THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 605
Biologie Santé
Spécialité : Epidémiologie, Analyse de Risque, Recherche Clinique

Par
Jonathan ROUX

Parcours de soins des patients atteints de sclérose en plaques à partir des données médico-administratives en France

Thèse présentée et soutenue à Rennes, le 22 novembre 2018
Thèse préparée dans le cadre du Réseau doctoral en santé publique animé par l'EHESP
Unités de recherche : Univ Rennes, EHESP, REPERES (Recherche en Pharmaco-Epidémiologie et Recours aux Soins) – EA 7449 & Inserm CIC-P 1414 (Centre d'Investigation Clinique Plurithématique), Centre Hospitalier Universitaire de Rennes

Rapporteurs avant soutenance :

Nicolas SAVY	Maître de conférences, Institut Mathématiques de Toulouse
Matthias STUDER	Senior researcher, Université de Genève

Composition du Jury :

Président :	Michel CLANET	Professeur des Universités – Praticien Hospitalier, CHU de Toulouse
Examinateurs :	Christine BINQUET	Professeur des Universités – Praticien Hospitalier, Inserm CIC 1432
	Laure COM-RUELLE	Directeur de recherche, Irdes
	Alexis ELBAZ	Directeur de recherche, Inserm Unité 1018
Directrice de thèse :	Emmanuelle LERAY	Professeur de l'EHESP, EA 7449 REPERES, Inserm CIC-P 1414
Co-directeur de thèse :	Olivier GRIMAUD	Professeur de l'EHESP, EA 7449 REPERES

THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 605
Biologie Santé
Spécialité : Epidémiologie, Analyse de Risque, Recherche Clinique

Par
Jonathan ROUX

Parcours de soins des patients atteints de sclérose en plaques à partir des données médico-administratives en France

Thèse présentée et soutenue à Rennes, le 22 novembre 2018
Thèse préparée dans le cadre du Réseau doctoral en santé publique animé par l'EHESP
Unités de recherche : Univ Rennes, EHESP, REPERES (Recherche en Pharmaco-Epidémiologie et Recours aux Soins) – EA 7449 & Inserm CIC-P 1414 (Centre d'Investigation Clinique Plurithématique), Centre Hospitalier Universitaire de Rennes

Rapporteurs avant soutenance :

Nicolas SAVY	Maître de conférences, Institut Mathématiques de Toulouse
Matthias STUDER	Senior researcher, Université de Genève

Composition du Jury :

Président :	Michel CLANET	Professeur des Universités – Praticien Hospitalier, CHU de Toulouse
Examinateurs :	Christine BINQUET	Professeur des Universités – Praticien Hospitalier, Inserm CIC 1432
	Laure COM-RUELLE	Directeur de recherche, Irdes
	Alexis ELBAZ	Directeur de recherche, Inserm Unité 1018
Directrice de thèse :	Emmanuelle LERAY	Professeur de l'EHESP, EA 7449 REPERES, Inserm CIC-P 1414
Co-directeur de thèse :	Olivier GRIMAUD	Professeur de l'EHESP, EA 7449 REPERES

“If you can not measure it, you can not improve it.”

William Thomson (Lord Kelvin) - 1883

Remerciements

Que serait une thèse sans remerciements aux personnes m'ayant soutenu tout au long de ce travail de trois ans.

Je souhaite commencer par mes deux encadrants, Emmanuelle Leray et Olivier Grimaud. Emmanuelle, sans toi l'ensemble de cette thèse n'aurait pas été possible, je te remercie pour m'avoir accepté en stage il y a maintenant trois ans et demi et pour m'avoir permis de continuer jusqu'à aujourd'hui. Grâce à toi, j'ai pu découvrir cette thématique passionnante qu'est la SEP et aussi la France (heureusement que sur la fin il y a eu un peu de congrès à l'international ;-)). J'ai également eu la chance d'avoir un projet avec un acronyme qui tenait encore la route avant la dérive, même si je suis quand même jaloux d'EMOJI. Je suis fier d'être officiellement ton premier doctorant, et je suis sûr que je ne suis pas le dernier, mais ça on en reparle demain…. Olivier, je souhaite te remercier pour l'ensemble de tes conseils et pour ta vision globale de la santé publique qui a été très utile lors de tes relectures attentives tout au long de ce travail. Merci également de m'avoir accueilli au sein du département

Je remercie M. Nicolas Savy et M. Matthias Studer de me faire l'honneur d'être rapporteurs de ce travail de thèse. Vos connaissances et vos commentaires me seront d'une grande utilité dans l'enrichissement de ce travail. Merci également à Mme Christine Binquet, Mme Laure Com-Ruelle, M. Michel Clanet et M. Alexis Elbaz d'avoir accepté d'être examinateurs de cette thèse. Vos connaissances de la sclérose en plaques, des parcours de soins, des données et des statistiques seront source de discussions fructueuses lors de la soutenance.

Je désire remercier les membres constitutifs de mon comité de thèse, M. Marc Debouverie, M. Yohann Foucher, M. Christophe Hitte et M. Alain Weill pour leurs remarques et leurs conseils au cours de mon suivi. M. Hitte je vous remercie également pour le temps que vous m'avez consacré quand j'en ai eu besoin.

Je souhaite remercier Emmanuel Oger, co-directeur de l'équipe d'accueil REPERES et directeur de la plate-forme PEPS. Merci d'avoir financé cette thèse et pour l'ensemble des discussions au cours des réunions de l'équipe. J'en profite pour remercier l'ensemble des membres de REPERES pour les échanges scientifiques, les galettes des rois et les virées lors des congrès.

Le département METIS a été un environnement propice au travail (et à la gourmandise) grâce à l'ensemble des personnes le composant. Merci pour les discussions dans le couloir, les bureaux et à la tisanerie où il y avait toujours quelque chose à manger ! Merci Nadine, grâce à toi je sais où se trouvent tous les services de l'école. A tous les stagiaires (Estelle, Maëliss, Hugo, Léo, François,

Anaïs, Asma, Marine, Camille, Léa, Maxime, Russel et ceux que j'aurai oubliés) et internes (Marion, Astrid, Faustine, Aurélie, Soline, Camille) qui sont venus dans le département, j'ai été très heureux de vous rencontrer pendant deux, trois, six mois ou plus longtemps. Un petit clin d'œil à la DRI qui a partagé le même couloir et qui a été très prolifique en pâtisseries.

Merci au Réseau Doctoral en santé publique, et plus particulièrement à Karine Laboux et Emmanuelle Guevara, et par extension Christophe Le Rat. Merci également à l'ensemble des doctorants que j'ai pu rencontrer durant ces trois années un peu partout en France à Paris, Rennes et Bordeaux. Annabelle, Gloria, Mélanie, Narimane et Solenn, on a fait un super SID ensemble ! J'ai été très heureux d'être membre de ce réseau et de faire toutes ces rencontres (Anne-Claire, Noriane, Samuel, Vincent, et tant d'autres !) en particulier avec Ph'Doc, l'asso des doctorants.

Je tiens à remercier Helen Tremlett et Elaine Kingwell pour m'avoir accueilli à l'autre bout du monde à Vancouver durant ma mobilité. Merci également aux membres de l'équipe PiMS pour m'avoir aidé à appréhender les données mises à disposition par Population Data BC et BC Data Stewards. It was just amazing! Mes remerciements vont également à REPERES, à l'Ecole des Docteurs de l'Université Bretagne Loire et la région Bretagne et au Réseau Doctoral pour m'avoir permis de réaliser cette mobilité.

Mathilde et Pascal, vous avez été toujours été présents pour un afterwork jusque très tard mais c'est aussi savoir décompresser la thèse ! Merci Pascal pour l'expérience inoubliable qu'ont été les cours IDEA durant les trois années. Et je n'oublie pas qu'il n'y a pas de post-doc sans doctorat. Yann, ça fait trois ans et demi qu'on est en coloc' dans le bureau et je pense qu'on arrive enfin à nous différencier !

Une énorme pensée à la team de la fête, Alice, Lola et Mathilde, qui a été d'un grand soutien notamment à la fin de la rédaction et pour le pédalo. Je pense que vous avez beaucoup de photos compromettantes me concernant acquises au cours de ces années. Lola, tu feras toujours partie de la MS team élargie, même si tu ne travailles pas sur la SEP. On s'est bien amusé (et on a surtout bien mangé !) à Lyon et je suis sûr qu'on a créé le tube de cet automne (Papayooo, Papayooo). Alice (Hi Hue), merci pour ton expérience de la thèse et pour notre petite virée à Palavas. Je ne sais pas si on a fait un encadrement de ouf mais ça a été une super expérience de le faire avec toi. Enfin, Régine, ou plutôt Mathilde, j'attends toujours notre soirée DBK, j'espère qu'elle aura lieu un jour ! On s'est tapé de bonnes barres, bichette, typiquement quand t'as eu tes chaussettes toutes mouillées ;-) Je n'oublierai jamais notre rencontre avec la voix de Quick et avec ce magnifique mannequin. Lola et Mathilde plus qu'un an ! Mais ne vous mettez pas la rate au court-bouillon ça va bien se passer !

A tous mes amis, Laura, Caroline, Jérémy, Delphine, Ophélie, Basile, Cédric, vous en avez entendu parler de loin ou de près de cette thèse et vous avez toujours été là pour les vacances et m'aider à

décompresser. Un merci spécial pour Thomas qui m'a accueilli tant de fois sur Paris dans son petit appartement, et pour nos fameuses visites culturelles du samedi.

Un grand merci à l'ensemble de ma famille, qui me soutient depuis toujours. Je vous parle de SEP depuis longtemps déjà et vous allez enfin pouvoir comprendre ce que je fais à Rennes depuis tout ce temps, par contre je ne serai jamais médecin ! Et ne vous inquiétez pas, vous serez toujours les bienvenus pour un weekend à Rennes. Guillaume, bientôt sous-directeur ! J'aurais quand même voulu que ce soit possible d'avoir une machine à McFlurry pour ma soutenance. A mes parents, merci pour votre soutien sans faille depuis le début de cette thèse et depuis le début de mes études, surtout que ça n'a pas toujours été facile. Je vais enfin commencer à travailler !

Enfin, à tous ceux que j'aurais oubliés dans ces remerciements sans le vouloir, merci. Et comme je le dis toujours, *food and science*!

Cette thèse a été financée par l'Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM) au travers du consortium de Pharmaco-Epidémiologie des Produits de Santé (PEPS) coordonné par le Pr. Emmanuel Oger.

Valorisations scientifiques

Articles

- *Articles liés à la thèse*

Roux J, Grimaud O, Leray E. 2018. Use of state sequence analysis for care pathway analysis: the example of multiple sclerosis. Statistical Methods in Medical Research. DOI: 10.1177/0962280218772068.

Roux J, Kingwell E, Tremlett H, Leray E [Liste des auteurs provisoire]. Care consumption of people with multiple sclerosis in Canada: an analysis of administrative data from British Columbia, Canada, using sequence analysis. [En préparation]

Roux J, Guilleux A, Lefort M, Leray E. Access to health care and use of disease-modifying treatments in multiple sclerosis over 2010-2015 in France: a nationwide population-based study using health administrative data. [En préparation]

- *Articles non liés à la thèse*

Jeanjean M, Bind, M-A, Roux J, Ongagna J-C, de Sèze J, Bard D, Leray E. 2018. Ozone, NO2 and PM10 are associated with the occurrence of multiple sclerosis relapses. Evidence from seasonal multi-pollutant analyses. Environmental Research 163, 43-52.

Roux J, Bard D, Le Pabic E, Segala C, Reis J, Ongagna J-C, de Sèze J, Leray E. 2017. Air pollution by particulate matter PM10 may trigger multiple sclerosis relapses. Environmental Research 156, 404–410.

Communications

- *Orales*

Roux J, Grimaud O, Leray E. L'analyse de séquences multidimensionnelle : une méthode innovante pour l'étude des parcours de soins – Application à la sclérose en plaques à partir des données du Système National des Données de Santé (SNDS). Congrès Biographie Santé. Paris, France, 18-19 octobre 2018.

Roux J, Grimaud O, Leray E. Étude des parcours de soins des patients ayant une sclérose en plaques en France sur la période 2007-2013 : Une analyse des bases médico-administratives utilisant la

méthode « multichannel sequence analysis ». ADELF-EMOIS Joint congress. Montpellier, France, 29-30 mars 2018.

Roux J, Grimaud O, Leray E. L'étude des parcours de soins grâce au « multichannel sequence analysis » - Application à la sclérose en plaques en France sur 2007-2013. 9ème Rencontres Scientifiques du Réseau Doctoral en Santé Publique. Bordeaux, France, 13-14 mars 2018.

Roux J, Grimaud O, Leray E. Peut-on mesurer le niveau de handicap moteur lié à la sclérose en plaques à partir des bases de l'Assurance Maladie ? ADELF-EPITER (Association des Epidémiologistes de Langue Française – Association des Epidémiologistes de Terrain). Rennes, France, 7-9 septembre 2016.

Roux J, Le Meur N, Grimaud O, Leray E. Care pathways of patients affected with multiple sclerosis in France from 2007 to 2013 using administrative databases and state sequence analysis. LaCOSA II (International Conference on Sequence Analysis and Related Methods). Lausanne, Suisse, 8-10 juin 2016.

Roux J, Le Meur N, Grimaud O, Leray E. Utilisation des bases médico-administratives pour l'étude des parcours de soins des patients atteints de sclérose en plaques en France de 2007 à 2013. ADELF-EMOIS Joint congress. Dijon, France, 10-11 mars 2016.

- *Affichées*

Roux J, Casey R, Vukusic S, Leray E, on behalf of OFSEP investigators. Complementary use of administrative and clinical data in studying disease-modifying therapies in MS in France in 2010-2015. 26th Annual Meeting of the European Charcot Foundation. Baveno, Italie, 15-17 novembre 2018.

Guilleux A, Roux J, Leray E. Effects of mental comorbidities on initiating disease-modifying therapies: a population-based study.26th Annual Meeting of the European Charcot Foundation. Baveno, Italie, 15-17 novembre 2018.

Roux J, Guilleux A, Grimaud O, Leray E. Care-seeking of patients with multiple sclerosis over 2010-2015 in France: a nationwide study using health administrative data. 34th ECTRIMS Congress (European Committee for Treatment and Research in Multiple Sclerosis). Berlin, Allemagne, 10-12 octobre 2018.

Guilleux A, Roux J, Leray E. Care-seeking and disease modifying therapies use among multiple sclerosis patients with and without mental disorders: a population-based study. 34th ECTRIMS

Congress (European Committee for Treatment and Research in Multiple Sclerosis). Berlin, Allemagne, 10-12 octobre 2018.

Leray E, Reilhac A, Kerbrat S, Roux J, Polard E. Incidence of pregnancy in women with multiple sclerosis treated with teriflunomide in France. 34th ECTRIMS Congress (European Committee for Treatment and Research in Multiple Sclerosis). Berlin, Allemagne, 10-12 octobre 2018.

Roux J, Guilleux A, Leray E. Impact of new oral disease-modifying therapies on the prescription of injectable and off-label drugs in multiple sclerosis in France over 2011-2015 using national administrative databases. 34th International Conference on Pharmacoepidemiology & Therapeutic Risk Management (ICPE). Prague, République Tchèque, 22-26 août 2018.

Roux J, Grimaud O, Leray E. Multichannel sequence analysis: an innovative method to study patterns of care pathways – Application to multiple sclerosis based on French health insurance data. European Congress of Epidemiology. Lyon, France, 4-6 juillet 2018.

Roux J, Grimaud O, Leray E. Care consumption and use of disease-modifying therapies in multiple sclerosis over 2010-2015 in France: a nationwide study using health administrative data. 27th Annual MS meeting ARSEP. Paris, France, 1 juin 2018.

Leray E, Rollot F, Casey R, Roux J, De Seze J, Laplaud D; Vukusic S. Quelle utilisation des traitements de fond chez des patients ayant une SEP rémittente en France sur la période 1990-2017 ? Eléments de réponse à partir des données de l'OFSEP. JNLF (Journées de Neurologie de Langue Française). Bordeaux, France, 10-13 avril 2018.

Jeanjean M, Bind MA, Roux J, Ongagna JC, De Sèze J, Bard D, Leray E. Short-term exposure to ambient air pollution and occurrence of multiple sclerosis relapses. 7th Joint ECTRIMS-ACTRIMS Meeting (European and American Committee for Treatment and Research in Multiple Sclerosis). Paris, France, 25-28 octobre 2017.

Roux J, Grimaud O, Leray E. Care consumption of multiple sclerosis patients in France: an analysis of health insurance administrative databases using multichannel sequence analysis from 2007 to 2013. 7th Joint ECTRIMS-ACTRIMS Meeting (European and American Committee for Treatment and Research in Multiple Sclerosis). Paris, France, 25-28 octobre 2017.

Roux J, Guilleux A, Leray E. Adherence to oral versus injectable disease-modifying therapies for multiple sclerosis using French national health insurance databases. 7th Joint ECTRIMS-ACTRIMS Meeting (European and American Committee for Treatment and Research in Multiple Sclerosis). Paris, France, 25-28 octobre 2017.

Guilleux A, Roux J, Leray E. Mood disorders and multiple sclerosis: a study based upon the French national health insurance databases. 7th Joint ECTRIMS-ACTRIMS Meeting (European and American Committee for Treatment and Research in Multiple Sclerosis). Paris, France, 25-28 octobre 2017.

Roux J, Grimaud O, Leray E. L'analyse de séquences pour l'étude des parcours de soins - Application à la sclérose en plaques. Prix du meilleur poster. 8ème Rencontres Scientifiques du Réseau Doctoral en Santé Publique. Rennes, France, 14-15 mars 2017.

Roux J, Grimaud O, Leray E. Can the motor disability level related to multiple sclerosis be measured from the French health insurance database? EAN (European Academy of Neurology). Copenhagen, Danemark, 28-31 mai 2016.

Roux J, Grimaud O, Leray E. Consommations de soins des patients ayant une sclérose en plaques traités par fingolimod à partir d'un échantillon au 1/97e des données de l'Assurance Maladie en France. JNLF (Journées de Neurologie de Langue Française). Nantes, France, 5-8 avril 2016.

