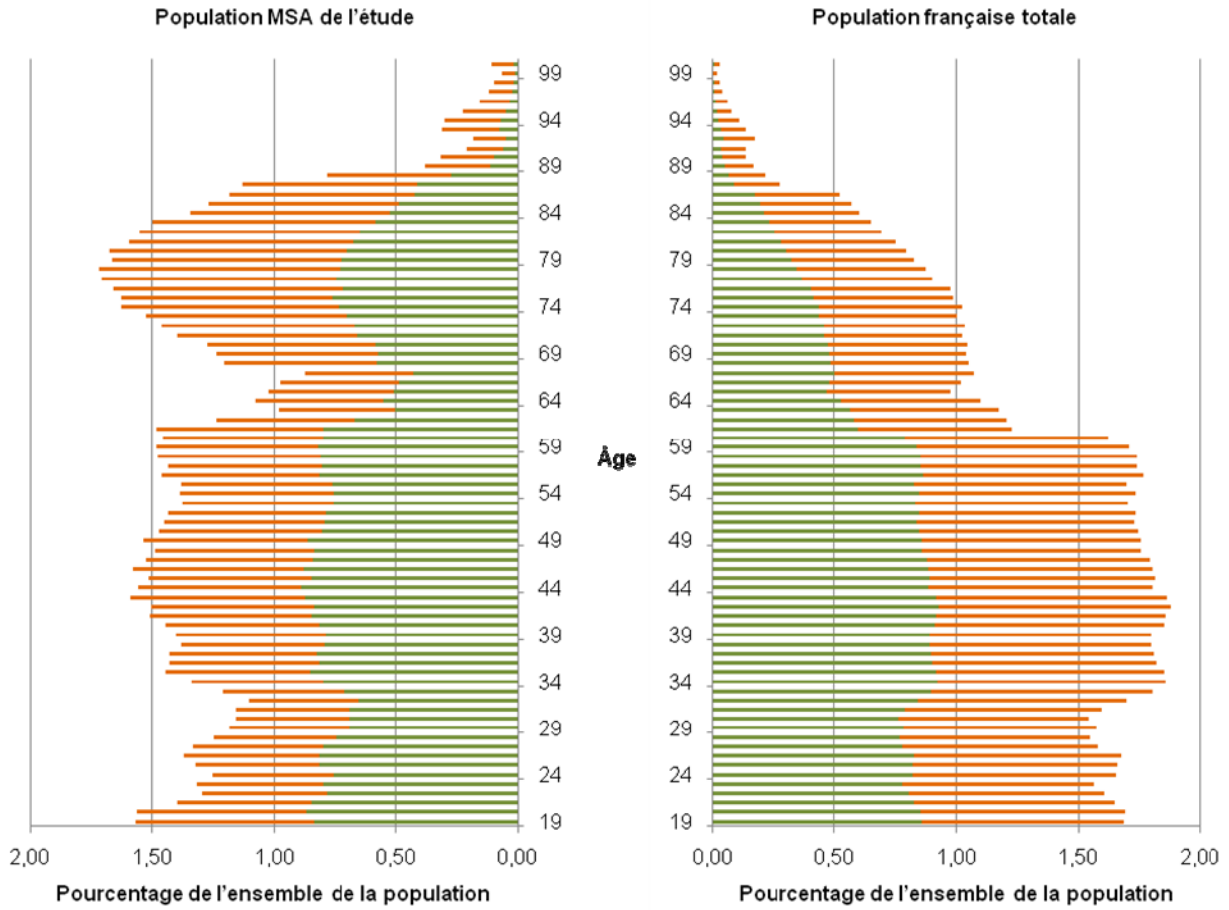
^bModèle développé à partir de 1 058 consommateurs de médicaments antiparkinsoniens (319 patients parkinsoniens et 739 patients non parkinsoniens).

^dCalculée pour le seuil optimal qui maximise le nombre d'individus correctement classés parmi les personnes ayant reçu au moins un médicament antiparkinsonien en 2007.

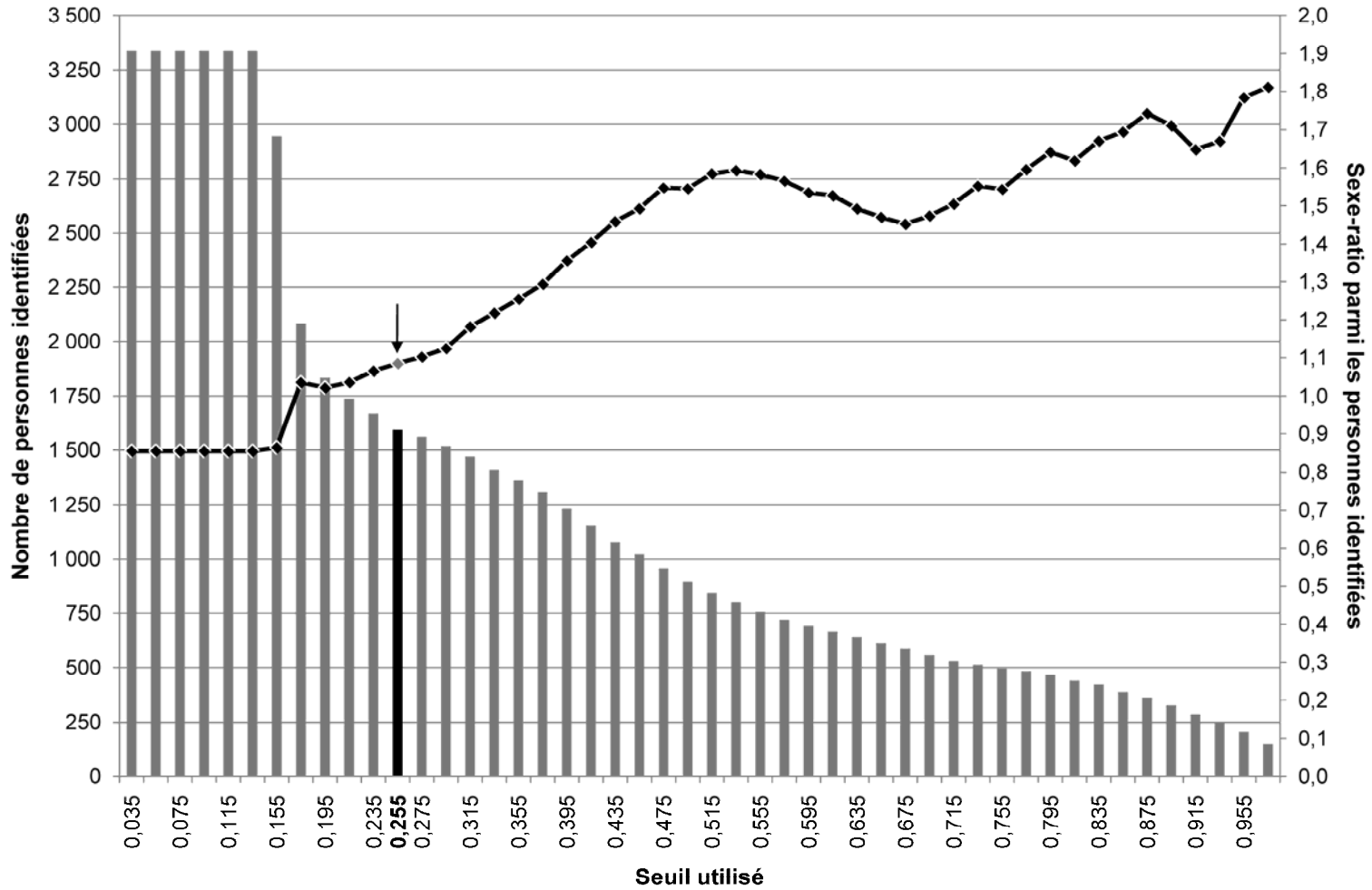
^eValidation interne réalisée par la méthode du bootstrap incluant l'étape de sélection.