Roux J, Le Meur N, Grimaud O, Leray E. Consommations de soins des patients atteints de sclérose en plaques en France de 2007 à 2013 à partir des bases de données médico-administratives. JNLF (Journées de Neurologie de Langue Française). Nantes, France, 5-8 avril 2016.

Leray E, Roux J, Le Meur N. Use of administrative database to describe physician services utilization and care patterns in multiple sclerosis in France. XXII World Congress of Neurology (WCN 2015). Santiago, Chili, 31 octobre – 5 novembre 2015.

Table des matières

LISTE DES FIGURES XXI

LISTE DES TABLES XXIII

GLOSSAIRE XXV

I. INTRODUCTION 1

I.1. LA SCLEROSE EN PLAQUES 1

I.1.1. Une maladie neurologique 1

I.1.2. Diagnostic et formes cliniques 2

I.1.3. Symptômes 3

I.1.4. Handicap 5

I.1.5. Traitements 6

I.1.5.1. Traitements de fond 6

I.1.5.2. Séquences thérapeutiques 7

I.1.5.3. Effets secondaires des traitements de fond 9

I.1.5.4. Traitements des poussées 9

I.1.5.5. Traitements symptomatiques 10

I.1.6. Epidémiologie 11

I.1.7. Etiologie 11

I.1.8. Comorbidités 12

I.2. LE PARCOURS DE SOINS 13

I.2.1. Concept 13

I.2.2. Affection de longue durée 14

I.2.3. Le parcours de soins dans la SEP 15

I.2.3.1. Travaux de la Haute Autorité de Santé 15

I.2.3.2. Le Plan Maladies Neurodégénératives 16

I.2.3.3. Prise en charge au niveau international 17

I.2.3.4. Prise en charge en France 17

I.2.4. Recours aux soins 21

I.2.5. Coûts 21

I.3. L'ETUDE QUANTITATIVE DES PARCOURS DE SOINS 22

I.3.1. Bases à disposition 22

I.3.2. Méthodes d'analyse des parcours 23

I.4. PROBLEMATIQUE DE LA THESE 23

I.5. OBJECTIFS 24

I.5.1. Objectif principal 24

I.5.2. Objectifs secondaires 25

II. MATERIELS ET METHODES 27

II.1. BASES MEDICO-ADMINISTRATIVES-DEFINITIONS, ACCES ET STRUCTURES 27

II.1.1. Présentation des bases médico-administratives françaises 27

II.1.1.1. Le Système National des Données de Santé (SNDS) 27

II.1.1.2. Le Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM) 28

II.1.1.3. Le Datamart de Consommation Inter-Régimes (DCIR) 30

II.1.1.4. Le Programme de Médicalisation des Systèmes d'Information (PMSI) 31

II.1.1.5. L'Echantillon Généraliste des Bénéficiaires (EGB) 32

II.1.2. Contexte réglementaire 33

II.1.2.1. Accès aux données 33

II.1.2.2. Données sensibles 34

II.1.3. Structure du SNDS 35

II.1.3.1. Référentiel des consommants 36

II.1.3.2. Structure du DCIR 36

II.1.3.3. Structure du PMSI 37

II.1.3.4. Jointure à des bases externes 39

II.2. POPULATION D'ETUDE 40

II.2.1. Algorithme de sélection 40

II.2.2. Identification dans l'EGB 41

II.2.3. Identification dans le SNDS 42

II.2.3.1. Algorithme 42

II.2.3.2. Identification des patients uniques 43

II.2.3.3. Consolidation de la population d'étude 45

II.3. PARAMETRES ETUDIES 45

II.3.1. Pré-traitement des données du SNDS 45

II.3.1.1. Recommandations de l'Assurance Maladie avant traitement 46

II.3.1.2. Gestion des quantités délivrées négatives 47
II.3.1.3. Gestion des séjours non uniques 48
II.3.2. Paramètres d'ajustement 48
II.3.2.1. Lieu de résidence 48
II.3.2.2. Variables socio-économiques contextuelles 49
II.3.2.3. Variables socio-économiques individuelles 50
II.3.2.4. Régime de rattachement 52
II.3.2.5. Affection de Longue Durée 53
II.3.2.6. Indice de comorbidités de Charlson 53

III. DESCRIPTIF DE LA POPULATION SEP EN FRANCE 57

III.1. POPULATION SEP DANS L'EGB 57
III.1.1. Mode d'identification dans l'EGB 57
III.1.2. Descriptif des populations de l'EGB 58
III.2. POPULATION SEP DANS LE SNDS 59
III.2.1. Mode d'identification dans le SNDS 59
III.2.2. Prévalence de la SEP en France 60
III.2.3. Descriptif de la population identifiée dans le SNDS 61

IV. TRAITEMENTS DE FOND DANS LA SEP 65

IV.1. TRAITEMENTS DE FOND DANS LE SNDS 65
IV.1.1. Méthodes 65
IV.1.1.1. Identification des traitements 65
IV.1.1.2. Représentations graphiques 66
IV.1.2. Résultats 67
IV.1.2.1. Descriptif de la population 67
IV.1.2.2. Distribution des traitements de 2010 à 2015 67
IV.1.2.3. Séquences thérapeutiques 69
IV.2. SUIVI DU PLAN DE GESTION DES RISQUES DANS L'EGB 72
IV.2.1. Méthodes 72
IV.2.1.1. Identification des actes d'imagerie dans l'EGB 73
IV.2.1.2. Identification des actes de biologie dans l'EGB 74
IV.2.2. Résultats 75

IV.3. DISCUSSION **76**

IV.3.1. Discussion des résultats 76

IV.3.2. Forces et limites 78

V. QUANTIFICATION DU NIVEAU DE HANDICAP MOTEUR **81**

V.1. METHODES **81**

V.1.1. Identification des aides techniques à la marche 81

V.1.2. Construction du score EDSS « estimé » 82

V.2. RESULTATS **83**

V.2.1. Aides techniques à la marche 83

V.2.2. Score EDSS « estimé » 84

V.3. DISCUSSION **86**

V.3.1. Discussion des résultats 86

V.3.2. Forces et limites 87

VI. PARCOURS DE SOINS **91**

VI.1. METHODES **91**

VI.1.1. Recours aux soins 91

VI.1.1.1. Consultations avec un spécialiste libéral ou dans un établissement de santé privé 91

VI.1.1.2. Consultations externes 93

VI.1.1.3. Ensemble des consultations 93

VI.1.1.4. Actes infirmiers et de kinésithérapie 93

VI.1.1.5. Séjours hospitaliers 94

VI.1.2. L'analyse de séquences 95

VI.1.2.1. Contexte 95

VI.1.2.2. Termes spécifiques à l'analyse de séquences 96

VI.1.2.3. Construction du parcours de soins 96

VI.1.2.4. Mesure de la dissimilarité entre les parcours 98

VI.1.2.5. Classification des parcours 102

VI.1.2.6. Représentation des séquences 104

VI.1.4. L'analyse multichannel sequence analysis (MCSA) 104

VI.1.5. Applications 106

VI.1.5.1. Application de l'analyse de séquences unidimensionnelle à l'EGB 106
VI.1.5.2. Application de la MCSA à l'EGB 108
VI.1.5.3. Application de la MCSA au SNDS 109
VI.2. RESULTATS 110
VI.2.1. Analyse de séquences unidimensionnelle sur l'EGB 110
VI.2.2. MCSA sur l'EGB 117
VI.2.3. MCSA sur le SNDS 124
VI.2.4. MCSA sur des données canadiennes 137
VI.2.4.1. Contexte 137
VI.2.4.2. Méthodes 137
VI.2.4.3. Résultats 139
VI.3. DISCUSSION 142
VI.3.1. Discussion des résultats 142
VI.3.2. Forces et limites 144
VI.3.2.1. Forces et limites des méthodes d'analyse de séquences 144
VI.3.2.2. Forces et limites des données de consommation de soins 147
VII. DISCUSSION GENERALE 149
VII.1. RESULTATS PRINCIPAUX 149
VII.2. INTERET DES DONNEES MEDICO-ADMINISTRATIVES 151
VII.3. REFLEXION SUR LES HYPOTHESES FORMULEES 155
VII.4. PERSPECTIVES 158
VIII. CONCLUSION 161
REFERENCES 163
ANNEXES 181
ANNEXE A-SYSTEMES FONCTIONNELS ET ECHELLE DU SCORE EDSS 181
ANNEXE B-TRAITEMENTS DE FOND DE LA SEP 187
ANNEXE C-CODES CIP DES TRAITEMENTS DE FOND DELIVRES EN VILLE 189
ANNEXE D-TRAITEMENTS DE FOND DANS L'OFSEP 193
ANNEXE E-LISTE DES CODES LPP CONSIDERES POUR L'ESTIMATION DU SCORE EDSS 197

ANNEXE F-ARTICLE : USE OF STATE SEQUENCE ANALYSIS FOR CARE PATHWAY ANALYSIS: THE EXAMPLE OF MULTIPLE SCLEROSIS **205**

ANNEXE G-RESULTATS DE L'APPLICATION DE LA MCSA SUR LES DONNEES CANADIENNES **219**

Liste des figures

Figure 1-Représentation du mécanisme d'action de la SEP (5) 2

Figure 2-Représentation schématique des trois formes cliniques principales de SEP (adapté de (12)) 3

Figure 3-Principaux symptômes de la SEP (23) 4

Figure 4-Echelle de l'Expanded Disability Status Scale (EDSS) (adapté de (27))..... 5

Figure 5-Frise chronologique des autorisations de mise sur le marché des traitements de fond de la SEP en France..... 6

Figure 6-Représentation schématique des stratégies thérapeutiques dans la SEP..... 8

Figure 7-Représentation des différentes catégories de parcours selon les domaines considérés (84) 14

Figure 8-Proposition de prise en charge des patients SEP selon la HAS (2006) (89) 15

Figure 9-Schéma de l'organisation des soins chez le patient SEP (95) 18

Figure 10-Algorithme FLUE-MS (First Line Urological Evaluation in Multiple Sclerosis) permettant l'évaluation et le traitement des troubles urinaires dans la SEP (95)..... 20

Figure 11-Synthèse des bases constitutives du SNDS en France 28

Figure 12-Calendrier synthétisant la démarche réglementaire d'accès aux données du SNDS pour ce travail de thèse..... 34

Figure 13-Modèle conceptuel du DCIR (143)..... 37

Figure 14-Flow-chart de la constitution de la population d'étude dans le SNDS 44

Figure 15-Diagramme de Venn du mode d'identification des personnes ayant la SEP dans l'EGB sur 2007-2012 (N=1 000)..... 57

Figure 16-Pyramide des âges de la population globale identifiée dans l'EGB (N=1 000) 58

Figure 17-Diagramme de Venn du mode d'identification des personnes ayant la SEP dans le SNDS sur 2010-2015 (N=112 745)..... 60

Figure 18-Carte des taux de prévalence standardisés de la SEP en France au 31 décembre 2015 (N=107 472) .. 61

Figure 19-Pyramide des âges de la population SEP identifiée dans le SNDS (N=112 745) 62

Figure 20-Chronogrammes de la répartition des traitements sur la période d'étude 2010-2015 selon (A) la substance active et (B) la voie d'administration (N=112 745)..... 68

Figure 21-Evolution du nombre annuel de nouveaux utilisateurs de chaque substance active sur 2010-2015.... 69

Figure 22-Sunburst des séquences thérapeutiques observées sur la période d'étude 2010-2015 (N=112 745) ... 70

Figure 23-Diagrammes de Sankey de la répartition des traitements sur la période d'étude 2010-2015 selon (A) la substance active et (B) la voie d'administration (N=112 745)..... 71

Figure 24-Recommandations de suivi des patients sous fingolimod d'après le PGR (Laboratoire Novartis) 72

Figure 25-Synthèse des résultats concernant le suivi des recommandations dans l'EGB suite à la délivrance de fingolimod (n=70)..... 76

Figure 26-Représentation schématique de la construction d'une séquence correspondant au parcours de soins d'un individu pour l'analyse de séquences 96

Figure 27-Opérations élémentaires de l'analyse de séquences présentées avec (A) un exemple avec des mots et (B) avec des parcours de soins assimilés à des séquences 100
Figure 28-Impact de la mesure de dissimilarité sur la comparaison des séquences (adapté de (207)) 101
Figure 29-Représentation des séquences sous forme (A) d'un index-plot et (B) d'un chronogramme..... 104
Figure 30-Représentation schématique de la construction d'une séquence correspondant au parcours de soins d'un individu pour la MCSA 106
Figure 31-Index plots des groupes obtenus après application de l'analyse de séquences unidimensionnelle sur la population des individus identifiés en 2007 dans l'EGB avec un coût d'indel de 0,9 et des coûts de substitution basés sur les transitions (n=648) 113
Figure 32-Index plots des groupes obtenus après application de la MCSA sur la population des individus identifiés en janvier 2007 dans l'EGB avec un coût d'indel de 1 et des coûts de substitution basés sur les transitions (n=543) 121
Figure 33-Index plots des clusters de la partition en cinq groupes obtenue suite à la MCSA sur l'échantillon aléatoire de patients SEP identifiés avant janvier 2010 dans le SNDS (N=35 000) 131
Figure 34-Chronogramme de la répartition des traitements sur la période d'étude 2010-2015 selon la substance active dans la cohorte de l'OFSEP (N=36 329)..... 195
Figure 35-Sunburst des séquences thérapeutiques observées sur la période d'étude 2010-2015 dans la cohorte de l'OFSEP (N=36 329) (DMT=traitement de fond) 195
Figure 36-Diagramme de Sankey de la répartition des traitements sur la période d'étude 2010-2015 selon la substance active dans la cohorte de l'OFSEP (N=36 329)..... 196

Liste des tables

Table 1-Informations disponibles et indisponibles dans les données du DCIR en 2016 (adapté de (131))..... 31

Table 2-Synthèse des jointures entre tables du SNDS et bases de connaissances externes 40

Table 3-Synthèse des populations sources construites dans l'EGB et des données de consommation correspondantes..... 42

Table 4-Liste des petits régimes sélectionnés selon le dispositif d'accès aux soins considéré..... 52

Table 5-Pathologies considérées dans le calcul de l'indice de comorbidités de Charlson et poids associés selon l'algorithme considérée..... 55

Table 6-Principales caractéristiques des trois populations identifiées dans l'EGB 59

Table 7-Descriptif de la population SEP identifiée dans le SNDS (N=112 745) 63

Table 8-Codes CCAM considérés pour l'identification des IRM 74

Table 9-Codes NABM considérés pour l'identification des actes de biologie 75

Table 10-Attribution des scores EDSS « estimés » selon les aides techniques à la marche considérées 83

Table 11-Prescription des aides techniques à la marche sur la période d'étude 2010-2015 dans la population identifiée dans le SNDS (N=112 745) 83

Table 12-Score EDSS « estimé » à l'inclusion et aux dernières nouvelles dans la population identifiée dans le SNDS (N=112 745) 84

Table 13-Répartition des patients en fonction de l'évolution du score EDSS « estimé » entre l'inclusion et les dernières nouvelles dans la population du SNDS (N=112 745)..... 85

Table 14-Répartition des patients en fonction de l'évolution du score EDSS « estimé » entre l'inclusion et les dernières nouvelles dans la population due l'EGB (N=650) 85

Table 15-Codes de nature de prestation considérés pour l'identification des consultations 92

Table 16-Codes des spécialités considérées pour la consommation de soins 93

Table 17-Codes de nature de prestation considérés pour la sélection des actes infirmiers et de kinésithérapie.. 94

Table 18-Critère de mesure de la qualité d'une partition (adapté de (190))..... 103

Table 19-Caractéristiques des populations pour l'étude du recours aux soins grâce à l'analyse de séquences unidimensionnelle identifiées dans l'EGB en 2007 (N=648) et après (N=352) 111

Table 20-Matrice des coûts de substitution pour l'analyse de séquences unidimensionnelle appliquée sur les patients SEP identifiés en 2007 dans l'EGB basés sur les transitions observées dans les données (N=1 000) 112

Table 21-Durée moyenne dans chacun des états en année (part de la durée totale) de chaque état selon la classification obtenue pour les patients identifiés en 2007 dans l'EGB (n=648) 112

Table 22-Caractéristiques des clusters de la partition en cinq groupes obtenue suite à l'analyse de séquences unidimensionnelle appliquée sur les patients SEP identifiés en 2007 dans l'EGB (n=648) 115

Table 23-Caractéristiques de la population identifiée en janvier 2007 dans l'EGB pour l'étude du recours aux soins grâce à la MCSA (N=543) 117

Table 24-Bornes de la catégorisation des dimensions considérées dans la MCSA sur la population identifiée dans l'EGB en janvier 2007 118
Table 25-Matrices des coûts de substitution pour la MCSA sur la population identifiée dans l'EGB en janvier 2007 basés sur les transitions observées dans les données (N=543) 118
Table 26-Caractéristiques des clusters de la partition en cinq groupes obtenue suite à la MCSA appliquée sur les patients SEP identifiés en janvier 2007 dans l'EGB (n=543) 122
Table 27-Caractéristiques des populations globale (N=112 745), identifiée avant le 1er janvier 2010 dans le SNDS (n=64 373) et de l'échantillon considéré pour l'étude du recours aux soins grâce à la MCSA (n=35 000). 124
Table 28-Bornes de la catégorisation des dimensions considérées dans la MCSA sur la population identifiée avant le 1er janvier 2010 dans le SNDS (n=64 373) 127
Table 29-Matrices des coûts de substitution pour la MCSA sur l'échantillon aléatoire de 35 000 patients issus de la population identifiée avant le 1er janvier 2010 dans le SNDS basés sur les transitions observées dans l'ensemble des données (N=64 373)..... 128
Table 30-Caractéristiques des clusters de la partition en cinq groupes obtenue suite à la MCSA appliquée sur l'échantillon de patients SEP identifiés avant janvier 2010 dans le SNDS (n=35 000)..... 132
Table 31-Bornes de la catégorisation des dimensions considérées dans la multichannel sequence analysis sur la population identifiée dans les bases canadiennes (cohortes 1 et 2) 139
Table 32-Répartition des patients dans les différents groupes suite à l'application de la multichannel sequence analysis sur les individus de la cohorte 1 (N=1 048)..... 140
Table 33-Répartition des patients dans les différents groupes suite à l'application de la multichannel sequence analysis sur les individus de la cohort 2 (N= 3180)..... 141
Table 34-Description des scores fonctionnels et des critères d'attribution des scores..... 181
Table 35-Echelle du score EDSS 184
Table 36-Informations sur les traitements de la SEP classés selon leur ordre d'admission et leur date d'autorisation de mise sur le marché 187
Table 37-Codes CIP utilisées pour le repérage des traitements de fond délivrés en ville 189
Table 38-Codes LPP des différents matériels considérés pour le calcul du score EDSS « estimé » 197
Table 39-Caractéristiques de la cohorte 1 (N=1 048) et de la cohorte 2 (N=3 180) 219
Table 40-Descriptif des comorbidités présentes des populations de la cohorte 1 (N=1 048) et de la cohorte 2 (N=3 180)..... 221
Table 41-Descriptif des comorbidités présentes des populations de la cohorte 1 (N=1 048) et de la cohorte 2 (N=3 180)..... 222
Table 42-Caractéristiques des clusters de la partition en douze groupes obtenue suite à la multichannel sequence analysis appliquée sur la cohorte 1 (n=1 048) 223
Table 43-Caractéristiques des clusters de la partition en six groupes obtenue suite à la multichannel sequence analysis appliquée sur la cohort 2 (N=3 180) 239