^fAppelé calibration-in-the-large dans la littérature anglophone.

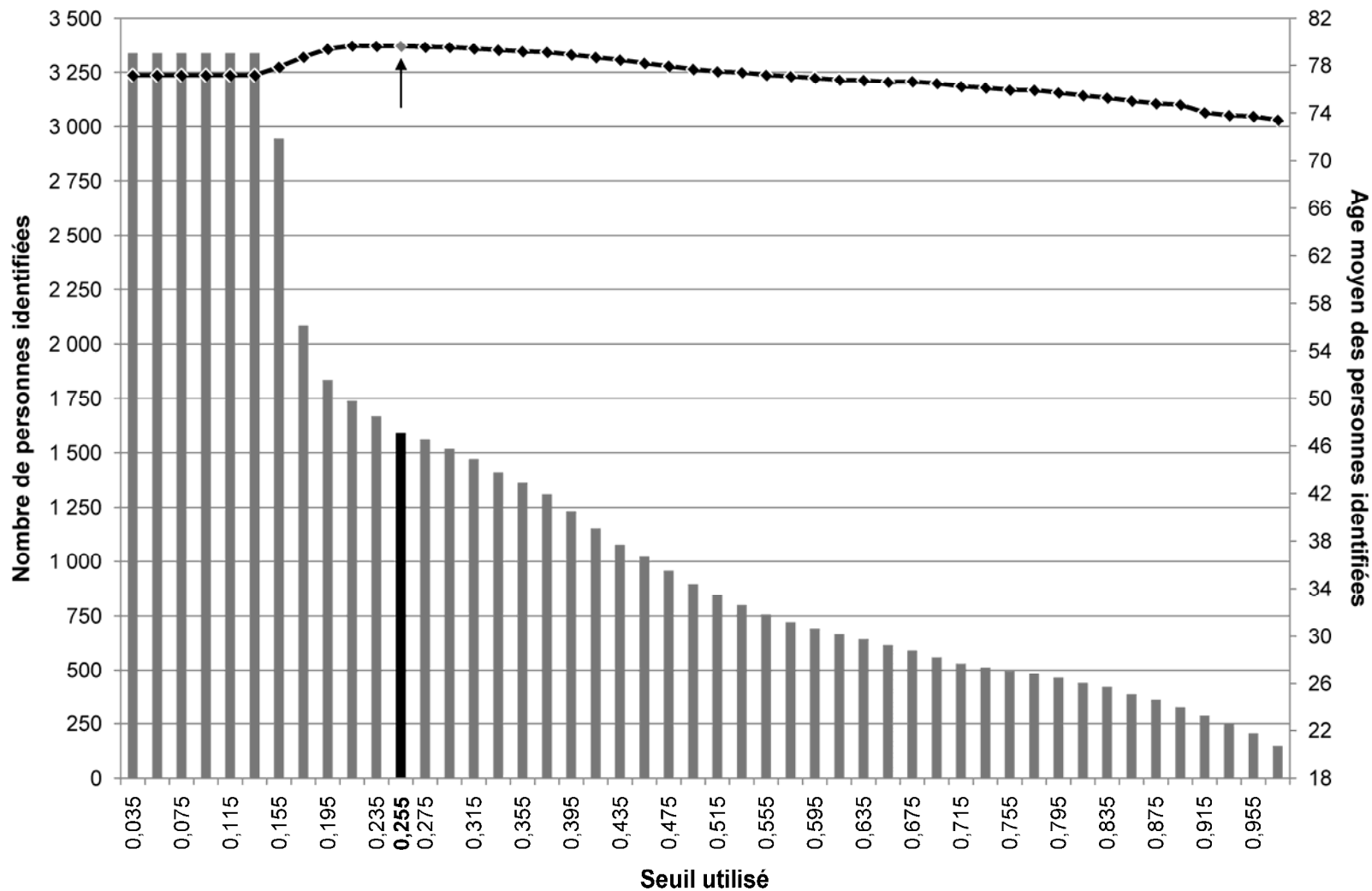
ANNEXE 9 - Structure d'âge de la population de la Mutualité sociale agricole (MSA) dans les cinq départements de l'étude et de l'ensemble de la population française majeure par sexe (hommes en vert et femmes en orange)



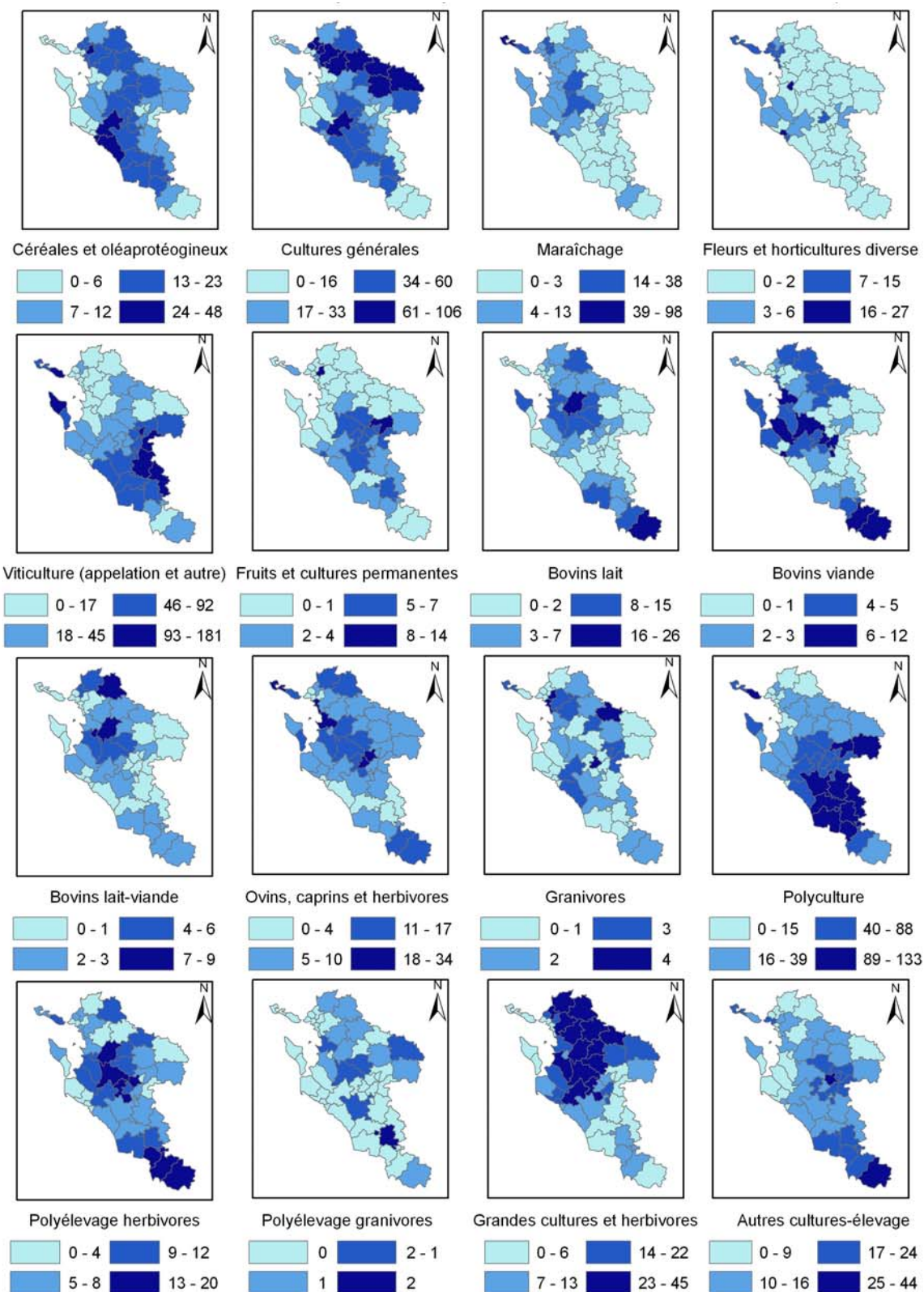
ANNEXE 10 - Evolution du nombre et du sexe-ratio (hommes/femmes) des personnes identifiées comme ayant la maladie de Parkinson en fonction du seuil utilisé lors de l'application du modèle prédictif (développé à partir des doses cumulées des médicaments antiparkinsoniens)