Glossaire

ACE	Actes et Consultations Externes
ACP	Analyse en Composantes Principales
ALD	Affection de Longue Durée
AME	Aide Médicale de l'Etat
AMM	Autorisation de Mise sur le Marché
ANOVA	Analyse de la Variance/Analysis of Variance
APHM	Assistance Publique des Hôpitaux de Marseille
APHP	Assistance Publique des Hôpitaux de Paris
APL	Accessibilité Potentielle Localisée
ARMSS	Age Related Multiple Sclerosis Severity Score
ASPA	Allocation de Solidarité aux Personnes Agées
ASWw	Weigthted Average Silhouette Width
ATC	Classification Anatomique, Thérapeutique et Chimique
ATIH	Agence Technique de l'Information sur l'Hospitalisation
ATU	Autorisation Temporaire d'Utilisation
BC	British Columbia
BCMS	British Columbia Multiple Sclerosis
CAH	Classification Ascendante Hiérarchique
CAMIEG	Caisse d'Assurance Maladie des Industries Electriques et Gazières
CANSSM	Caisse Autonome Nationale de la Sécurité Sociale dans les Mines
CAVIMAC	Caisse d'Assurance Vieillesse, Invalidité et Maladie des Cultes
CCAM	Classification Commune des Actes Médicaux
CCAS RATP	Caisse de Coordination aux Assurances Sociales de la RATP
CCTIRS	Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé
CépiDC	Centre d'épidémiologie sur les causes médicales décès
CEREES	Comité d'Expertise pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé
CHRU	Centre Régional Hospitalier Universitaire
CHU	Centre Hospitalier Universitaire
CIM-10	Classification Internationale des Maladies – 10ème version
CIM-9	Classification Internationale des Maladies – 9ème version
CIP	Code Identifiant de Présentation
CMU	Couverture Maladie Universelle

CMU-C	Couverture Maladie Universelle - Complémentaire
CNAMTS	Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés
CNIL	Commission Nationale Informatique et Libertés
CNMSS	Caisse Nationale Militaire de Sécurité Sociale
CPM Banque de France	Caisse de Prévoyance Maladie de la Banque de France
CPPAB	Caisse de Prévoyance du Port Autonome de Bordeaux
CPRPSNCF	Caisse de Prévoyance et de Retraite du Personnel de la SNCF
CRCSEP	Centre de Ressources et de Compétences sur la Sclérose En Plaques
CRPCEN	Caisse de Retraite et de Prévoyance des Clercs et Employés de Notaire
CSARR	Catalogue Spécifique des Actes de Rééducation et Réadaptation
DAS	Diagnostic Associé Significatif
DCIR	Datamart de Consommation Inter-Régimes
DMI	Dispositif Médical Implantable
DP	Diagnostic Principal
DR	Diagnostic Relié
DREES	Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques
DROM	Départements et Régions d'Outre-Mer
EDMUS	European Database for Multiple Sclerosis
EDSS	Expanded Disability Satus Scale
EGB	Echantillon Généraliste des Bénéficiaires
EHESP	Ecole des Hautes Etudes en Santé Publique
EHPAD	Etablissement d'Hébergement pour Personnes Agées Dépendantes
ENIM	Etablissement National des Invalides de la Marine
ESPIC	Etablissement de santé privé d'intérêt collectif
FDep	Indice de défavorisation sociale
FINESS	Fichier National des Établissements Sanitaires et Sociaux
FISH	Faisabilité de l'Identification des Situations de Handicap
FLUE-MS	First Line Urological Evaluation in Multiple Sclerosis
FSSAN	Fonds de Sécurité Sociale de l'Assemblée Nationale
GHJ	Groupe Homogène Journalier
GHM	Groupe Homogène de Malades
GHS	Groupe Homogène de Séjour
HAS	Haute Autorité de Santé
HC	Hubert's C
HCL	Hospices Civils de Lyon

HG	Hubert's Gamma
IC95%	Intervalle de confiance à 95%
IDE	Infirmier Diplômé d'Etat
IDS	Institut des Données de Santé
INDS	Institut National des Données de Santé
Insee	Institut national de la statistique et des études économiques
Inserm	Institut national de la santé et de la recherche médicale
IR	Incidence Ratio
Irdes	Institut de recherche et documentation en économie de la santé
Iris	Ilots regroupés pour l'information statistique
IRM	Imagerie par Résonance Magnétique
kml	k-means appliquées aux trajectoires
LASSO	Least Absolute Shrinkage and Selection Operator
LCA	Latent Class Analysis/Analyse en classes latentes
LEMP	Leucoencéphalopathie Multifocale Progressive
LORSEP	Réseau Lorrain pour la prise en charge de la Sclérose En Plaques
LPP	Liste des Produits et Prestations remboursables
MCSA	Multichannel Sequence Analysis/analyse séquentielle multidimensionnelle
MDPH	Maisons Départementales Pour le Handicap
MG	Médecin Généraliste
MIP	Module Interprofessionnel de santé publique
MPR	Médecine Physique et de Réadaptation
MSA	Mutualité Sociale Agricole
NABM	Nomenclature des Actes de Biologie Médicale
NFS	Numération de la Formule Sanguine
NGAP	Nomenclature Générale des Actes Professionnels
NIR	Numéro d'Inscription au Répertoire des personnes physiques
NMS	Number of Matching Subsequences
NOEMIE	Norme Ouverte d'Échanges entre la Maladie et les Intervenants Extérieurs
OCDE	Organisation de Coopération et de Développement Economique
OFSEP	Observatoire Français de la Sclérose En Plaques
PAM	Partitioning Around Medoids
PBC	Point Biserial Correlation

PGR	Plan de Gestion des Risques
PiMS	Pharmacoepidemiology in Multiple Sclerosis
PMND	Plan Maladies Neurodégénératives
PMSI	Programme de Médicalisation des Systèmes d'Information
PMSI HAD	Programme de Médicalisation des Systèmes d'Information Hospitalisations A Domicile
PMSI MCO	Programme de Médicalisation des Systèmes d'Information Médecine Chirurgie Obstétrique Odontologie
PMSI PSY	Programme de Médicalisation des Systèmes d'Information Psychiatrique
PMSI SSR	Programme de Médicalisation des Systèmes d'Information Soins de Suite et de Réadaptation
PUMA	Protection Universelle Maladie
RA3S	Régime Autonome de Sécurité Sociale du Sénat
RATP	Régie autonome des transports parisiens
RHA	Résumé Hebdomadaire de sortie Anonymisé
RIM-P	Recueil d'Informations Médicalisé pour la Psychiatrie
RNIPP	Répertoire National d'Identification des Personnes Physiques
RR	Risque Relatif
RSA	Résumé de Sortie Anonyme
RSI	Régime Social des Indépendants
SEP	Sclérose En Plaques
SEP PP	Sclérose En Plaques Progressive d'emblée
SEP RR	Sclérose En Plaques Rémittente
SEP SP	Sclérose En Plaques Secondairement Progressive
SLM	Section Locale Mutualiste
SNC	Système Nerveux Central
SNCF	Société Nationale des Chemins de Fers français
SNDS	Système National des Données de Santé
UBC	University of British Columbia
UBL	Université Bretagne-Loire
UCD	Unité Commune de Dispensation
UM	Unité Médicale
VHP	Véhicule pour Handicapé Physique
VZV	Virus Varicelle-Zona

I. Introduction

I.1. La sclérose en plaques

I.1.1. Une maladie neurologique

Le système nerveux central (SNC), constitué du cerveau et de la moelle épinière et opposé au système nerveux périphérique formé des nerfs, est le siège de l'activité neurologique dans le corps humain (1). Il permet le contrôle et la régulation de l'ensemble de l'organisme, notamment la motricité, la perception de nos cinq sens, l'équilibre, mais également les fonctions cognitives et intellectuelles. C'est également grâce au SNC que fonctionnent certains organes comme l'intestin et la vessie par exemple. Son bon fonctionnement est donc essentiel dans le corps humain. Deux cellules composent principalement le SNC : les neurones et les cellules gliales. Les premiers sont l'unité fonctionnelle principale du système nerveux et les secondes permettent leur fonctionnement et leur entretien (2). Les neurones reçoivent, analysent, produisent et transfèrent les informations par influx nerveux, une suite d'impulsions électriques aussi appelés potentiels d'action, jusqu'aux muscles, organes ou glandes (Figure 1). Pour ce faire, un neurone est formé d'un corps cellulaire muni de prolongements appelés dendrites, servant à réceptionner l'influx nerveux des autres neurones, et d'un axone unique servant au transport de l'information vers la cellule cible. Afin de protéger l'axone et d'accélérer la vitesse de propagation de l'influx nerveux, une gaine de myéline l'entoure. Cette dernière est composée de protéines et de lipides et est produite par des cellules gliales dédiées, les oligodendrocytes. Lorsque cette gaine est endommagée, on parle de démyélinisation. La propagation de l'influx nerveux peut alors être perturbée (fourmillements, troubles urinaires, de la vision, …) voire stoppée (paralysie, perte de la vue, …), entraînant un handicap plus ou moins important.

La sclérose en plaques (SEP), découverte en 1868 par Jean-Martin Charcot (3), est une maladie auto-immune inflammatoire résultant de cette démyélinisation du SNC. Elle se caractérise par l'altération anormale par le propre système immunitaire du patient, principalement les lymphocytes B et T, de la gaine de myéline protégeant les axones créant une inflammation du système nerveux. Ce phénomène se traduit par l'apparition de lésions épaisses et dures, appelées scléroses, le plus souvent sous forme de plaques d'où le nom de la maladie. Ces lésions pouvant apparaître dans différentes localisations du SNC, les symptômes de la SEP sont donc divers et diffèrent selon chaque personne. La gaine de myéline peut être recréée par les oligodendrocytes dans certains cas principalement au début de la

maladie (on parle alors de remyélinisation), le retour de la communication nerveuse est alors possible faisant disparaître les symptômes. Cette évolution avec l'apparition et la disparition de symptômes a conduit à l'utilisation du terme de « poussée » pour caractériser l'apparition de nouveaux ou à l'aggravation d'anciens symptômes pendant une durée supérieure à 24 heures, en dehors d'une période de fièvre (4).

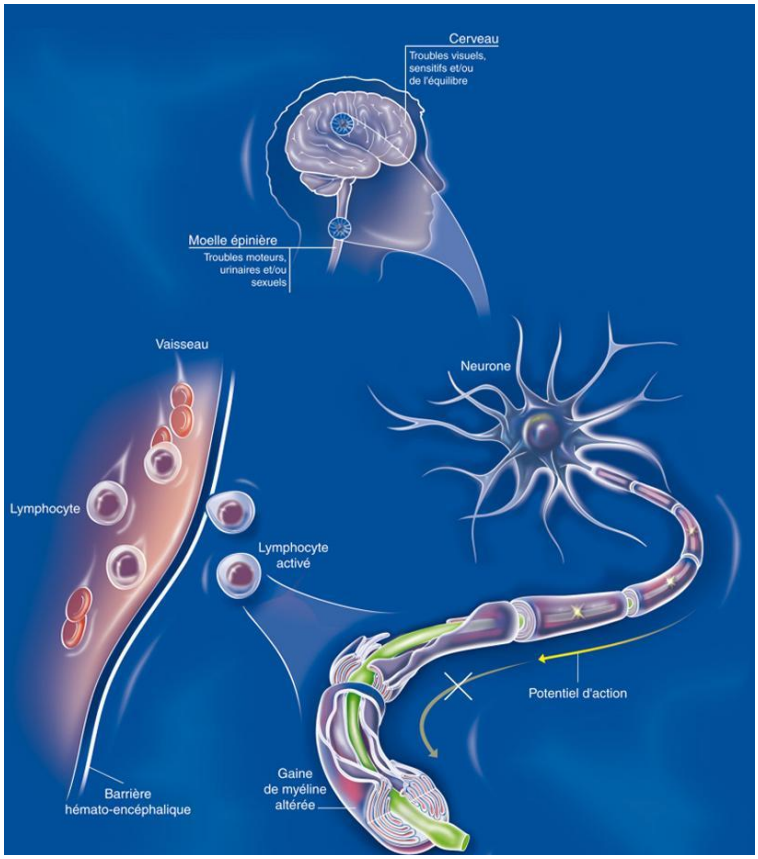
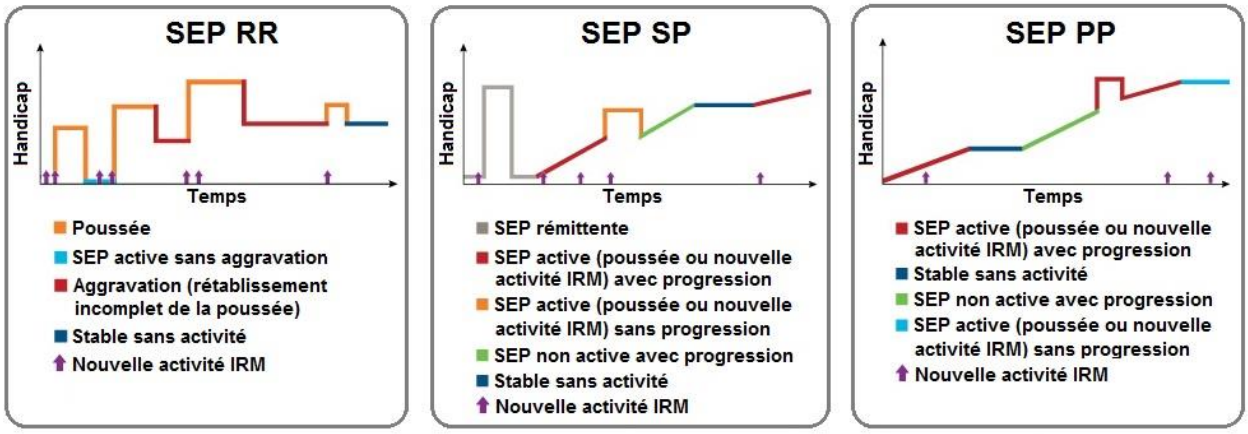


Figure 1-Représentation du mécanisme d'action de la SEP (5)

I.1.2. Diagnostic et formes cliniques

Les premiers critères diagnostiques de la SEP ont été proposés par Poser et al. en 1983 (6). Ces derniers étant devenus obsolètes avec le développement de l'Imagerie par Résonance Magnétique (IRM), de nouveaux critères, appelés critères de Mc Donald, ont été proposés à partir de 2001 par un groupe d'experts internationaux (7). Ces critères, révisés en 2005 (8), en 2010 (9) et pour la dernière fois en 2017 (10),ont pour but de permettre une identification de plus en plus précoce de la maladie. Ils reposent sur les notions de disséminations temporelle (plusieurs lésions au cours du temps) et spatiale (plusieurs lésions à différentes localisations sur le SNC), révélées par IRM, mais aussi sur la notion d'inflammation, mise en évidence par la réalisation d'une ponction lombaire. Une fois le diagnostic de la SEP posé, sa forme évolutive peut être déterminée. Ainsi, comme défini par Lublin et al., il existe la

forme rémittente et la forme progressive (11,12) (Figure 2). La forme rémittente (SEP RR), représentant 85% des cas, est caractérisée par des poussées successives, avec ou sans séquelles, espacées par des périodes sans symptômes ni évolution de la maladie (4,13). La forme progressive correspond à une évolution lente et progressive de la SEP, avec ou sans poussées surajoutées durant laquelle le handicap de la personne atteinte ne peut que s'amplifier. Cette dernière peut être progressive d'emblée (SEP PP), correspondant à 15% des formes de début et majoritairement à des personnes démarrant la maladie après 40 ans. Dans ce cas, la personne atteinte développe un handicap permanent dès le début de la maladie. Ou alors elle peut être secondairement progressive (SEP SP), si elle succède à une phase initialement rémittente (14), la personne atteinte gardant des séquelles de ses précédentes poussées.



Source: Lublin et al., 2014.

Figure 2-Représentation schématique des trois formes cliniques principales de SEP (adapté de (12))

I.1.3. Symptômes

Comme évoqué précédemment, les mécanismes inhérents à la SEP conduisent à des symptômes différents pour chaque individu et même différents pour un même individu au cours du temps. Ces symptômes peuvent être définitifs ou alors temporaires pour des durées allant de quelques heures à plusieurs semaines, la durée moyenne étant de deux à six semaines. Malgré une diversité de symptômes (Figure 3), plusieurs grandes familles peuvent être identifiées.