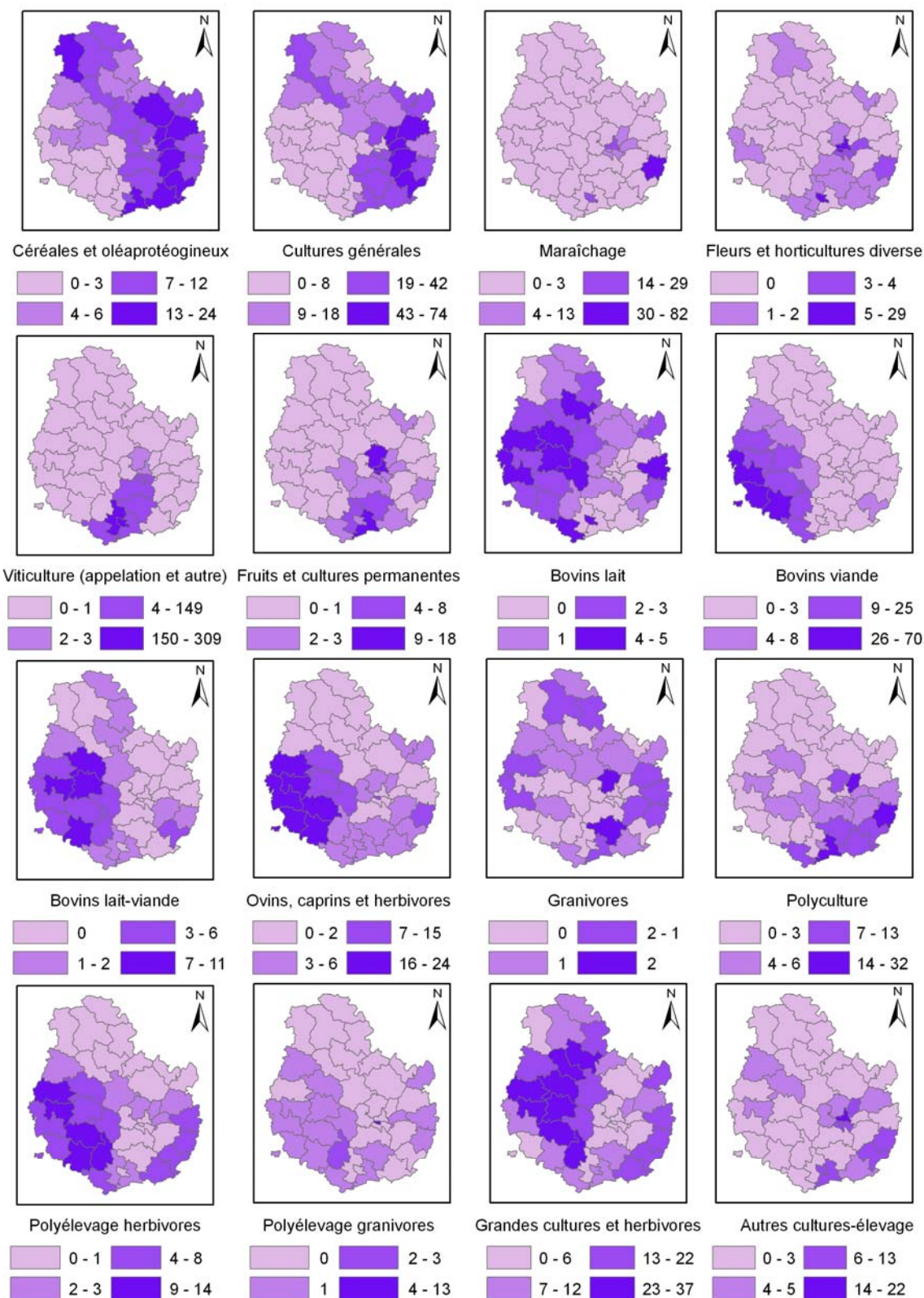
ANNEXE 11 - Evolution du nombre et de l'âge moyen des personnes identifiées comme ayant la maladie de Parkinson en fonction du seuil utilisé lors de l'application du modèle prédictif (développé à partir des doses cumulées des médicaments antiparkinsoniens)