Les troubles liés à l'atteinte pyramidale, se traduisant principalement par des troubles de la motricité, et de la spasticité, c'est-à-dire une raideur excessive dans les muscles, sont majoritaires dans la SEP. En effet, environ 90% des personnes ayant une SEP ont des troubles de la spasticité au cours de leur vie (15), et environ 90% expriment avoir des troubles moteurs dans les 10 ans suivant le diagnostic (16). L'atteinte cérébelleuse, c'est-à-dire du cervelet, peut conduire à des faiblesses musculaires, des

troubles de l'équilibre ou de l'ataxie, soit un manque de coordination des mouvements. Présents dans 20% des symptômes initiaux de la maladie, les troubles sensitifs peuvent se traduire par des sensations de brûlures, de ruissellement, des douleurs neurologiques ou des fourmillements sur tout ou partie du corps (17). Les troubles visuels représentent environ 25% des premiers symptômes se traduisant notamment par des névrites inflammatoires optiques caractérisant une inflammation du nerf optique conduisant à une baisse de l'acuité visuelle voire une perte d'une partie de la vision (17). Les troubles urinaires (incontinence, dysurie (difficultés à uriner), …) atteignent plus de 75% des personnes ayant la SEP au cours de leur suivi du fait de la forte innervation du système vésico-sphinctérien (15,18).
Parmi les autres symptômes de la SEP, on trouve les troubles du langage, les troubles cognitifs (perte de mémoire, troubles de l'attention, difficultés de réflexion, …), ces derniers ayant une prévalence estimée entre 43% et 70% (19). Les troubles intestinaux affectent environ 68% des personnes ayant une SEP, notamment la constipation dans près de 50% des cas (15). La prévalence des troubles sexuels est également élevée avec entre 50% et 90% des hommes touchés et entre 40% et 80% des femmes. Enfin, la fatigue est le symptôme le plus répandu dans la maladie avec une prévalence oscillant entre 50% et près de 90% des patients (15,20,21) et est pour plus de la moitié des patients le symptôme le plus problématique (22).

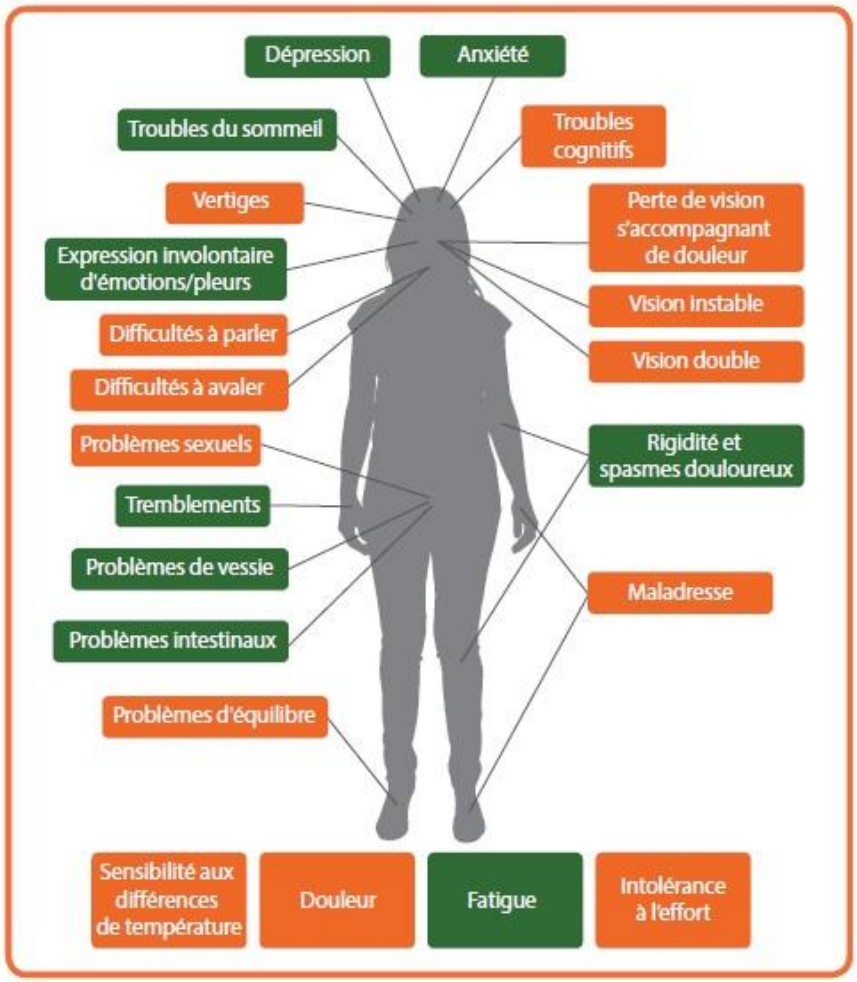


Figure 3-Principaux symptômes de la SEP (23)

I.1.4. Handicap

La principale conséquence des symptômes évoqués ci-dessus est l'apparition progressive d'un handicap. En effet, dans la majorité des cas, la SEP aboutit à plus ou moins long terme à un handicap principalement moteur, le délai médian avant une aide à la marche étant de 20 à 25 ans (24,25). Huit systèmes fonctionnels permettent de caractériser ce handicap : quatre majeurs (fonction pyramidale, fonction cérébelleuse, fonctions du tronc cérébral et fonction sensitive) et quatre mineurs (fonctions sphinctérienne et intestinale, fonction visuelle, fonction cérébrale ou mentale et autres fonctions) (Annexe A). Les quatre systèmes majeurs font référence principalement à l'appareil locomoteur, à la coordination des mouvements et aux troubles sensitifs. En complément les systèmes mineurs concernent les troubles liés aux appareils urinaire et intestinal, à la vision, à la capacité de réflexion et aux autres perturbations neurologiques pouvant être attribuées à la SEP. Afin de quantifier le handicap lié à la SEP, l'échelle EDSS (Expanded Disability Status Scale) a été développée par Kurtzke à partir de 1983 (26). Cette dernière est basée sur l'évaluation de chacun des huit systèmes fonctionnels précédents et conduit selon le niveau d'atteinte des systèmes à un score allant de 0 à 10 (Figure 4). Les points de repère sur cette échelle sont le score 0 correspondant à un examen neurologique normal, 6 signifiant le besoin d'une aide unilatérale à la marche (canne, béquille, canne anglaise), et 10 le décès du patient à cause de la SEP (27). Les systèmes fonctionnels étant majoritairement liés aux troubles locomoteurs, le score EDSS reflète principalement le handicap moteur dont la limitation du périmètre de marche des personnes atteintes de la SEP.

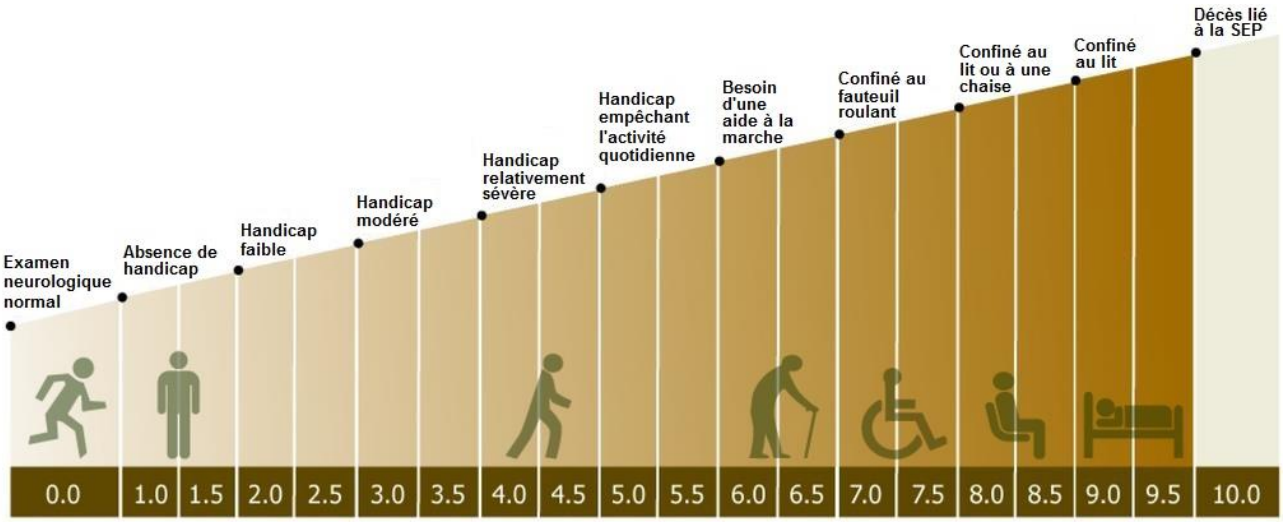


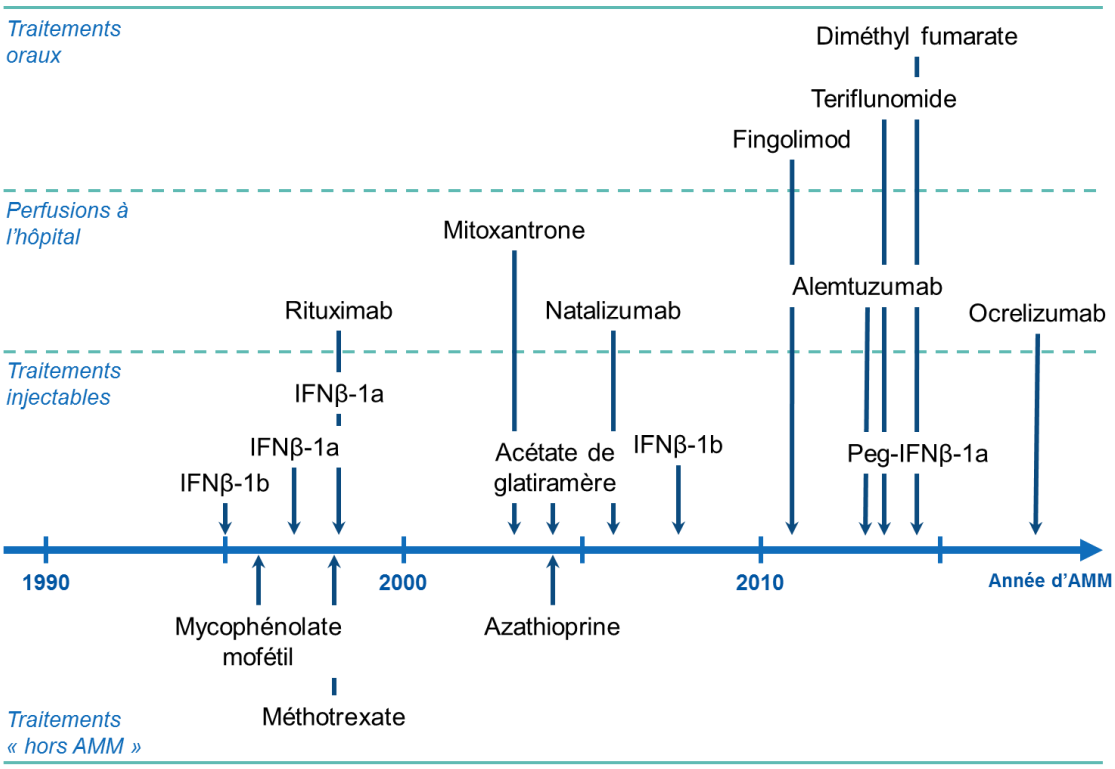
Figure 4-Echelle de l'Expanded Disability Status Scale (EDSS) (adapté de (27))

I.1.5. Traitements

I.1.5.1. Traitements de fond

Actuellement la SEP est une maladie pour laquelle aucun traitement curatif n'est disponible. Pendant longtemps, seuls des traitements des symptômes de la maladie et non de la maladie elle-même ont été délivrés aux patients. Grâce aux progrès de la médecine, des traitements spécifiques à la SEP ont été disponibles à partir du milieu des années 1990. Ainsi, les interférons ont été les premiers traitements approuvés en France à partir de 1995. Ces traitements de fond visent à diminuer la fréquence des poussées et à limiter le handicap. Agissant sur l'inflammation, ils sont prescrits majoritairement pour les SEP rémittentes, alors qu'il n'existe à l'heure actuelle qu'un seul traitement pour la SEP progressive primaire uniquement ayant reçu une autorisation de mise sur le marché (AMM) en France permettant sa commercialisation.

Durant les dernières décennies, l'arsenal thérapeutique, à prescription réservée uniquement aux neurologues, s'est considérablement élargi. Au total, 13 médicaments spécifiques à la maladie sont disponibles en 2018, certains étant délivrés à l'hôpital uniquement, d'autres en ville dans les pharmacies (Figure 5 et Annexe B). Ces traitements peuvent être classés selon leur voie d'administration (injectable, orale ou en perfusion), leur mécanisme d'action (immuno-modulateur ou immuno-suppresseur) et leur ordre dans la séquence thérapeutique, formée de l'ensemble des traitements pris successivement par un patient (première ligne, deuxième ligne et troisième ligne).



IFNβ : Interféron-β. AMM : Autorisation de Mise sur le Marché

Figure 5-Frise chronologique des autorisations de mise sur le marché des traitements de fond de la SEP en France

Ainsi, en première ligne, des immuno-modulateurs, ayant pour but de réguler le système immunitaire, à délivrance en officine, peuvent être prescrits. Ces derniers se présentent sous forme injectable et sont présents sur le marché pour certains depuis plus de vingt ans. Ils ont pour substance active les interférons bêta-1a (Avonex® et Rebif®) ou bêta-1b (Betaferon® et Extavia®), ou encore l'acétate de glatiramère (Copaxone®). Afin d'améliorer la qualité de vie des patients et limiter les effets secondaires liés aux injections, de nouveaux traitements sont arrivés sur le marché depuis 2013. Ainsi, les derniers traitements ayant obtenu une AMM en 2013 et 2014 sont sous forme de gélule, donc orale, comme le diméthyl fumarate (Tecfidera®) et le teriflunomide (Aubagio®) ou sous forme injectable seulement deux fois par mois (peginterféron bêta-1a (Plegridy®)), contre des injections quotidiennes ou pluri-hebdomadaires pour les interférons.

Des traitements immuno-suppresseurs administrés oralement utilisés dans d'autres maladies chroniques telles que la polyarthrite rhumatoïde ou le cancer, et donc non spécifiques à la SEP, peuvent également être prescrits en première ligne, et ont alors comme substance active l'azathioprine (Imurel®), le méthotrexate (Methotrexate®) ou le mycophénolate mofétil (Cellcept® et génériques). Ces trois médicaments sont prescrits hors AMM dans le cadre de la SEP. Ces médicaments ont été fortement prescrits en France des années 70 aux années 90 mais le sont beaucoup moins à l'heure actuelle (en termes d'initiation tout du moins).

Pour des SEP actives avec un nombre important de lésions, ou lorsqu'un traitement de première ligne ne fait pas effet, un traitement de seconde ligne peut-être proposé. D'un côté, le Gilenya®, se présentant sous forme orale et ayant pour substance active le fingolimod, est le seul délivré en pharmacie. De l'autre, le natalizumab (Tysabri®), l'alemtuzumab (Campath®) et le rituximab (Mabthera®), ce dernier étant non spécifique à la SEP, sont prescrits en perfusion à l'hôpital par voie intraveineuse. Enfin, en cas de SEP agressive, le traitement par mitoxantrone (Novantrone® puis Elsep®) peut être proposé en milieu hospitalier. Par ailleurs, plusieurs autres traitements sont en attente d'AMM en France et/ou en autorisation temporaire d'utilisation (ATU), comme le Lemtrada® ayant pour substance active l'alemtuzumab ou encore la biotine (Quizenday®), l'ocrelizumab (Ocrevus®) et la cladribine (Mavenclad®).
Le conditionnement de l'ensemble des traitements de fond de la SEP est tel qu'une boîte correspond à un mois de traitement avec un prix variant de 700 € à près de 2000 € la boîte.

I.1.5.2. Séquences thérapeutiques

La grande variété de traitements de fond disponibles entraîne de plus nombreux changements de traitements, appelés également switches, qu'au début des années 90 où un patient pouvait rapidement

se trouver en échec thérapeutique, c'est-à-dire ne répondant à aucun traitement disponible. Les raisons principales de ces switches sont 1) la tolérabilité du traitement avec l'apparition ou non d'effets secondaires et 2) l'inefficacité du traitement, correspondant à la progression de la maladie avec l'apparition ou la résurgence de nouvelles plaques sur l'imagerie (28). L'arrêt peut aussi correspondre à un choix du patient, pour lui permettre une amélioration de sa qualité de vie ou à la survenue d'une grossesse (29). La majorité des traitements de fond disposant de l'AMM ne nécessitent pas une période de wash-out, c'est-à-dire une période sans aucun traitement, permettant un switch immédiat entre traitements, excepté le teriflunomide (29). En effet, pour ce dernier, une procédure d'élimination à base de charbon actif existe, la demi-vie du principe actif étant très longue.

Dépendant de la sévérité et de l'avancée de la maladie, deux stratégies thérapeutiques principales peuvent être envisagées dans le traitement de la SEP : l'escalade thérapeutique et l'induction (28–31) (Figure 6). La première consiste en un début de traitement avec un immuno-modulateur de première ligne, suivi d'un traitement de seconde ligne, puis de troisième ligne dans une dernière mesure, dans le cas d'une faible efficacité sur la progression de la maladie. La stratégie d'induction emprunte le chemin inverse, c'est-à-dire un premier traitement immunosuppresseur (de troisième ligne) suivi d'un traitement de seconde et/ou de première ligne (31). Contrairement à l'escalade thérapeutique, l'induction est beaucoup moins fréquente et reste réservée à des situations bien particulières.

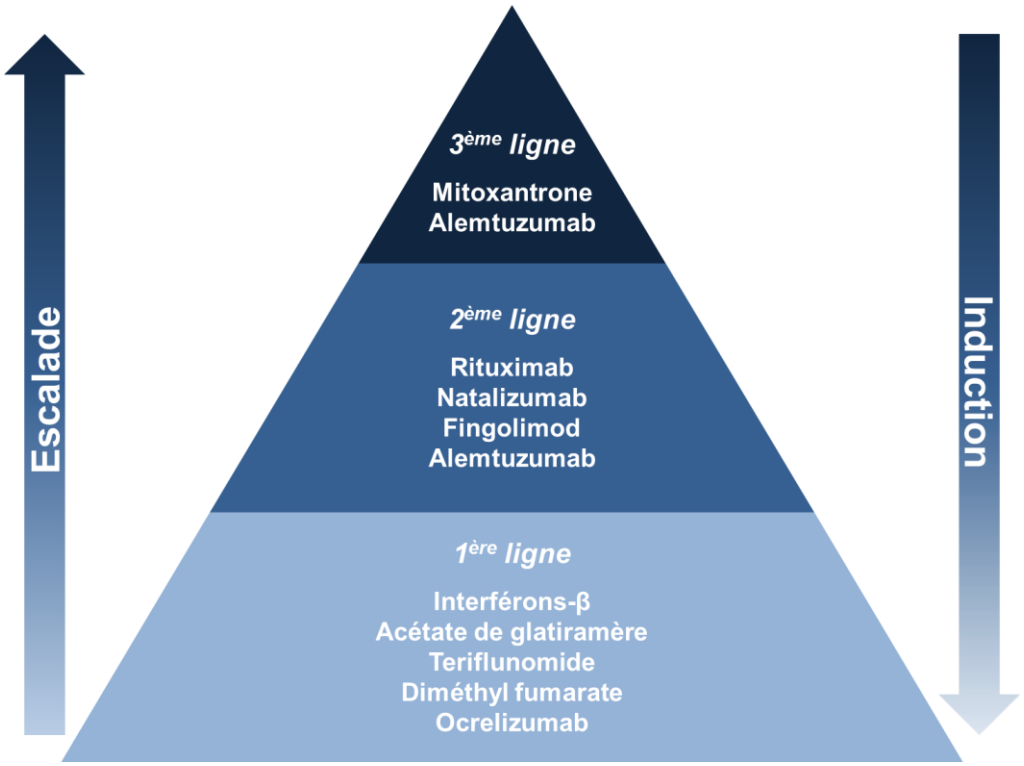


Figure 6-Représentation schématique des stratégies thérapeutiques dans la SEP

I.1.5.3. Effets secondaires des traitements de fond

Les traitements de fond disponibles à l'heure actuelle présentent comme tout traitement des effets indésirables pouvant être plus ou moins graves selon les principes actifs (29). Ainsi, les interférons ont pour effet secondaire principal un syndrome grippal suivant l'injection, les traitements oraux pouvant causer des nausées et douleurs abdominales, notamment pour le diméthyl fumarate, et la perte de cheveux concernant le teriflunomide. Les traitements de seconde et de troisième ligne entraînent quant à eux des effets indésirables semblables à ceux provoqués par les chimiothérapies.

Concernant les effets indésirables graves, pour les plus anciens traitements comme les interférons ou l'acétate de glatiramère, le recul et le nombre de patients traités en vie réelle sont suffisamment importants pour quantifier le risque et ainsi adapter les modalités de prescription, de suivi et de prise en charge. Ceci n'est pas le cas pour les traitements plus récents, pour lesquels quelques événements graves et rares ont été documentés (par exemple, 15 cas de LEMP (leucoencéphalopathie multifocale progressive) ont été observés suite à un traitement par fingolimod, alors que cet évènement indésirable grave n'a jamais été observé dans les essais thérapeutiques (15 cas sur 217 000 patients traités au 31 août 2017)) (32). Le fingolimod peut également entraîner des troubles du rythme cardiaque pouvant conduire au décès du patient. Le teriflunomide, quant à lui, peut conduire à des lésions hépatiques et est à fort risque tératogène, c'est-à-dire dangereux pour le fœtus dans une grossesse, nécessitant l'utilisation de la procédure d'élimination accélérée. A ce jour, ces deux molécules orales ainsi que la troisième (le diméthyl fumarate) font l'objet d'un plan de gestion de risques (PGR) à la demande des autorités de santé.

I.1.5.4. Traitements des poussées

Lors d'une poussée, un traitement peut être administré après confirmation diagnostic par un neurologue (33). Actuellement, le seul traitement est la délivrance de corticoïdes à forte dose. Ces derniers étant des anti-inflammatoires, ils permettent une diminution de l'intensité, de la gravité et de la durée de la poussée et peuvent accélérer la récupération suite à celle-ci. La posologie fréquemment utilisée est de 1g/jour de méthylprednisolone sur 3 à 5 jours consécutifs. Plusieurs modes d'administration du traitement sont désormais possibles : en bolus en intraveineuse à l'hôpital, avec le Solumedrol® 1000 mg, ou alors per os, avec du Medrol® 100 mg (34). Il est également possible d'effectuer des échanges plasmatiques, au nombre de 5 à 6 et s'étalant sur 2 à 3 semaines. Mais ce traitement est beaucoup moins fréquent avec seulement 5 à 10 patients par an par CHU (Centre Hospitalier Universitaire).

I.1.5.5. Traitements symptomatiques

En parallèle des traitements de fond de la SEP, des traitements symptomatiques peuvent être prescrits. Ils peuvent être délivrés tout au long de la maladie qu'importe la phase de cette dernière et traitent principalement les séquelles des poussées.

Traitements pharmacologiques

Un seul traitement pharmacologique est disponible sur le marché pour lutter contre les troubles de la marche, la fampridine (Fampyra®) (35). Son principe actif permet à l'influx nerveux de se propager pour stimuler les muscles et ainsi améliorer la marche. Le baclofène (Lioresal®) et le dantrolène (Dantrium®) sont deux molécules per os utilisées contre la spasticité (15,35,36). En cas d'échec de ces traitements, la tizanidine (Sirdalud®), en ATU uniquement, peut être délivrée (36). Lorsque la spasticité est localisée, des injections de toxine botulique peuvent être prescrites, alors que lorsqu'elle est diffuse et à partir d'un EDSS important (EDSS>7) le baclofène peut être administré par voie intrathécale au moyen d'une pompe implantée chez le patient avec rechargement de la substance active à l'hôpital (15,36). Les troubles sensitifs, notamment les douleurs neurologiques, peuvent être traités avec des spécialités anti-épileptiques (gabapentine, prégabaline, carbamazépine) ou anti-dépressives (duloxétine, amitriptyline) (35,36). Les anticholinergiques sous forme orale sont utilisés pour lutter contre les symptômes urinaires, comme l'oxybutynine ou la solifénacine dans le cadre de l'hyperactivité vésicale (incontinence), et les alpha-bloquants comme la tamsulosine dans le cadre de rétention urinaire due à des problèmes sphinctériens (15,35,36). Si les symptômes persistent, des injections intra-vésicales de toxine botulique peuvent être envisagées. Concernant les troubles intestinaux et sexuels, les traitements principalement proposés sont des laxatifs et des traitements du trouble de l'érection, respectivement (36).

Traitements non-pharmacologiques

En complément de médicaments, des traitements non-pharmacologiques peuvent être proposés aux patients (15,35,36). Ces derniers sont majoritairement proposés par les spécialistes de médecine physique et de réadaptation (MPR) au vu des symptômes présentés. Ainsi, des séances de kinésithérapie peuvent être prescrites pour lutter contre la spasticité ou les troubles de la marche. Des bains froids, ou des vestes réfrigérantes peuvent permettre de lutter contre les troubles sensitifs et la spasticité. Lorsque les troubles urinaires deviennent trop importants et que l'évacuation naturelle de l'urine n'est plus possible, le passage à l'auto-sondage peut être envisagé. De même, lorsque les troubles moteurs deviennent handicapants, des aides techniques à la marche (canne, déambulateur, véhicules pour handicapé physique (VHP) manuel ou électrique, ….) peuvent être prescrites. Des recours à des ergothérapeutes peuvent également être envisagés afin d'orienter le patient dans sa vie

quotidienne et d'aménager son domicile. Enfin, des stages dédiés à la gestion de la fatigue sont proposés par certains réseaux de soins spécialisés dans la SEP afin de permettre aux patients touchés de partager leur expérience.

I.1.6. Epidémiologie

Environ 2,3 millions de personnes ont une SEP dans le monde (37). Cette maladie affecte plus de femmes que d'hommes avec un sex-ratio de 2 à 2,5 femmes pour 1 homme (38–40). La SEP se déclare principalement chez le jeune adulte, âgé de 20 à 40 ans, il s'agit de la première cause de handicap non liée à un accident traumatique dans cette classe de la population (14). La maladie ne diminue en moyenne que de sept ans l'espérance de vie par rapport à la population générale (41), ce qui en fait une maladie au long cours.

La SEP est présente sur l'ensemble du globe, mais sa prévalence varie fortement selon la région (42). En France, deux études récentes ont estimé la prévalence de la SEP selon différents algorithmes de sélection des cas. Ainsi, en 2010, Fromont et al. ont estimé une prévalence française de 94.7 malades pour 100 000, soit 49 417 personnes atteintes, au 31 octobre 2004, en identifiant les cas au moyen du statut d'affection de longue durée (ALD) uniquement (43). Plus récemment, Foulon et al. ont évalué, quant à eux, la prévalence à 151.2 cas pour 100 000, soit 99 123 personnes touchées, au 31 décembre 2012 (40). Dans cette dernière étude, les cas ont été identifiés à partir de l'ALD, mais aussi des traitements de fond spécifiques à la SEP et des diagnostics relatifs aux séjours hospitaliers, permettant une identification plus importante des cas comparativement à la première étude. Ces deux études, ainsi qu'une étude plus ancienne de Kurtzke et al., montrent un gradient décroissant Nord-Est/Sud-Ouest en France (40,43,44). Quelques études ont été réalisées sur l'incidence de la SEP en France aux niveaux national et régional. La dernière au niveau national, effectuée sur les données d'ALD de l'Assurance Maladie, fait état d'une incidence de 7,5 pour 100 000 sur l'année 2004 (45). Deux études ont été réalisées au niveau régional, une en Bretagne et une en région Lorraine toutes deux basées sur des données issues de registres régionaux (46,47). L'étude lorraine estimait en 2008 une incidence brute de 4,4 pour 100 000 réévaluée à 8,5 pour 100 000 grâce à une méthode de capture-recapture appliquée sur les données du registre Lorrain (LORSEP - Réseau Lorrain pour la prise en charge de la Sclérose en Plaques) (47). L'étude bretonne avait quant à elle trouvé une incidence brute similaire de 4,3 pour 100 000 en 2000-2001 (46).

I.1.7. Etiologie

L'étiologie de la SEP reste pour une grande part inexpliquée, mais il est admis qu'il s'agit d'une maladie multifactorielle (48), dont on peut regrouper les causes en deux catégories majeures. D'un

côté, la SEP peut toucher plusieurs membres d'une même famille, suggérant une origine génétique. En effet, des études menées sur des jumeaux et des fratries ont montré une augmentation du risque de développer la maladie si un des membres l'avait (49,50). De plus, de nombreux gènes semblent être liés à la maladie, les majeurs étant dans la région HLA du chromosome 6 jouant un rôle dans la gestion de l'immunité, mais d'autres d'ordre mineur ont été identifiés ailleurs dans le génome (51,52). De l'autre, la variation importante de la prévalence de la SEP dans le monde suppose un rôle de l'environnement dans son étiologie (42). Une des causes environnementales probables, proposée dès 1974, est un déficit en vitamine D, dont la production est liée à l'exposition au soleil (53,54). En outre, d'autres facteurs environnementaux semblent augmenter le risque de déclencher la maladie, comme une infection au virus Epstein-Barr, appartenant à la famille des herpès, ou encore la consommation de tabac (55–58).

L'étiologie des poussées de la maladie est encore moins bien comprise. Les infections virales (59–61), l'exposition aux allergènes (62), le stress (63), la vaccination (64) et la grossesse (48) ont été étudiés et ont conduit à des résultats variables. Par ailleurs, les variations saisonnières des poussées pointent vers le rôle possible de paramètres météorologiques (65) et de la pollution de l'air (66–68).

I.1.8. Comorbidités

Les comorbidités physiques et mentales, c'est-à-dire d'autres maladies se déclarant avant ou après le diagnostic, sont communes dans la SEP. Ainsi, la prévalence vie entière de la dépression est estimée à 50% chez les patients ayant la maladie, soit plus de trois fois la valeur en population générale (69), et celle de l'anxiété à 36%, même si ces deux troubles restent sous-diagnostiqués (70). D'après une méta-analyse de Marrie et al. sur des études populationnelles (71), les prévalences de la dépression et de l'anxiété chez les individus ayant une SEP sont de 24% et 22%, respectivement. Les autres comorbidités les plus présentes sont l'hypertension avec une prévalence de 19%, l'hyperlipidémie et les maladies pulmonaires chroniques avec chacune une prévalence de 10% (71). Par ailleurs, le risque de fracture est augmenté chez les patients ayant une SEP en raison d'un risque accru d'ostéoporose et d'un risque élevé de chutes (69). L'ensemble de ces comorbidités peut avoir une influence sur l'histoire de la maladie (évolution du handicap (72–74), mortalité (75), poussées (76)), sur la prise en charge dans le cadre de soins (traitements (75), hospitalisations (77), recours aux soins (77), …) et sur la qualité de vie des patients (69).

I.2. Le parcours de soins

I.2.1. Concept

L'allongement de l'espérance de vie et les progrès constants de la médecine entraînent une hausse continue du nombre de personnes souffrant de maladie chronique (78). Ainsi, en France, 15 millions de personnes souffrent d'au moins une maladie chronique, soit un peu plus de 20% de la population (79). Ces affections nécessitent de nombreux soins médicaux et paramédicaux et peuvent avoir diverses conséquences sur la vie quotidienne et la vie professionnelle. Afin de prendre en compte la coordination nécessaire entre les différents professionnels de santé pour ces maladies chroniques, le concept de parcours de soins a été défini (80). Utilisé pour la première fois en 1985 (81), son but est d'être un outil pour l'amélioration continue de la qualité de soins en augmentant le rapport bénéfice-risque et la satisfaction des patients, en promouvant la sécurité des patients et en optimisant l'utilisation des ressources, d'après la définition donnée par l'Association Européenne sur les Parcours (*European Pathway Association*) (80,82). En France, cette notion a été pour la première fois évoquée avec la réforme Douste-Blazy en 2004 avec la création du parcours de soins coordonné et du concept de médecin traitant[1]. Depuis lors, la définition communément admise est « la prise en charge globale du patient et de l'usager dans un territoire donné au plus près de son lieu de vie, avec une meilleure attention portée à l'individu et à ses choix » (83–85). S'incluant dans des notions plus larges de parcours de santé et de parcours de vie, un parcours de soins se définit alors comme l'enchaînement pour un patient de différentes compétences professionnelles liées directement ou indirectement aux soins (consultations, actes techniques ou biologiques, traitements médicamenteux et non-médicamenteux, ...) (83) (Figure 7).

Ce concept reflète donc le besoin d'une coordination des professionnels de santé médicaux et paramédicaux, notamment entre les soins en établissement et en ambulatoire. Afin de faciliter cette coordination, la Haute Autorité de Santé (HAS) est chargée de définir des recommandations de parcours pour différentes maladies chroniques et affections de longue durée, ce qui a déjà été fait pour la maladie de Parkinson (86) ou la maladie rénale chronique (87) par exemple.

[1] Loi n° 2004-810 du 13 août 2004 relative à l'Assurance Maladie.

Figure 7-Représentation des différentes catégories de parcours selon les domaines considérés (84)

I.2.2. Affection de longue durée

Dès 1945, année de la création de l'Assurance Maladie en France, a été mis en place le dispositif spécifique d'ALD. Ce dernier a pour but de permettre la prise en charge des patients ayant une maladie chronique comportant un traitement prolongé et une thérapeutique particulièrement coûteuse. Au total, trente pathologies font partie de la liste des ALD, dont la SEP, correspondant à l'ALD numéro 25[2] (88). Cette dernière, auparavant permanente est à renouveler tous les cinq ans depuis 2011[3], et permet une exonération du ticket modérateur (part restant à charge à l'assuré après remboursement par l'Assurance Maladie) pour les frais liés à la maladie, c'est-à-dire une prise en charge à 100%. Afin de bénéficier l'ALD, au moins un des deux critères suivants doit être rempli :

- un traitement immunomodulateur de fond est prescrit à l'issue du bilan diagnostique, même en l'absence de handicap permanent,
- il existe un handicap permanent (parfois seulement constitué d'une asthénie ou de troubles cognitifs) nécessitant un traitement symptomatique et justifiant une prise en charge au long cours.

[2] Article D. 322-1 du code de la sécurité sociale, modifié par les décrets n°2004-1049 du 4 octobre 2004 publié au JO du 5 octobre 2004 et n°2011-77 du 19 janvier 2011 publié au JO du 21 janvier 2011.
[3] Décrets n°2011-74-75-77 du 19 janvier 2011et n°2011-726 du 24 juin 2011.

I.2.3. Le parcours de soins dans la SEP

I.2.3.1. Travaux de la Haute Autorité de Santé

Actuellement, il n'existe pas en France de recommandations précises concernant les parcours de soins pour les individus ayant une SEP. La HAS a édité en 2006 un guide affection de longue durée sur la maladie (89) recensant l'ensemble des professionnels pouvant être impliqués dans le parcours ainsi que les traitements de fond et symptomatiques (Figure 8). Cependant aucune recommandation concernant la fréquence de recours au neurologue, ou à d'autres spécialistes n'a été émise, faute de données de prise en charge à disposition. Ce document a eu pour suite la rédaction dix ans plus tard en 2015 d'un manuel récapitulant les actes et prestations pouvant être pris en charge dans le cadre de l'ALD SEP (90). Dans le cadre de la poursuite de son travail de recommandations de prise en charge des maladies chroniques, la HAS a inscrit dans son programme de travail 2018 un travail intitulé « Parcours du patient atteint de sclérose en plaques : Comment organiser la collaboration entre les différents intervenants du parcours » par auto-saisine (91). Les livrables concernant ce travail sont attendus pour la fin d'année 2018.