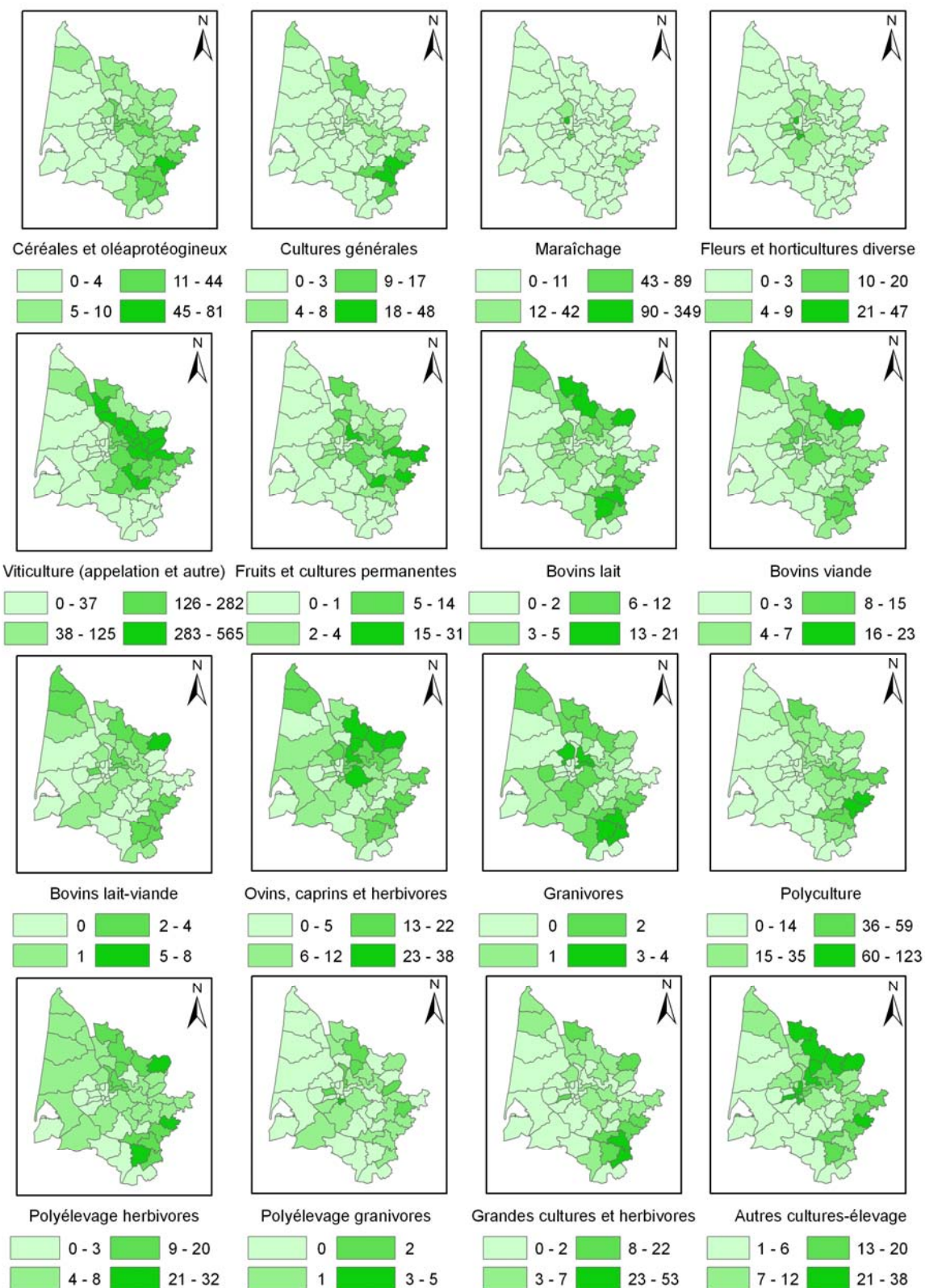
ANNEXE 12 - Distribution spatiale de la densité (nombre d'exploitations pour 100 km²) des 16 orientations technico-économiques (Otex) des exploitations en Charente-Maritime en 1988 (échelle : 1 cm=34 km)



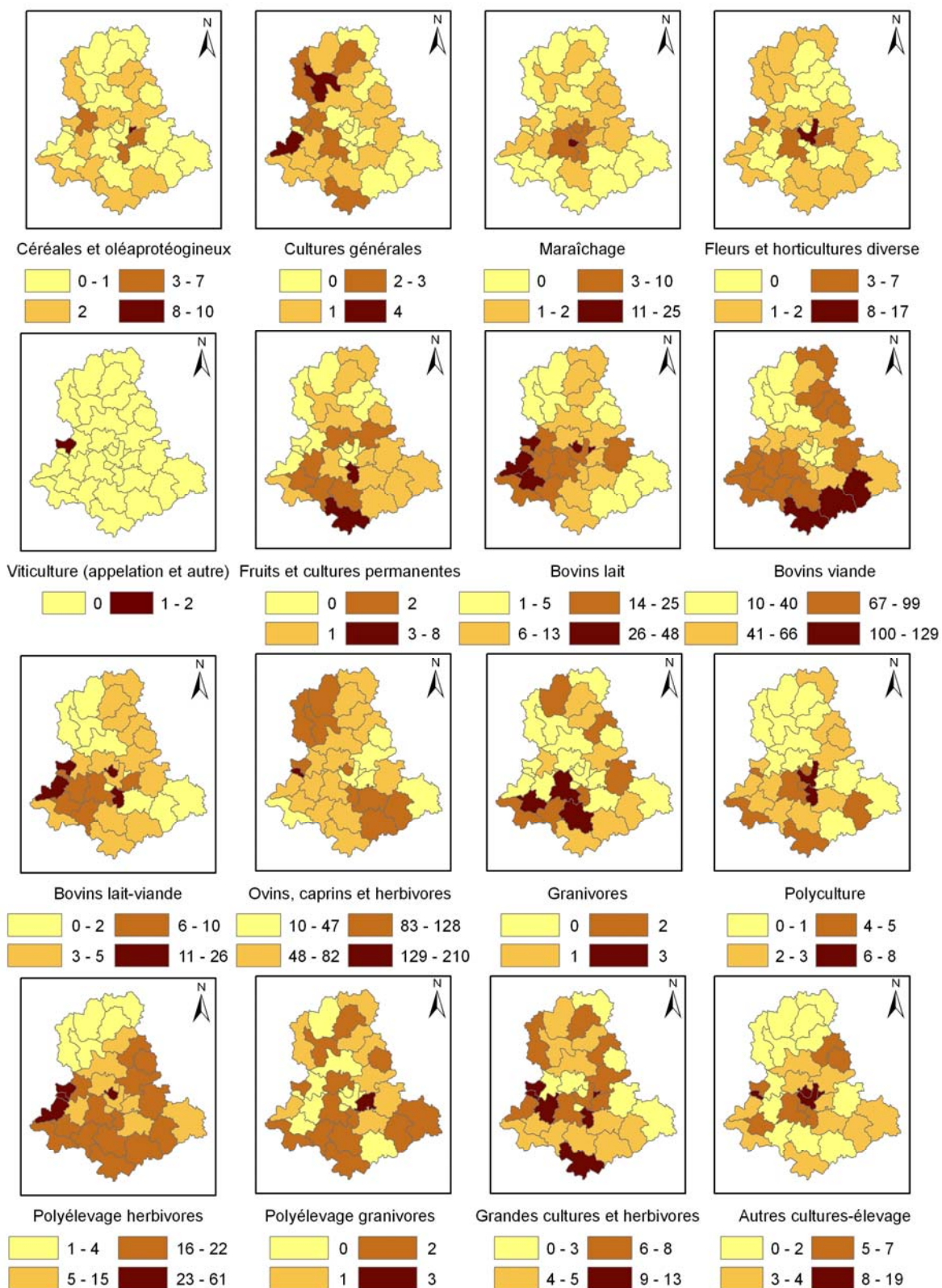
ANNEXE 13 - Distribution spatiale de la densité (nombre d'exploitations pour 100 km²) des 16 orientations technico-économiques (Otex) des exploitations en Côte-d'Or en 1988 (échelle : 1 cm=30 km)



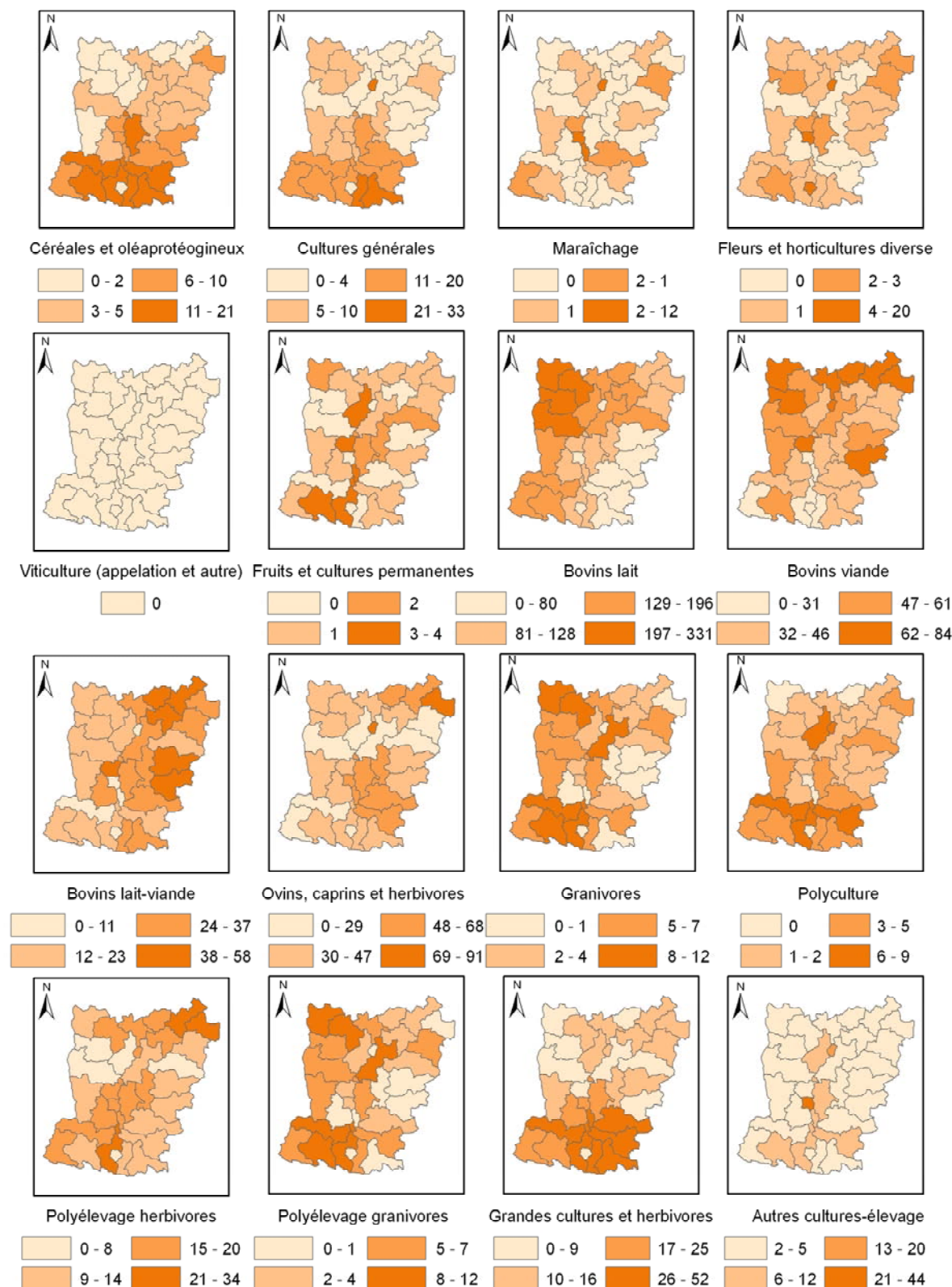
ANNEXE 14 - Distribution spatiale de la densité (nombre d'exploitations pour 100 km²) des 16 orientations technico-économiques (Otex) des exploitations en Gironde en 1988 (échelle : 1 cm=37 km)



ANNEXE 15 - Distribution spatiale de la densité (nombre d'exploitations pour 100 km²) des 16 orientations technico-économiques (Otex) des exploitations en Haute-Vienne en 1988 (échelle : 1 cm=26 km)



ANNEXE 16 - Distribution spatiale de la densité (nombre d'exploitations pour 100 km²) des 16 orientations technico-économiques (Otex) des exploitations en Mayenne en 1988 (échelle : 1 cm=24 km)



ANNEXE 17 - Relation entre la prévalence de la maladie de Parkinson en 2007 et la densité des orientations technico-économiques dans les cinq départements de l'étude (Charente-Maritime, Côte-d'Or, Gironde, Haute-Vienne, Mayenne) chez les hommes

Orientations technico-économiques des exploitations (Otex)	C	Densité (/100 km ²)	Cas parkinsoniens N (%)	Sujets non parkinsoniens N (%)	OR (IC 95 %) ^a	p ^b
Céréales et oléoprotéagineux	1	0-1	172 (20,75)	24 242 (19,83)	Référence	
	2	2-4	156 (18,82)	25 046 (20,49)	0,99 (0,78-1,25)	
	3	4-8	140 (16,89)	24 081 (19,70)	1,04 (0,80-1,36)	
	4	8-14	175 (21,11)	23 707 (19,39)	1,09 (0,84-1,42)	
	5	14-81	186 (22,44)	25 184 (20,60)	0,98 (0,75-1,29)	0,815
Cultures générales	1	0-0	116 (13,99)	24 638 (20,15)	Référence	
	2	0-2	181 (21,83)	24 161 (19,76)	1,02 (0,80-1,29)	
	3	2-8	162 (19,54)	24 541 (20,07)	0,86 (0,66-1,13)	
	4	9-26	148 (17,85)	23 887 (19,54)	0,83 (0,62-1,11)	
	5	26-106	222 (26,78)	25 033 (20,48)	1,11 (0,81-1,52)	0,153
Maraîchage	1	0-0	285 (34,38)	36 376 (29,75)	Référence	
	2	0-1	148 (17,85)	22 231 (18,18)	0,86 (0,70-1,06)	
	3	1-2	159 (19,18)	21 424 (17,52)	1,03 (0,81-1,30)	
	4	2-6	120 (14,48)	20 580 (16,83)	0,93 (0,72-1,21)	
	5	6-349	117 (14,11)	21 649 (17,71)	0,94 (0,74-1,20)	0,824
Fleurs et horticulture diverse	1	0-0	193 (23,28)	24 731 (20,23)	Référence	
	2	0-1	168 (20,27)	25 221 (20,63)	0,94 (0,76-1,17)	
	3	1-2	206 (24,85)	24 444 (19,99)	1,15 (0,92-1,42)	
	4	2-3	132 (15,92)	23 855 (19,51)	0,98 (0,77-1,25)	
	5	3-47	130 (15,68)	24 009 (19,64)	1,07 (0,84-1,35)	0,652
Viticulture (appellation et autre)	1	0-0	371 (44,75)	46 139 (37,74)	Référence	
	2	0-24	119 (14,35)	18 748 (15,33)	1,05 (0,75-1,48)	
	3	24-76	131 (15,80)	18 924 (15,48)	1,20 (0,83-1,73)	
	4	78-203	122 (14,72)	18 335 (15,00)	1,12 (0,81-1,56)	
	5	213-565	86 (10,37)	20 114 (16,45)	0,98 (0,68-1,41)	0,590
Fruits et autres cultures permanentes	1	0-0	140 (16,89)	24 460 (20,01)	Référence	
	2	0-1	158 (19,06)	25 163 (20,58)	1,14 (0,90-1,43)	
	3	1-2	201 (24,25)	24 210 (19,80)	1,15 (0,92-1,44)	
	4	2-4	165 (19,90)	24 112 (19,72)	1,17 (0,92-1,48)	
	5	4-31	165 (19,90)	24 315 (19,89)	1,40 (1,10-1,79)	0,009
Bovins lait	1	0-1	122 (14,72)	23 884 (19,54)	Référence	
	2	1-3	175 (21,11)	24 053 (19,67)	1,23 (0,97-1,56)	
	3	3-6	143 (17,25)	25 463 (20,83)	1,01 (0,79-1,30)	
	4	6-19	183 (22,07)	23 678 (19,37)	1,09 (0,85-1,40)	
	5	19-331	206 (24,85)	25 182 (20,60)	0,99 (0,69-1,43)	0,626