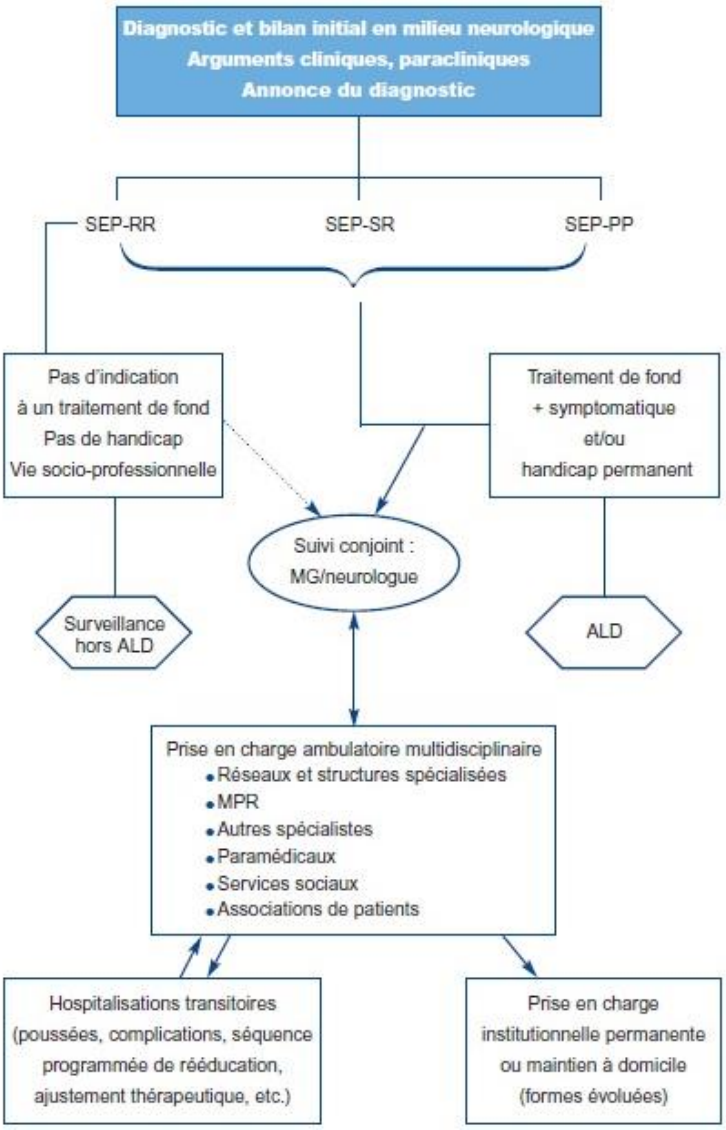


Figure 8-Proposition de prise en charge des patients SEP selon la HAS (2006) (89)

I.2.3.2. Le Plan Maladies Neurodégénératives

En 2014, le Ministère de l'Education Nationale, de l'Enseignement supérieur et de la Recherche et le Ministère des Affaires Sociales, de la Santé et des Droits des Femmes ont décidé de mettre un place conjointement un plan dédié aux maladies neurodégénératives dont fait partie la SEP. Ce Plan Maladies Neurodégénératives (PMND) lancé en novembre 2014 s'étend sur la période 2014-2019 et s'inscrit dans la Stratégie Nationale de Santé et la Stratégie Nationale de la Recherche (92). Cette démarche nouvelle, car englobant plusieurs maladies « modèles » et non une seule comme auparavant avec le plan Alzheimer par exemple (93), repose sur une démarche d'intégration sanitaire et médico-sanitaire, de décloisonnement et sur une démarche participative des patients et de leurs représentants (94). Au travers de quatre axes, dont un dédié à la continuité du parcours de santé, 12 enjeux et 96 mesures concrètes et communes aux trois maladies « modèles » (la SEP, la maladie d'Alzheimer et la maladie de Parkinson) sont proposés.

Centres de ressources et de compétences SEP

Parmi les mesures influençant la prise en charge de la SEP, la mesure 11 a conduit à la création des centres de ressources et de compétences SEP (CRCSEP)[4], correspondant à une officialisation des centres experts de la SEP déjà existants. Ces centres basés dans 23 CHU ont pour but de regrouper les professionnels de santé travaillant sur la maladie autour de trois thématiques : la prise en charge médicale, médico-sociale et la recherche. Le rapprochement de ces professionnels peut permettre la mise en place de consultations pluridisciplinaires comme c'est déjà le cas depuis 1996 à Rennes avec la clinique de la SEP (95) et dans d'autres centres ensuite.

Réseaux de santé SEP

Les réseaux de santé SEP, le plus souvent basés au niveau régional et au nombre de 17 en France (96), ont pour objectif de favoriser et de coordonner l'accès aux soins du patient, de le rendre acteur et partenaire dans la prise en charge de sa maladie, et de développer une approche globale à la fois médicale et psychosociale des conséquences au quotidien de la maladie (95). Le but de ces réseaux est double. Pour les professionnels, le réseau permet la mise en place de protocoles de soins régionaux et propose des formations. Pour les patients, le réseau propose des séances d'éducation thérapeutique mais aussi du soutien psychologique et des réunions d'informations pour les malades et leur entourage. Une hétérogénéité forte entre les réseaux dans les actions menées a pu être remarquée suite à un état des lieux réalisé dans le cadre du PMND. Ce dernier a conduit à une volonté d'harmonisation

[4] Instruction n° DGOS/R4/2016/176 du 27 mai 2016 relative aux modalités de labellisation des centres de ressources et de compétences SEP.

des objectifs vers de l'appui aux pratiques professionnelles pour aller progressivement vers des pôles de ressources régionales[5] proposant l'ensemble des actions citées ci-dessus.

Outils épidémiologiques

Afin d'améliorer la connaissance de l'histoire naturelle de la maladie, de l'utilisation des traitements et de la prise en charge clinique, le PMND encourage également l'utilisation de la base de données française EDMUS (European Database for Multiple Sclerosis) de l'Observatoire Français de la Sclérose En Plaques (OFSEP) (97). Cette base multicentrique contient les informations cliniques sur près de 50 000 personnes ayant une SEP en France représentant un socle riche de connaissances. Néanmoins, elle contient majoritairement des données cliniques (score EDSS, traitements délivrés, poussées,…) et donc aucune donnée sur le recours aux soins autre qu'auprès du neurologue, limitant son utilisation dans l'étude des parcours de soins.

I.2.3.3. Prise en charge au niveau international

Depuis une dizaine d'années, le concept de « comprehensive care », c'est-à-dire de soin global, proche de celui d' « integrated care », correspondant à la notion de parcours de soins en français, se développe dans le cadre de la SEP au niveau international. Ce concept repose sur une prise en charge pluridisciplinaire à la fois médicale, mais également cognitive, psycho-sociale et paramédicale de la SEP (98–100). On peut alors parler d'approche holistique, c'est-à-dire considérer la maladie comme un tout et non chaque item (symptômes, traitements, rééducation, …) la constituant séparément les uns des autres.

Certains pays ont déjà défini des référentiels de prise en charge, comme par exemple l'Espagne qui a défini un ensemble documentaire basé sur une matrice-temps recensant l'ensemble des examens, traitements et consultations effectuables selon la phase de la maladie (diagnostic, suivi post-diagnostic, …) (101).

I.2.3.4. Prise en charge en France

Comme évoqué dans les différents guides de la HAS (79,80), de nombreux professionnels de santé sont susceptibles d'intervenir dans la prise en charge de la SEP (Figure 9). Cette prise en charge pluridisciplinaire vise à mieux comprendre l'évolution de la maladie afin de proposer à chaque patient un programme de soins personnalisé (95).

[5] Instruction n°SG/PMND/DGOS/DGS/DGCS/2017/56 du 15 février 2017.

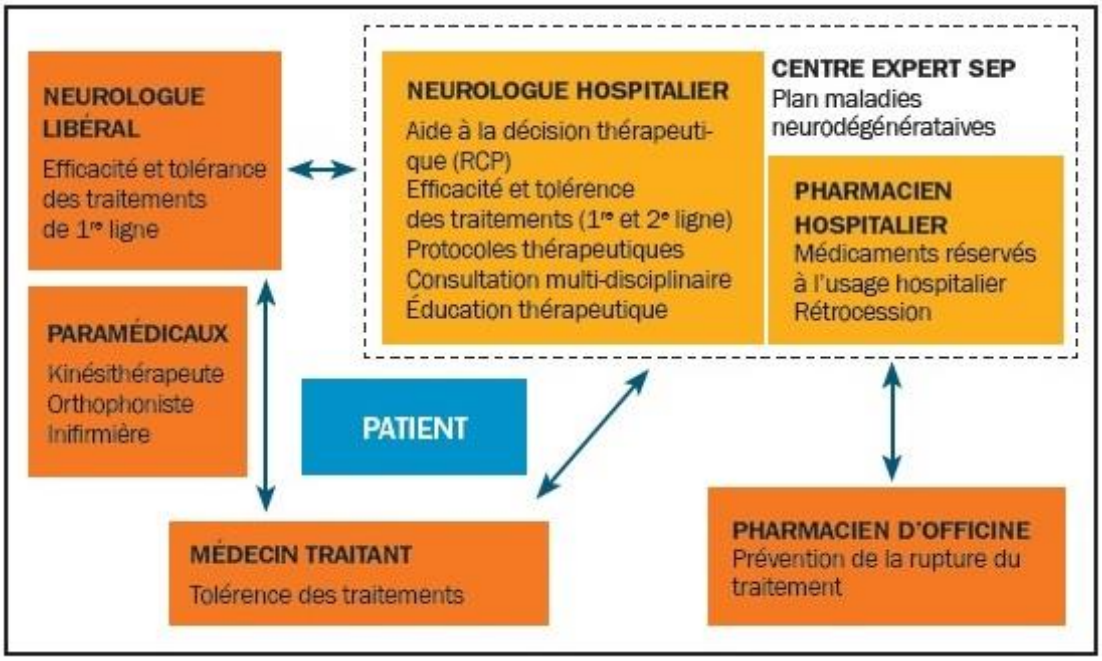


Figure 9-Schéma de l'organisation des soins chez le patient SEP (95)

La prise en charge médicale

Le duo médecin généraliste-neurologue

Parmi les professionnels médicaux, deux constituent la pierre angulaire de la prise en charge : le médecin généraliste et le neurologue. En effet, ils interviennent tous deux dès le début de la maladie, le généraliste étant le plus souvent le premier à observer les symptômes et à orienter vers le neurologue, seul capable de poser le diagnostic (89,102). Le neurologue, pour lequel la HAS recommande au moins une visite annuelle (90), est le référent dans la SEP. Par ailleurs, il est le seul habilité à prescrire un traitement de fond, tous réservés à une prescription neurologique. Le médecin généraliste, quant à lui, permet un suivi au quotidien du patient notamment concernant les symptômes, la tolérance des traitements et la prise en charge des autres problèmes de santé.

Etant donné l'étendue des pathologies du système nerveux, tout neurologue ne peut pas développer un haut niveau d'expertise vis-à-vis de la SEP. C'est pourquoi, la majorité des neurologues français ne suit qu'une dizaine de patients ayant une SEP parmi l'ensemble de sa patientèle. Dans ce cas, il peut faire appel à un neurologue hospitalier expert de la SEP situé dans l'un des CRCSEP pour un avis ponctuel, pour une instauration de traitement ou une confirmation du diagnostic par exemple. Cette complémentarité dans la décision permet alors au patient de bénéficier de la meilleure prise en charge, du meilleur traitement et aussi d'accéder aux essais cliniques éventuellement.

Autres spécialistes impliqués dans le parcours de soins

Parallèlement à ce duo médecin généraliste/neurologue, d'autres spécialistes peuvent être impliqués notamment concernant le traitement des symptômes de la maladie ou dans le cadre des PGR des traitements.

Le spécialiste de MPR est un de ceux le plus sollicité en raison du handicap fréquemment entraîné par la SEP. En créant un programme personnalisé, il a pour but d'aider au maintien de la qualité de vie et de l'autonomie du patient (102). Ainsi, son rôle consiste à optimiser la récupération suite aux poussées, aider à la conservation de la marche, et gérer la spasticité et les troubles urinaires et sexuels. Afin de garantir cette prise charge, il coordonne l'ensemble des professions paramédicales (kinésithérapeute, infirmier, ergothérapeute, orthophoniste, …) et peut également orienter le patient vers d'autres spécialistes. Par ailleurs, lorsque le handicap moteur s'installe chez un patient, il est souvent orienté vers un MPR qui est alors l'interlocuteur privilégié pour la prescription d'une aide technique à la marche.

Dans le cadre de troubles urinaires, un patient peut être amené à avoir recours à un urologue en plus du MPR, en cas de nécessité d'actes invasifs (injection de toxine botulique dans la vessie, chirurgie de la vessie). Des recommandations de prise en charge spécifiques à la SEP ont par ailleurs été définies concernant le traitement des troubles urinaires (Figure 10). Par ailleurs, un ophtalmologue peut intervenir dans la phase diagnostique en cas de symptômes initiaux affectant la vision (névrite optique) ou tout au long de l'histoire naturelle de la SEP. Le recours à l'ophtalmologue est également motivé par le suivi d'effets secondaires liés à certains traitements de fond (fingolimod). La prise en charge psychologique ou psychiatrique est souvent nécessaire dans la SEP, via le recours aux psychologues, neuropsychologues, aussi impliqués dans le dépistage et le suivi des troubles cognitifs, ou aux psychiatres.

Enfin, le pharmacien est également un interlocuteur dans le parcours de soins puisqu'il dispense la majorité des traitements de fond et des traitements symptomatiques. Son rôle est important sur l'éducation thérapeutique, la gestion des effets indésirables et l'observance.

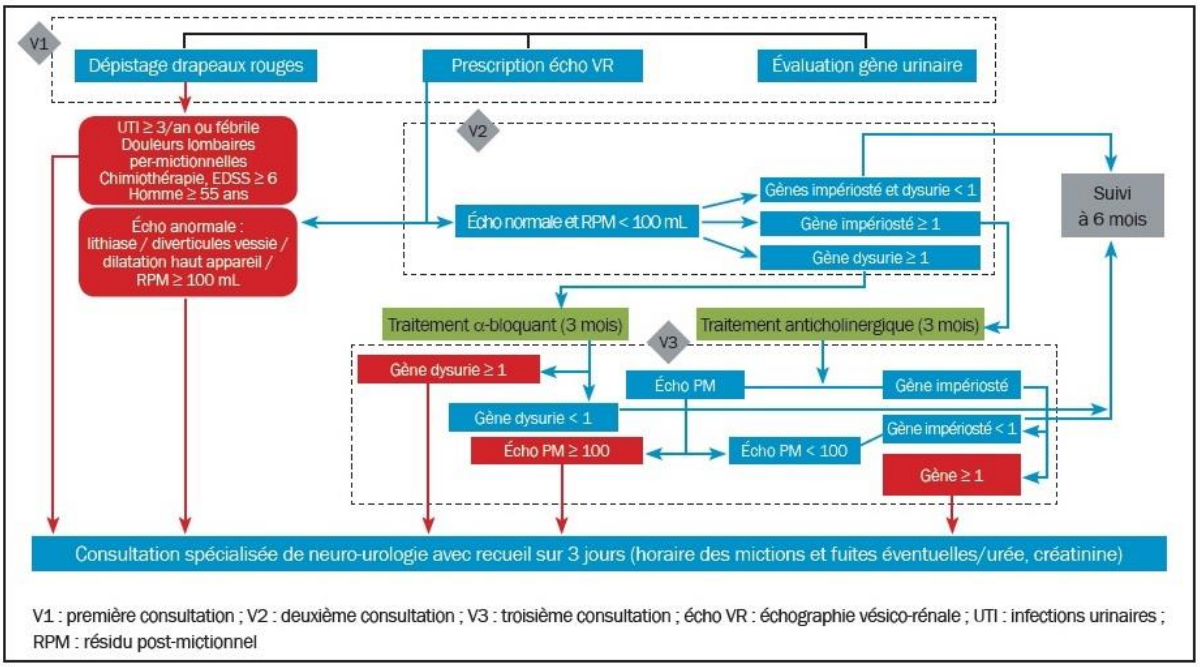


Figure 10-Algorithme FLUE-MS (First Line Urological Evaluation in Multiple Sclerosis) permettant l'évaluation et le traitement des troubles urinaires dans la SEP (95)

La prise en charge paramédicale

Les troubles moteurs étant très fréquents dans la SEP, le masseur-kinésithérapeute en lien avec le MPR les évalue et propose au patient une prise en charge de rééducation afin qu'il puisse maintenir son autonomie (102). Son action englobe également la majeure partie des symptômes de la SEP, puisqu'il prend en charge la gestion de la spasticité, des troubles de l'équilibre, de la préhension et de la fatigue (95). Son intervention peut apparaître très tôt dans la maladie dès qu'une gêne apparaît (103).

L'infirmier diplômé d'état (IDE) a également une place majeure dans le parcours de soins de la SEP. Pouvant intervenir dès l'annonce diagnostique, l'IDE peut prendre le relais du neurologue dans l'explication de la maladie et de ses conséquences, la reconnaissance des poussées, la gestion de la maladie au quotidien, et des symptômes notamment dans les CRCSEP. Au travers de l'éducation à la maladie, il enseigne au patient la prise des traitements de fond, notamment les injectables et les perfusions en établissements de santé, l'observance, la gestion de la fatigue mais également la pratique de l'auto-sondage en cas de troubles urinaires (104). L'IDE peut intervenir également dans la réalisation de soins à domicile, notamment l'administration des traitements, ou la prévention des escarres en cas de handicap avancé par exemple.

I.2.4. Recours aux soins

En France, peu de données sont disponibles sur la consommation de soins des patients ayant une SEP. A notre connaissance, seule une étude de l'Assurance Maladie, réalisée sur 90 288 individus ayant une SEP identifiés dans les bases médico-administratives, fait état du recours aux soins de ces patients en 2013 (105). Le nombre médian de consultations avec un généraliste et un neurologue étaient de 6 et 2, respectivement. Cependant, aucune comparaison avec la population française ou l'international n'a été effectué, l'article se concentrant sur les coûts.

Au niveau international, plusieurs études sur le recours aux soins ont été effectuées (77,106–109). Il apparaît sans surprise que les patients SEP ont un recours aux soins plus important que la population générale. Ainsi, une étude canadienne basée sur les bases médico-administratives (107) a montré qu'en 2008, le nombre moyen de consultations médicales était de 12,9 par an pour les personnes ayant une SEP versus 8,4 en population générale (RR (Risque Relatif) 1,53, IC95% (Intervalle de Confiance à 95%) [1,52-1,55]). La différence significative persistait après ajustement sur l'âge et le sexe. Dans les 5 années précédant le diagnostic de la SEP, le nombre de consultations annuel était déjà significativement plus élevé dans le groupe SEP, mais de façon moins importante (RR 1,15, IC95% [1,10-1,21]). Un résultat concordant a été retrouvé plus récemment par Wijnands et al. sur une base plus importante au niveau national suggérant un prodrome de la maladie (108). L'étude de Pohar et al. basée sur des données déclaratives a également montré un recours plus important des patients ayant une SEP avec 8,4 consultations en moyenne par an avec un spécialiste versus 4,5 en population générale (106). Des différences de prise en charge entre patients ont également été mises en évidence au Canada entre les patients suivis dans des cliniques spécialisées et ceux ne l'étant pas (109). En effet, les patients n'étant pas suivis dans ces établissements avaient plus de consultations toutes causes et plus d'hospitalisations que ceux suivis dans les cliniques (IR (Incidence Ratio) 1,14, IC95% [1,08-1,20] et 1,72 [1,56-1,90], respectivement). Par ailleurs, plusieurs facteurs peuvent influencer le recours aux soins, notamment la présence d'un handicap (110) ou de comorbidités (77,111,112), et la fatigue (77) augmentant le recours alors qu'une bonne observance thérapeutique le diminue (113).

I.2.5. Coûts

D'un point de vue économique, le nombre de personnes atteintes par la SEP, le prix des traitements ainsi que le fort recours aux soins entraînent un coût élevé en constante augmentation pour l'Assurance Maladie. Il représentait en France 1,083 milliards d'euros en 2013, et 1,296 milliards en 2015 (environ 0,8% du budget total de l'Assurance Maladie), dont environ 65% de ce budget dédiés au remboursement des soins en ville (114,115). Cette augmentation est liée entre autres à l'arrivée des nouveaux traitements. Ainsi les dépenses liées à ces derniers ont augmenté de 62 millions d'euros et 33 millions respectivement en 2013 et en 2014 (116,117). De plus, une étude économique de Lefeuvre

et al. a montré que 38% des dépenses étaient liés aux traitements de fond et 9% aux traitements symptomatiques (105). Le coût moyen par patient est estimé à 11 920 euros par la CNAMTS (Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés) (115) avec une forte hétérogénéité dans la répartition par types de dépenses en fonction de l'âge notamment. Ainsi, la part des coûts représentée par les traitements diminue avec l'âge (71% entre 20 et 29 ans *vs* 12% après 80 ans) alors qu'à l'inverse les parts des hospitalisations (de 15% (20-29 ans) à 35% (70-79 ans)) et des soins paramédicaux (de 2% entre 20 et 29 ans à 24% entre 70 et 79 ans) augmentent (105). Des disparités de coûts selon le niveau de handicap ont également été observées sur de plus petits échantillons (22 600 € pour EDSS 0-3 *vs* 48 100 € pour EDSS 7-9) (118).

I.3. L'étude quantitative des parcours de soins

L'étude des parcours de soins est en plein essor depuis les dix dernières années en France (119–127) et au niveau international (101,128–130). En effet, dans un but de meilleure prise en charge et d'efficience, une meilleure connaissance des parcours de soins est un préalable à l'atteinte d'une situation idéale dans laquelle une population reçoit les bons soins par les bons professionnels dans les bonnes structures au bon moment, le tout au meilleur coût. Par ailleurs, caractériser les parcours de certaines populations offre la possibilité de vérifier leur conformité vis-à-vis des recommandations de prise en charge et de vérifier leur application. Il est donc nécessaire de disposer d'outils visant à décrire et comparer ces parcours, d'identifier des bases de données permettant de les construire ainsi que des méthodes statistiques pour les analyser.

I.3.1. Bases à disposition

Afin de construire le parcours de soins d'un patient, deux sources de données peuvent être employées : des données recueillies auprès du patient (129) ou des professionnels de santé (120) ou des données médico-administratives (119,122,128,131). Les bases médico-administratives, dont la principale en France est le SNDS (Système National des Données de Santé), sont ouvertes depuis quelques années à la recherche et leur utilisation est encouragée (132), notamment par la HAS (133). Elles sont exhaustives et concernent de grandes populations, alors que le recueil auprès des professionnels ou des patients ne permet de toucher qu'un faible échantillon et expose à différents problèmes méthodologiques dont le biais de mémoire. A l'inverse, le recueil direct permet d'avoir plus de données sur les caractéristiques démographiques, cliniques et socio-économiques, que dans les bases médico-administratives où ces données sont souvent absentes, excepté le sexe et l'âge.

I.3.2. Méthodes d'analyse des parcours

Malgré un certain engouement autour du concept de parcours de soins, la majorité des études citées précédemment n'analysent pas réellement les parcours à proprement parler mais plutôt le recours aux soins. Ainsi, il s'agit de descriptifs du nombre de recours à des spécialistes (129) et de ré-hospitalisations au cours d'une période donnée (123), ou encore du délai entre différentes consultations (120). Le parcours de soins étant défini comme une succession de recours, une méthodologie adaptée prenant en compte la dimension temporelle doit être employée. Seules quatre études en France (124–126,128) ont utilisé à notre connaissance une méthodologie adaptée. Ainsi, trois méthodes issues à l'origine du domaine de l'informatique ou de la bio-informatique ont ainsi pu être relevées : le *mining multidimensional itemsets sequential patterns* (126), la *symbolic data analysis* (125), et l'analyse de séquences (*state sequence analysis*) (124,128). Les deux premières méthodes précédemment citées sont complexes à mettre en œuvre, car nécessitant des connaissances en programmation informatique et elles n'ont été testées que sur des données exclusivement hospitalières. *A contrario*, l'analyse de séquences a été utilisée sur des données ambulatoires et hospitalières et bénéficie d'une prise en main plus rapide et d'une large bibliographie en sciences sociales (134,135). Cependant, les deux études épidémiologiques identifiées n'ont pas clairement défini la procédure à suivre, ni montré les forces et limites de la méthode pour l'étude des parcours de soins.

I.4. Problématique de la thèse

Le contexte de prise en charge de la SEP a considérablement changé au cours des dernières décennies avec l'apparition des nouvelles thérapeutiques, la montée d'une prise en charge pluridisciplinaire et le développement de politiques en faveur d'une harmonisation des pratiques. L'ensemble de ces changements pose plusieurs questions dans le champ de la santé publique, en particulier de l'accès aux soins, de la consommation de soins et de pharmaco-épidémiologie. Par ailleurs, l'ouverture des bases médico-administratives aux chercheurs est un des outils permettant la réponse à ces questions.

L'arsenal thérapeutique à disposition entraîne de potentielles modifications et des variations de pratiques médicales et de prise en charge qui restent à évaluer. En effet, le choix du traitement par le neurologue est rendu plus compliqué du fait des multiples molécules disponibles, des indications qui se recoupent en partie et des niveaux de risque différents et plus ou moins bien évalués et/ou connus à l'heure actuelle. Le but du médecin est de « choisir le bon traitement pour le bon patient au bon moment ». Le spécialiste faisant face à plus de possibilités, il doit comparer les rapports bénéfices-risques pour le patient. A ce jour, ces médicaments font l'objet d'un PGR à la demande des autorités de santé, ce qui justifie l'étude de données observationnelles en vie réelle de prescription et d'utilisation et l'évaluation de leur impact sur la consommation de soins et sur les hospitalisations. De plus, la prescription des traitements de fond de la SEP est partiellement connue en France et un

descriptif précis de la situation actuelle permettrait de mieux comprendre les séquences de traitement et possiblement d'améliorer la prescription des traitements pour les neurologues ou d'aider à la prise de décision.

En outre, la SEP entraîne à long terme un handicap principalement moteur. Comme évoqué précédemment, ce handicap modifie le type de prise en charge dont va bénéficier un patient avec alors une augmentation des soins paramédicaux et de l'utilisation de traitements symptomatiques. De plus, lorsque le handicap est installé et que le patient entre dans une phase progressive de la maladie, une baisse de l'utilisation des traitements de fond est à prévoir. Ces conséquences modifient alors l'ensemble du recours aux soins et les coûts qui y sont associés, ce qui justifie pleinement le fait que le niveau de handicap soit considéré dans la majorité des analyses sur l'accès et le recours aux soins aux niveaux français et international. Cependant, actuellement, il n'existe aucun indicateur permettant de quantifier le handicap dans les bases de données médico-administratives telles que le SNDS. La construction d'un tel indicateur serait utile dans ce travail sur le recours aux soins mais également dans d'autres cadres, tels que l'étude de la dépendance et de la perte d'autonomie, ou maladies.

La SEP nécessite une prise en charge pluridisciplinaire du patient, faisant intervenir plusieurs professionnels de santé. Cependant, aucune donnée de recours aux soins des personnes ayant une SEP en France n'est disponible au niveau national à notre connaissance. De plus, aucun parcours de soins clairement défini n'existant pour la maladie, un descriptif de ces derniers permettrait de fournir des données quantitatives objectives pour aider à leur élaboration. La maladie évoluant en différentes phases, il est probable qu'il existe des différences dans la prise en charge et donc dans les parcours de soins.

I.5. Objectifs

I.5.1. Objectif principal

L'objectif principal de cette thèse est d'étudier les parcours de soins des personnes ayant une SEP en France extraites du SNDS afin de décrire les consommations de soins et de mettre en évidence des typologies de parcours. Afin d'atteindre cet objectif, le travail s'est déroulé en plusieurs étapes :

(i) L'identification des cas de SEP en France dans le SNDS,

(ii) L'extraction et la création pour ces personnes des éléments représentatifs de leur recours aux soins,

(iii) La création des parcours de soins grâce à l'application d'une méthode innovante dans le champ de la recherche en santé : l'analyse de séquences,

(iv) Le regroupement en groupes homogènes de ces parcours,

(v) La caractérisation de cette typologie en termes de recours aux soins,

(vi) L'identification de facteurs individuels ou contextuels liés à la typologie de parcours.

I.5.2. Objectifs secondaires

En parallèle de cet objectif principal, deux autres axes de recherche secondaires ont été définis afin de fournir des données permettant la caractérisation de la typologie de parcours de soins. Le premier axe a porté sur les traitements de fond de la SEP et leur utilisation (fréquence et place dans la séquence thérapeutique). Une étude a également été effectuée sur la faisabilité du suivi dans l'EGB (Echantillon Généraliste des Bénéficiaires) du PGR pour le fingolimod (Gilenya®). Le second axe a concerné la mesure du handicap dans le SNDS, c'est-à-dire étudier la faisabilité d'estimer le niveau de handicap moteur dans les bases médico-administratives en l'absence de données cliniques.

II. Matériels et méthodes

II.1. Bases médico-administratives-Définitions, accès et structures

II.1.1. Présentation des bases médico-administratives françaises

II.1.1.1. Le Système National des Données de Santé (SNDS)

L'ensemble du travail de la thèse a été mené sur les données issues du SNDS. Ce dernier a été créé par la loi de modernisation du système de santé du 26 janvier 2016[6] et est une évolution de l'ancien Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM), lui-même mis en place à partir de 1998 (131). Il recueille prospectivement et de manière systématique l'ensemble des données individuelles de remboursement de soins des principaux régimes de sécurité sociale en France. Ces données sont d'ordre médical (statut ALD, diagnostics médicaux), mais sont principalement d'ordre économique, d'où l'emploi du terme de base médico-administratives. En effet, les buts du SNDS ont été élargis à sa création en 2006 par rapport à ceux du SNIIRAM, dont l'objectif principal était de permettre une gestion comptable pour les régimes de sécurité sociale. Le SNDS a ainsi pour but de contribuer à la définition et l'évaluation des politiques de santé, des pratiques et des parcours de soins afin de permettre une meilleure compréhension de la nature des dépenses de santé et d'informer les différents professionnels de santé sur leur activité.

Le SNDS, actuellement en déploiement, sera composé de cinq bases de données qui seront chaînées les unes aux autres au niveau individuel : le Datamart de Consommation Inter-Régimes (DCIR), le Programme de Médicalisation des Systèmes d'Information (PMSI), tous deux formant initialement le SNIIRAM, auxquelles viendront s'ajouter la base Inserm (Institut national de la santé et de la recherche médicale) des causes médicales de décès (CépiDC-Centre d'épidémiologie sur les causes médicales décès), la base de données médico-sociales des Maisons Départementales Pour le Handicap (MDPH) et un échantillon représentatif des données de remboursement par les mutuelles (Figure 11). A ce jour, seules les trois premières bases (SNIIRAM, PMSI et CépiDC) sont effectivement chaînées. Le travail de cette thèse ayant eu lieu de 2015 à 2018 a quant à lui été réalisé seulement sur les données du SNIIRAM (DCIR et PMSI).

---

[6] Article 193 de la loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé.

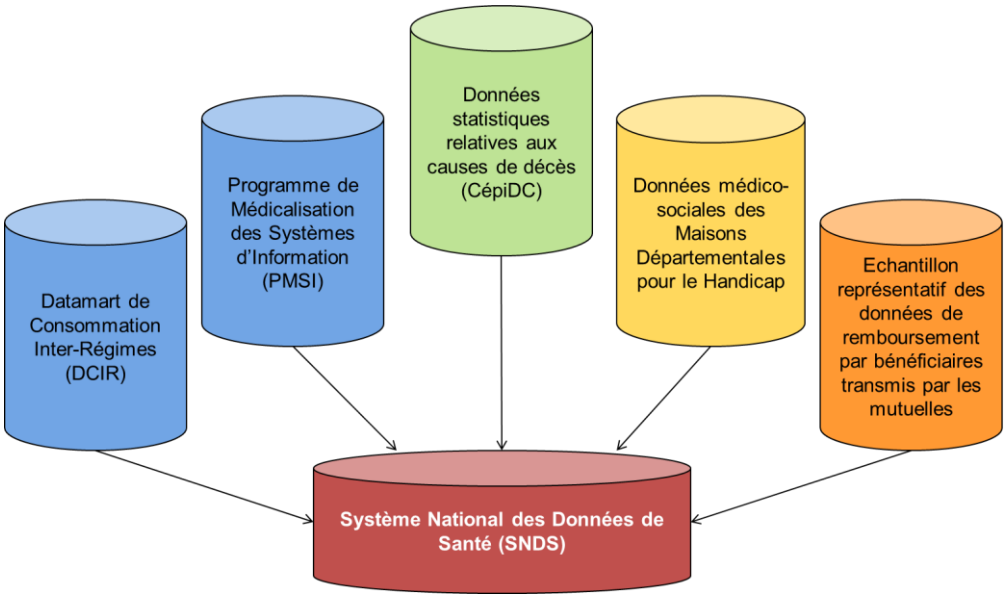


Figure 11-Synthèse des bases constitutives du SNDS en France

II.1.1.2. Le Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM)

Créé par la loi de financement de la Sécurité Sociale du 23 décembre 1998[7], le SNIIRAM correspond au recueil de l'ensemble des remboursements de soins effectués en ville, stockés dans le DCIR et gérés par la CNAMTS et à l'hôpital, rassemblés dans le PMSI et gérés par l'ATIH (Agence Technique de l'Information sur l'Hospitalisation). Ses objectifs principaux fixés par l'arrêté ministériel du 11 avril 2002[8] étaient d'améliorer la qualité des soins, contribuer à une meilleure gestion de l'Assurance Maladie, contribuer à une meilleure gestion des politiques de santé et transmettre aux prestataires de soins les informations pertinentes relatives à leur activité. Il s'agit d'une base quasi-exhaustive pseudonymisée avec un recueil systématique et exhaustif des consommations chaînées de soins individuelles sur tous les régimes couverts. Le SNIIRAM est qualifié de recueil systématique puisque les données sont récupérées par le biais des feuilles de soins et des cartes vitales sans l'intervention du patient lui-même. Afin de prendre en compte les remboursements dus à l'envoi et au traitement des feuilles de soins, un délai de 27 mois suivant la fin de la période d'étude, correspondant au délai maximal pour le remboursement d'un acte de soins, est conseillé dans le cadre de l'utilisation des données. Ces dernières sont exhaustives concernant l'ensemble des soins remboursés par l'Assurance

[7] Article 21 de la loi n° 98-1194 du 23 décembre 1998 codifié aux articles L. 161-28-1 à L. 161-28-4 du code de la sécurité sociale.
[8] Arrêté du 11 avril 2002 relatif à la mise en œuvre du système national d'information inter-régimes de l'Assurance Maladie.

Maladie. Elles sont pseudonymisées, c'est-à-dire que le numéro de sécurité sociale du patient est anonymisé mais ses caractéristiques en termes d'âge et de sexe sont préservées (136).

En France, l'ensemble de la population française se répartit entre différents régimes d'Assurance Maladie :

- Le Régime Général y compris les Sections Locales Mutualistes (SLM – Fonctionnaires d'Etat, étudiants, fonctionnaires hospitaliers et territoriaux),
- La Mutuelle Sociale Agricole (MSA),
- Le Régime Social des Indépendants (RSI),
- Le régime de la SNCF (CPRPSNCF – Caisse de Prévoyance et de Retraite du Personnel de la SNCF (Société Nationale des Chemins de Fers français)),
- Le régime de la RATP (CCAS RATP – Caisse de Coordination aux Assurances Sociales de la RATP (Régie Autonome des Transports Parisiens)),
- Le régime des marins et inscrits maritimes (ENIM – Etablissement National des Invalides de la Marine),
- Le régime des mines (CANSSM – Caisse Autonome Nationale de la Sécurité Sociale dans les Mines),
- Le régime des militaires (CNMSS – Caisse Nationale Militaire de Sécurité Sociale),
- Le régime des clercs et employés de notaire (CRPCEN – Caisse de Retraite et de Prévoyance des Clercs et Employés de Notaire),
- Le régime du Port autonome de Bordeaux (CPPAB – Caisse de Prévoyance du Port Autonome de Bordeaux),
- Le régime des ministres du culte (CAVIMAC – Caisse d'Assurance Vieillesse, Invalidité et Maladie des Cultes),
- Le régime d'EDF-GDF (CAMIEG – Caisse d'Assurance Maladie des Industries Electriques et Gazières),
- Le régime de la Banque de France (CPM Banque de France – Caisse de Prévoyance Maladie de la Banque de France),
- Le régime de l'Assemblée Nationale (FSSAN – Fonds de Sécurité Sociale de l'Assemblée Nationale),
- Le régime du Sénat (RA3S – Régime Autonome de Sécurité Sociale du Sénat).