Suite Annexe 17

Orientations technico-économiques des exploitations (Otex)	C	Densité (/100 km²)	Cas parkinsoniens N (%)	Sujets non parkinsoniens N (%)	OR (IC 95 %) ^a	p ^b
Bovins viande	1	0-1	173 (20,87)	24 867 (20,34)	Référence	0,801
	2	1-4	132 (15,92)	24 609 (20,13)	0,79 (0,63-1,00)	
	3	4-7	137 (16,53)	23 702 (19,39)	0,86 (0,68-1,09)	
	4	8-39	144 (17,37)	22 689 (18,56)	0,81 (0,63-1,04)	
	5	40-129	243 (29,31)	26 393 (21,59)	0,91 (0,65-1,28)	
Bovins lait-viande	1	0-0	199 (24,00)	35 415 (28,97)	Référence	0,651
	2	0-1	122 (14,72)	22 034 (18,02)	1,03 (0,81-1,30)	
	3	1-2	143 (17,25)	20 891 (17,09)	0,98 (0,78-1,22)	
	4	2-9	174 (20,99)	21 318 (17,44)	1,04 (0,83-1,31)	
	5	9-58	191 (23,04)	22 602 (18,49)	0,89 (0,59-1,35)	
Ovins, caprins et autres herbivores	1	0-4	133 (16,04)	24 993 (20,44)	Référence	0,471
	2	4-8	166 (20,02)	24 116 (19,73)	1,06 (0,83-1,35)	
	3	8-14	129 (15,56)	23 694 (19,38)	0,98 (0,76-1,28)	
	4	14-34	166 (20,02)	23 756 (19,43)	1,06 (0,82-1,37)	
	5	34-210	235 (28,35)	25 701 (21,02)	1,15 (0,81-1,62)	
Granivores	1	0-0	189 (22,80)	33 548 (27,44)	Référence	0,259
	2	0-1	140 (16,89)	22 542 (18,44)	0,94 (0,75-1,18)	
	3	1-1	143 (17,25)	22 095 (18,07)	0,98 (0,78-1,22)	
	4	1-2	167 (20,14)	21 814 (17,84)	1,06 (0,85-1,32)	
	5	2-12	190 (22,92)	22 261 (18,21)	1,12 (0,87-1,43)	
Polyculture	1	0-1	178 (21,47)	25 070 (20,51)	Référence	0,601
	2	1-5	191 (23,04)	25 101 (20,53)	1,06 (0,86-1,31)	
	3	5-18	137 (16,53)	24 608 (20,13)	1,18 (0,91-1,54)	
	4	19-39	148 (17,85)	23 630 (19,33)	1,08 (0,79-1,47)	
	5	42-133	175 (21,11)	23 851 (19,51)	1,03 (0,74-1,42)	
Polyélevage à orientation herbivores	1	0-3	126 (15,20)	24 657 (20,17)	Référence	0,676
	2	3-5	146 (17,61)	24 809 (20,29)	1,09 (0,84-1,41)	
	3	5-9	167 (20,14)	24 239 (19,83)	0,99 (0,77-1,28)	
	4	10-15	190 (22,92)	24 335 (19,90)	1,05 (0,81-1,36)	
	5	15-61	200 (24,13)	24 220 (19,81)	0,98 (0,76-1,27)	
Polyélevage à orientation granivores	1	0-0	304 (36,67)	51 843 (42,40)	Référence	0,276
	2	0-1	125 (15,08)	17 648 (14,43)	1,06 (0,86-1,31)	
	3	1-1	109 (13,15)	17 265 (14,12)	1,01 (0,81-1,27)	
	4	1-4	129 (15,56)	17 671 (14,45)	1,01 (0,80-1,28)	
	5	4-27	162 (19,54)	17 833 (14,59)	1,32 (0,80-2,17)	

Suite Annexe 17

Orientations technico-économiques des exploitations (Otex)	C	Densité (/100 km²)	Cas parkinsoniens N (%)	Sujets non parkinsoniens N (%)	OR (IC 95 %) ^a	p ^b
Grandes cultures et herbivores	1	0-2	109 (13,15)	24 181 (19,78)	Référence	
	2	2-4	138 (16,65)	24 354 (19,92)	0,88 (0,67-1,14)	
	3	4-8	206 (24,85)	24 352 (19,92)	0,99 (0,76-1,31)	
	4	8-16	175 (21,11)	24 351 (19,92)	0,90 (0,68-1,18)	
	5	16-53	201 (24,25)	25 022 (20,47)	0,95 (0,71-1,26)	0,999
Autres combinaisons cultures-élevage	1	0-3	179 (21,59)	25 221 (20,63)	Référence	
	2	3-5	176 (21,23)	24 397 (19,96)	1,05 (0,83-1,31)	
	3	5-10	168 (20,27)	25 052 (20,49)	1,38 (1,08-1,77)	
	4	10-16	161 (19,42)	23 952 (19,59)	1,17 (0,89-1,55)	
	5	16-44	145 (17,49)	23 638 (19,33)	1,14 (0,87-1,50)	0,907

Abréviations : IC : intervalle de confiance ; OR : odds ratio ; C : catégorie.

Les catégories d'exposition sont définies à partir des quintiles de la distribution des densités sauf pour les Otex suivantes: maraîchage, viticulture, bovin lait-viande, granivore, polyélevage à orientation granivore. Pour ces dernières, la catégorie de référence correspond aux cantons sans ce type d'exploitation et les catégories suivantes sont définies à partir des quartiles de la distribution des densités.

^a*Les OR et IC sont ajustés sur l'âge (terme linéaire et quadratique), le département et le revenu médian des ménages du canton.*

^b*Test de tendance des OR.*

ANNEXE 18 - Relation entre la prévalence de la maladie de Parkinson en 2007 et la densité des orientations technico-économiques dans les cinq départements de l'étude (Charente-Maritime, Côte-d'Or, Gironde, Haute-Vienne, Mayenne) chez les femmes

Orientations technico-économiques des exploitations (Otex)	C	Densité (/100 km ²)	Cas parkinsoniens N (%)	Sujets non parkinsoniens N (%)	OR (IC 95 %) ^a	p ^b
Céréales et oléoprotéagineux	1	0-1	166 (21,73)	22 297 (19,34)	Référence	
	2	2-4	147 (19,24)	22 938 (19,90)	1,04 (0,81-1,33)	
	3	4-8	135 (17,67)	23 832 (20,68)	1,12 (0,85-1,46)	
	4	8-14	151 (19,76)	22 859 (19,83)	1,13 (0,86-1,48)	
	5	14-81	165 (21,60)	23 341 (20,25)	1,12 (0,85-1,47)	0,532
Cultures générales	1	0-0	119 (15,58)	22 718 (19,71)	Référence	
	2	0-2	159 (20,81)	22 892 (19,86)	0,98 (0,77-1,25)	
	3	2-8	151 (19,76)	23 334 (20,24)	0,88 (0,67-1,15)	
	4	9-26	168 (21,99)	22 926 (19,89)	1,03 (0,77-1,37)	
	5	26-106	167 (21,86)	23 397 (20,30)	0,96 (0,69-1,32)	0,980
Maraîchage	1	0-0	230 (30,10)	31 841 (27,62)	Référence	
	2	0-1	147 (19,24)	19 826 (17,20)	1,10 (0,89-1,36)	
	3	1-2	139 (18,19)	20 722 (17,98)	1,25 (0,98-1,60)	
	4	2-6	139 (18,19)	21 601 (18,74)	1,39 (1,07-1,80)	
	5	6-349	109 (14,27)	21 277 (18,46)	1,05 (0,82-1,34)	0,714
Fleurs et horticulture diverse	1	0-0	162 (21,20)	21 521 (18,67)	Référence	
	2	0-1	181 (23,69)	22 545 (19,56)	1,18 (0,95-1,46)	
	3	1-2	143 (18,72)	23 774 (20,63)	0,88 (0,69-1,12)	
	4	2-3	135 (17,67)	23 234 (20,16)	1,10 (0,85-1,41)	
	5	3-47	143 (18,72)	24 193 (20,99)	1,10 (0,86-1,39)	0,491
Viticulture (appellation et autre)	1	0-0	348 (45,55)	40 819 (35,41)	Référence	
	2	0-24	92 (12,04)	17 394 (15,09)	0,95 (0,65-1,37)	
	3	24-76	136 (17,80)	19 002 (16,49)	1,43 (0,97-2,10)	
	4	78-203	111 (14,53)	18 438 (16,00)	1,23 (0,86-1,75)	
	5	213-565	77 (10,08)	19 614 (17,02)	1,00 (0,68-1,47)	0,730
Fruits et autres cultures permanentes	1	0-0	151 (19,76)	22 651 (19,65)	Référence	
	2	0-1	150 (19,63)	22 420 (19,45)	1,12 (0,89-1,41)	
	3	1-2	162 (21,20)	23 008 (19,96)	0,90 (0,71-1,13)	
	4	2-4	166 (21,73)	23 319 (20,23)	1,24 (0,99-1,56)	
	5	4-31	135 (17,67)	23 869 (20,71)	1,14 (0,89-1,47)	0,166
Bovins lait	1	0-1	103 (13,48)	22 814 (19,79)	Référence	
	2	1-3	151 (19,76)	21 800 (18,91)	1,42 (1,10-1,84)	
	3	3-6	128 (16,75)	23 943 (20,77)	1,11 (0,85-1,45)	
	4	6-19	181 (23,69)	23 322 (20,23)	1,31 (1,01-1,70)	
	5	19-331	201 (26,31)	23 388 (20,29)	1,05 (0,73-1,52)	0,514

Suite Annexe 18

Orientations technico-économiques des exploitations (Otex)	C	Densité (/100 km²)	Cas parkinsoniens N (%)	Sujets non parkinsoniens N (%)	OR (IC 95 %) ^a	p ^b
Bovins viande	1	0-1	137 (17,93)	22 506 (19,53)	Référence	0,582
	2	1-4	133 (17,41)	22 894 (19,86)	1,08 (0,85-1,38)	
	3	4-7	112 (14,66)	23 947 (20,78)	0,85 (0,65-1,10)	
	4	8-39	150 (19,63)	21 842 (18,95)	1,03 (0,79-1,34)	
	5	40-129	232 (30,37)	24 078 (20,89)	1,11 (0,78-1,56)	
Bovins lait-viande	1	0-0	179 (23,43)	34 092 (29,58)	Référence	0,111
	2	0-1	116 (15,18)	19 594 (17,00)	1,19 (0,93-1,52)	
	3	1-2	133 (17,41)	21 147 (18,35)	1,04 (0,82-1,30)	
	4	2-9	159 (20,81)	19 886 (17,25)	1,12 (0,88-1,43)	
	5	9-58	177 (23,17)	20 548 (17,83)	0,70 (0,45-1,07)	
Ovins, caprins et autres herbivores	1	0-4	114 (14,92)	21 933 (19,03)	Référence	0,890
	2	4-8	150 (19,63)	23 408 (20,31)	1,12 (0,87-1,46)	
	3	8-14	114 (14,92)	23 158 (20,09)	0,98 (0,74-1,29)	
	4	14-34	166 (21,73)	22 951 (19,91)	1,06 (0,81-1,39)	
	5	34-210	220 (28,80)	23 817 (20,66)	1,02 (0,72-1,44)	
Granivores	1	0-0	182 (23,82)	31 901 (27,68)	Référence	0,384
	2	0-1	121 (15,84)	20 471 (17,76)	0,98 (0,77-1,23)	
	3	1-1	128 (16,75)	20 860 (18,10)	0,98 (0,78-1,23)	
	4	1-2	160 (20,94)	21 186 (18,38)	1,15 (0,92-1,44)	
	5	2-12	173 (22,64)	20 849 (18,09)	1,09 (0,85-1,41)	
Polyculture	1	0-1	168 (21,99)	22 272 (19,32)	Référence	0,094
	2	1-5	174 (22,77)	21 850 (18,96)	1,07 (0,86-1,33)	
	3	5-18	120 (15,71)	22 648 (19,65)	1,18 (0,90-1,55)	
	4	19-39	129 (16,88)	24 332 (21,11)	1,09 (0,79-1,50)	
	5	42-133	173 (22,64)	24 165 (20,96)	1,32 (0,95-1,84)	
Polyélevage à orientation herbivores	1	0-3	118 (15,45)	22 427 (19,46)	Référence	0,767
	2	3-5	130 (17,02)	22 730 (19,72)	1,16 (0,89-1,51)	
	3	5-9	132 (17,28)	22 771 (19,76)	0,90 (0,69-1,17)	
	4	10-15	179 (23,43)	23 223 (20,15)	1,03 (0,79-1,33)	
	5	15-61	205 (26,83)	24 116 (20,92)	1,07 (0,82-1,38)	
Polyélevage à orientation granivores	1	0-0	292 (38,22)	49 492 (42,94)	Référence	0,851
	2	0-1	100 (13,09)	16 046 (13,92)	0,97 (0,77-1,22)	
	3	1-1	98 (12,83)	16 716 (14,50)	1,01 (0,80-1,28)	
	4	1-4	121 (15,84)	16 763 (14,54)	0,97 (0,76-1,24)	
	5	4-27	153 (20,03)	16 250 (14,10)	1,04 (0,65-1,66)	

Suite Annexe 18

Orientations technico-économiques des exploitations (Otex)	C	Densité (/100 km ²)	Cas parkinsoniens	Sujets non parkinsoniens	OR (IC 95 %) ^a	p ^b
			N (%)	N (%)		
Grandes cultures et herbivores	1	0-2	90 (11,78)	22 864 (19,84)	Référence	
	2	2-4	150 (19,63)	23 328 (20,24)	1,33 (1,01-1,74)	
	3	4-8	173 (22,64)	23 475 (20,37)	1,10 (0,82-1,47)	
	4	8-16	182 (23,82)	22 910 (19,88)	1,25 (0,94-1,66)	
	5	16-53	169 (22,12)	22 690 (19,68)	1,14 (0,85-1,54)	0,915
Autres combinaisons cultures-élevage	1	0-3	177 (23,17)	21 447 (18,61)	Référence	
	2	3-5	163 (21,34)	22 597 (19,60)	0,86 (0,68-1,09)	
	3	5-10	129 (16,88)	23 429 (20,33)	0,99 (0,77-1,29)	
	4	10-16	143 (18,72)	23 029 (19,98)	0,93 (0,71-1,23)	
	5	16-44	152 (19,90)	24 765 (21,48)	1,01 (0,77-1,32)	0,694

Abréviations : IC : intervalle de confiance ; OR : odds ratio ; C : catégorie.

Les catégories d'exposition sont définies à partir des quintiles de la distribution des densités sauf pour les Otex suivantes: maraîchage, viticulture, bovin lait-viande, granivore, polyélevage à orientation granivore. Pour ces dernières, la catégorie de référence correspond aux cantons sans ce type d'exploitation et les catégories suivantes sont définies à partir des quartiles de la distribution des densités.

^a*Les OR et IC sont ajustés sur l'âge (terme linéaire et quadratique), le département et le revenu médian des ménages du canton.*

^b*Test de tendance des OR.*

Maladie de Parkinson parmi les affiliés à la Mutualité sociale agricole dans cinq départements

Modèles prédictifs à partir des bases de remboursement de médicaments

Les bases de remboursement de médicaments de l'Assurance maladie représentent une opportunité pour la surveillance épidémiologique en permettant d'identifier des patients atteints de certaines maladies à partir de leur consommation de médicaments. Avant d'envisager de les utiliser pour cet objectif, il est nécessaire d'évaluer leur validité. Nous avons étudié leur validité pour identifier des patients traités pour maladie de Parkinson dans cinq caisses de Mutualité sociale agricole en 2007. Nous avons comparé les profils de consommation de médicaments antiparkinsoniens chez des patients dont le diagnostic de maladie de Parkinson a été confirmé par un neurologue et chez des patients traités pour une autre raison. Nous avons développé différents modèles prédictifs de maladie de Parkinson permettant d'estimer la probabilité qu'ont les consommateurs d'antiparkinsoniens d'être traités pour cette maladie. Parmi 1 114 participants ayant consommé un médicament antiparkinsonien en 2007, un diagnostic de maladie de Parkinson a été confirmé pour 320 (29 %) d'entre eux tandis que 794 (71 %) étaient traités pour une autre raison. Différents modèles ont été comparés en termes de qualité des prédictions. Le modèle logistique incluant les doses cumulées des différents types d'antiparkinsoniens et la régularité du traitement est caractérisé par de bonnes performances (statistique $c=0,953$; sensibilité=92,5 % ; spécificité=86,4 %). D'après ce modèle, la prévalence de la maladie de Parkinson parmi les personnes âgées de 18 ans et plus est de 6,66/1 000 ; la correction par la sensibilité et la spécificité du modèle conduit à une estimation proche (6,01/1 000). Nous présentons également un exemple d'étude de la relation entre la prévalence ainsi estimée et une exposition définie au niveau géographique. Nous discutons les avantages et inconvénients des bases de remboursement de médicaments pour surveiller la fréquence de certaines maladies et étudier leurs facteurs de risque. Nos résultats montrent que, pour certaines pathologies, un système de surveillance à partir des bases de remboursement de médicaments, et sans recueil d'informations supplémentaires auprès des sujets, pourrait être mis en place au niveau national.

Mots clés : maladie Parkinson, surveillance épidémiologique, prévalence, Mutualité sociale agricole, modèle prédictif, bases administratives

Parkinson's disease among affiliates of the Mutualité sociale agricole in five French districts

Prediction model based on drug claims

Drug claims databases represent an opportunity for epidemiologic surveillance because they contribute to identifying patients with certain diseases based on their drug use. Before using them for this purpose, it is necessary to assess their validity. We investigated their validity to identify patients treated for Parkinson's disease among members of the Mutualité sociale agricole in five French districts in 2007. We compared patterns of antiparkinsonian drugs use in patients for whom a neurologist confirmed the diagnosis and in patients treated for another reason, and developed prediction models in order to estimate the probability that antiparkinsonian drug users were treated for Parkinson's disease. Among 1,114 persons who used antiparkinsonian drugs in 2007, 320 (29%) were considered to have Parkinson's disease, while 794 (71%) were treated for another reason. Different models were compared in terms of quality of predictions. A logistic model including cumulative doses of different antiparkinsonian drugs and regularity of treatment displayed good performance (c statistic=0.953, sensitivity=92.5%, specificity=86.4%). According to this model, the prevalence of Parkinson's disease among persons aged 18 years old and over was 6.66/1,000; a correction of this figure using sensitivity and specificity estimates of the model led to a similar estimate (6.01/1,000). We also present an example of a study that examined the relationship between prevalence estimated through this prediction model and an exposure defined geographically. We discuss the advantages and disadvantages of drug claims databases to monitor the frequency of certain diseases and study their risk factors. Our results show that, for some conditions, monitoring systems based on drug claims databases could be implemented at the national level without gathering additional individual data.

Citation suggérée :

Moisan F, Elbaz A. Maladie de Parkinson parmi les affiliés à la Mutualité sociale agricole dans cinq départements. Modèles prédictifs à partir des bases de remboursement de médicaments. Saint-Maurice: Institut de veille sanitaire; 2012. 66 pages. Disponible à partir de l'URL: <http://www.invs.sante.fr>