Le SNIIRAM (et donc par extension le SNDS) couvre l'ensemble de ces régimes hormis les régimes des marins, des mines, de la SNCF, de la RATP, ceux de l'Assemblée Nationale et du Sénat et celui du Port autonome de Bordeaux correspondant à un peu plus de 2% de la population française. On peut donc dire que le SNIIRAM est une base quasi-exhaustive en France, couvrant actuellement 97% de la

population française. Par ailleurs, les principaux régimes de l'Assurance Maladie étant concernés, un patient pouvant passer d'un régime à un autre au cours de sa vie, il est peu probable qu'il sorte de la base de données, limitant ainsi les perdus de vue. Il faut également noter que pour des raisons de confidentialité et pour éviter des pressions sur les personnes ayant un rôle politique, les données des personnes affiliées aux régimes du Sénat et de l'Assemblée Nationale ne sont pas mises à disposition pour analyse ou consultation dans le SNIIRAM.
Cette base de données offre donc de nombreux avantages : quasi-exhaustivité de la population, absence de perdus de vue pendant le suivi, données souvent fiables.

II.1.1.3. Le Datamart de Consommation Inter-Régimes (DCIR)

Comme évoqué précédemment, le DCIR recense l'ensemble des données de consommation de soins remboursées effectuées en ville, c'est-à-dire hors établissements de santé publics et ESPIC (Etablissement de santé privé d'intérêt collectif). Concernant les soins effectués dans les établissements de santé privés à but lucratif, plus communément appelés cliniques, les actes et consultations externes (ACE-passage dans l'établissement de santé pour une période inférieure à 8 heures) sont tracés dans le DCIR. Quant aux informations concernant les séjours hospitaliers effectués dans les cliniques (passage dans l'établissement de santé pour une période supérieure à 8 heures), elles sont retranscrites à la fois dans le DCIR et dans le PMSI.

Les données contenues dans le DCIR couvrent les consultations avec l'ensemble des professionnels de santé médicaux et paramédicaux, les délivrances de médicaments remboursés, de matériels techniques, les versements d'invalidités et d'indemnités journalières ou encore les transports pris en charge dans le cadre des soins. L'historique à disposition pour ces données est de 4 ans, dont l'année en cours, avec un archivage pour 10 ans. Sont présentes également dans le DCIR des données médicales propres à chaque patient, comme les ALD, les accidents du travail et les maladies professionnelles (Table 1).

Table 1-Informations disponibles et indisponibles dans les données du DCIR en 2016 (adapté de (131))

Information disponible	Information indisponible
Informations relatives au patient	*Autres caractéristiques socio-économiques*
Année et mois de naissance	Emploi, revenus, statut marital,…
Sexe	*Facteurs de risque*
Date exacte de décès	Tabagisme, historique familial,…
Statut CMU, CMU-C, AME	*Résultats d'examens cliniques*
Statut AAH	Pression artérielle, taille, poids,…
Commune de résidence	Radiographie, Anatomo-pathologie,…
Indice de défavorisation sociale	*Consultations médicales et paramédicales*
ALD	Motifs diagnostics de recours
Pensions d'invalidité	Actes non facturés
Maladie professionnelle/Arrêts de travail	*Médicaments*
Informations relatives aux remboursements (date des soins et nombre de procédures)	Médicaments non remboursés ou non-prescrits
Médicaments	Observance du patient
Procédures	Posologie pour certains médicaments
Professionnels de santé (spécialité, mode d'exercice, …)	
Actes de biologie	
Aides techniques	

CMU : Couverture Maladie Universelle. CMU-C : Couverture Maladie Universelle Complémentaire. AME : Aide Médicale de l'Etat. AAH : Allocation aux Adultes Handicapés. ALD : Affection de Longue Durée.

II.1.1.4. Le Programme de Médicalisation des Systèmes d'Information (PMSI)

Le DCIR est chaîné via un identifiant unique avec les différentes bases du PMSI, correspondant aux séjours hospitaliers et ACE effectuées dans les établissements de santé publics et les ESPIC, et les séjours hospitaliers effectués dans les établissements de santé privés à but lucratif. Les bases du PMSI, gérées par l'ATIH, sont au nombre de quatre et correspondent chacune à un type de soins. Ainsi, le PMSI MCO (Médecine Chirurgie Obstétrique et Odontologie), est le plus important des quatre et correspond aux séjours de soins de courte et moyenne durées par opposition aux trois autres bases du PMSI portant sur les séjours de longue durée. Il contient les informations des séjours hospitaliers de 12,2 millions de patients pour 10,9 millions de séjours en hospitalisation complète (hors ambulatoire et séances) au niveau national en 2016 (137). Le PMSI SSR (Soins de Suite et de Réadaptation)

correspond majoritairement aux séjours de convalescence, pour maladie à évolution prolongée, ou encore pour rééducation fonctionnelle et réadaptation. Enfin, le PMSI PSY, aussi appelé Recueil d'Informations Médicalisé pour la Psychiatrie (RIM-P), correspond aux hospitalisations psychiatriques et aux internements, et le PMSI HAD aux soins liés aux hospitalisations à domicile. Dans ce travail, seuls les PMSI MCO et SSR ont été utilisés. Les PMSI HAD et PSY ne seront donc pas détaillés ultérieurement.

Pour chacune de ces bases, le détail des soins dispensés au cours du séjour hospitalier (actes d'imagerie, de laboratoire, …) ainsi que lors des ACE est disponible. Par ailleurs, des informations propres au patient sont spécifiques à certaines bases, comme le degré de dépendance mesuré dans le SSR ou l'indice de Karnofsky (138) codant pour l'état général du patient disponible dans le HAD. L'ensemble de ces informations est disponible sur un historique de 10 ans.

II.1.1.5. L'Echantillon Généraliste des Bénéficiaires (EGB)

L'EGB est un échantillon au 1/97ème du SNIIRAM, soit environ 680 000 personnes, créé en 2005[9] (139–141). Les personnes inclues dans cet échantillon sont sélectionnées par rapport à la clé suivant leur numéro d'inscription au répertoire des personnes physiques (NIR) unique. Cette dernière correspond au reste de la division du nombre formé par les treize chiffres du NIR modulo 97. Les bénéficiaires de l'EGB sont alors ceux ayant une clé, parmi les 97 clés possibles allant de 01 à 97, identique à la valeur confidentielle choisie à la création de l'échantillon. Créé en 2004, il contiendra à terme les données de consommation de soins sur une période de 20 ans rétrospectifs. L'EGB est représentatif de la population française au niveau national pour l'âge et le sexe (140). Cependant, il n'est pas représentatif à une échelle plus fine, comme la région ou le département. Le modèle conceptuel est le même que celui du SNIIRAM avec quelques simplifications notamment dans la jointure des tables. L'EGB est une cohorte dynamique. Ainsi, l'échantillon est mis à jour de manière trimestrielle avec l'ajout de nouveaux bénéficiaires, la sortie de ceux décédés depuis plus de 27 mois, correspondant au délai maximal pour le remboursement d'un acte de soins, et la modification des informations concernant les régimes de rattachement des bénéficiaires. Contrairement au SNIIRAM, seul le PMSI MCO est à disposition dans l'échantillon, limitant ainsi les données hospitalières disponibles.

[9] Arrêté du 20 juin 2005 paru au Journal Officiel du 19 août 2005.

II.1.2. Contexte réglementaire

II.1.2.1. Accès aux données

Afin de pouvoir répondre à des questions de recherche d'« intérêt public », les bases de données de remboursement de l'Assurance Maladie sont ouvertes au monde de la recherche. Deux modalités d'accès sont possibles : un accès à distance via un portail sécurisé en ligne ou un accès physique sécurisé via une extraction de la base. Pour accéder au moyen de l'une ou l'autre des modalités, un processus réglementaire doit être entrepris par le chercheur. Ce dernier a évolué avec la création du SNDS et de nouvelles instances. Par conséquent, le processus décrit ci-après, qui a été suivi pour l'obtention des données de la thèse, est désormais obsolète.

Dans le cadre d'une extraction physique des données, c'est-à-dire une extraction du SNDS basée localement et donc hors portail, le parcours réglementaire est le suivant. Premièrement, l'étude doit être validée par l'Institut des Données de Santé (IDS), devenu depuis Institut National des Données de Santé (INDS), sur le critère de son « intérêt public ». Si besoin, chaînage avec une base clinique externe notamment, un avis scientifique peut être demandé au CCTIRS (Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé), remplacé actuellement par le CEREES (Comité d'Expertise pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé). Suite à ces avis, la CNIL (Commission Nationale Informatique et Libertés) doit se prononcer sur le projet, notamment au niveau de l'exploitation des données et de la sécurité de ces dernières. Dans le cas d'un avis favorable de la CNIL, une convention de mise à disposition des données est alors rédigée entre la CNAMTS et l'investigateur principal du projet afin de cerner les modalités de l'extraction avant une mise à disposition des données.

S'agissant de notre étude, un protocole a été rédigé dans un premier temps. Ce dernier contenait une description des données demandées pour une extraction physique et précisait les analyses à réaliser. Ce protocole a été soumis à l'IDS en date du 17 mai 2016, après relecture et avis de plusieurs experts externes au comité de suivi de la thèse (Figure 12). Parallèlement, ce protocole a été soumis à la CNIL le 19 mai 2016 (Demande d'autorisation n°1959035). L'accord de l'IDS a été obtenu le 25 mai 2016 (Décision d'approbation n°191) sans demande d'avis au CCTIRS du fait de la relecture externe par des experts effectuée en amont. L'accord CNIL a été obtenu le 21 mars 2017 (Décision DE-2017-026) avec une signature de la convention entre l'EHESP (Ecole des Hautes Etudes en Santé Publique) et la CNAMTS le 23 août 2017 pour une mise à disposition effective des données le 12 octobre 2017, soit 17 mois après le début de la procédure.

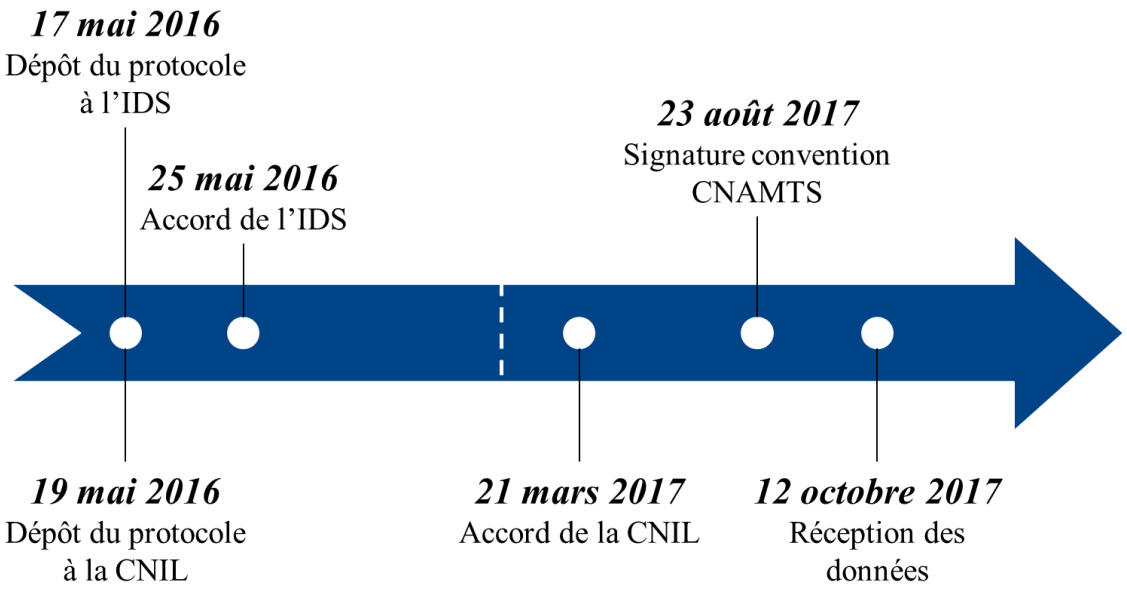


Figure 12-Calendrier synthétisant la démarche réglementaire d'accès aux données du SNDS pour ce travail de thèse

En l'attente de la livraison des données, le travail des première et deuxième années de thèse a été effectué sur l'EGB (Décision d'approbation n°143 de l'IDS en date du 8 septembre 2015).

II.1.2.2. Données sensibles

Même si les données du SNDS sont pseudonymisées, une ré-identification des patients est toujours possible au moyen de variables décrites comme étant sensibles. Ces dernières sont au nombre de quatre, à savoir le mois de naissance, la commune de résidence, la date exacte de décès et la date exacte de soins. Afin d'empêcher le croisement des variables entre elles, des profils d'accès ont été créés, chacun permettant de ne travailler que sur une seule donnée sensible à la fois. Lorsqu'une donnée sensible est choisie, les autres variables restent à disposition mais à une granularité plus large. Ainsi, l'utilisateur a toujours accès à l'année de naissance, le département de résidence, le mois de décès et le mois de chaque soin.

Dans le cadre de cette thèse sur les parcours de soins, il est apparu important d'avoir à disposition la granularité de temps la plus fine à disposition. C'est pourquoi, le travail sur l'EGB a été réalisé en utilisant le profil donnant accès aux dates exactes de soins. Les autres variables à disposition ont donc été l'année de naissance, le département de résidence et le mois de décès. L'extraction physique du SNDS est quant à elle une extraction brute de la base globale. Par conséquent, l'ensemble des variables sensibles est disponible. Lors de la signature de la convention avec la CNAMTS et dans la déclaration CNIL, les investigateurs se sont engagés à ne pas croiser ces informations. Cependant,

elles peuvent être utilisées séparément afin de collecter le plus d'informations possibles sur les patients.

II.1.3. Structure du SNDS

Le SNDS est une base de données complexe rentrant dans le cadre du *big data*. En effet, elle respecte les 5V caractéristiques de ce type de base (142), à savoir :

- le volume (plus de 8,9 milliards de remboursements traités en 2015 et plus de 20 milliards de lignes disponibles dans l'historique),
- la variété (plusieurs centaines de variables à disposition sous différents formats avec chacune une table de valeurs de référence),
- la vélocité ou rapidité (flux quotidien de réception des données pour le régime général, mensuel pour les autres régimes et mensuel avec consolidation annuelle pour les données du PMSI),
- la véracité (présence de plusieurs contrôles qualité tout au long du processus permettant de garantir une bonne qualité des données (131)),
- la valeur (données de consommations de soins exhaustives utiles pour l'étude des parcours de soins individuels, la pharmaco-vigilance, les études médico-économiques entre autres).

Un ensemble de termes spécifiques au vocabulaire de l'Assurance Maladie sont liés au SNDS et nécessitent d'être définis pour une meilleure compréhension des choix effectués et des algorithmes développés dans ce travail.

Chaque ligne contenue dans le DCIR correspond à une prestation, c'est-à-dire à tout acte de soins conduisant à un remboursement. Le SNDS étant une base économique, des erreurs dans le remboursement d'un soin peuvent avoir lieu se traduisant par des régularisations dans la base. Ces dernières consistent en l'ajout d'une nouvelle ligne annulant la prestation en erreur et ayant les mêmes caractéristiques excepté des montants de remboursement, des nombres d'actes ou quantités délivrées négatifs, et en l'ajout d'une ou plusieurs nouvelles lignes correspondant au remboursement régularisé.
Chaque prestation est liée à un acte de base, c'est-à-dire l'acte nécessitant un remboursement, comme une consultation avec un spécialiste, la réalisation d'une glycémie, ou une échographie cardiaque par exemple. A ces actes peuvent s'ajouter des majorations, telles que les majorations de nuit ou d'urgence, ou des compléments d'actes ayant entraîné une surfacturation de ce dernier. Ces derniers donnent alors lieu à la création d'une ou plusieurs autres lignes de prestations codant chacune pour un complément ou une majoration de l'acte de base. Le nombre de lignes ne correspond donc pas au nombre total de prestations effectuées.

II.1.3.1. Référentiel des consommants

Les informations concernant les individus sont rassemblées dans une base commune au DCIR et au PMSI, appelée référentiel des consommants et notée *IR_BEN_R*. Ce référentiel ne contient que quelques informations démographiques basées sur le NIR du patient, tels que l'année et le mois de naissance et le sexe. D'autres informations mises à jour régulièrement sans sauvegarde dans un historique les complètent, notamment la commune de résidence, l'organisme de sécurité sociale d'affiliation et la date de décès le cas échéant, d'après les données du RNIPP (Répertoire National d'Identification des Personnes Physiques). Chaque ligne du référentiel possède deux clés uniques, l'identifiant anonyme du patient dans le SNDS (BEN_NIR_PSA) et le rang gémellaire du patient, permettant de distinguer les jumeaux. Or, plusieurs lignes peuvent correspondre à un même patient du fait de l'évolution de ses droits à la sécurité sociale (passage d'ayant-droit à assuré principal par exemple). Par conséquent, une clé unique, basée sur le couple BEN_NIR_PSA et rang gémellaire, a été créée afin de distinguer les patients uniques dans le référentiel. Cette dernière correspond au NIR anonymisé du patient, d'où son nom BEN_NIR_ANO, et sert de jointure entre l'ensemble des données portant sur le patient. Cette clé indispensable dans une analyse longitudinale pour comptabiliser l'ensemble des éléments de recours aux soins du parcours d'un individu.

En complément du référentiel patient, le DCIR contient un référentiel médicalisé, appelé *ER_IMB_F*, qui permet de compléter les données sur les patients avec des données d'invalidité et d'ALD. Contrairement au premier, ce dernier recense les informations sous forme d'historique permettant ainsi de voir l'évolution des informations pour chaque patient. Ces deux référentiels sont joints grâce à la clé BEN_NIR_ANO.

II.1.3.2. Structure du DCIR

Le DCIR a été construit sur le format d'une base en étoile avec une table centrale : la table des prestations. Autour de cette dernière, gravitent 13 tables affinées thématiques permettant d'ajouter de l'information à la table principale (Figure 13). Par exemple, la table des prestations contient une ligne indiquant qu'il y a eu une prestation de délivrance de médicament(s) dans une officine, et la table affinée pharmacie permet de connaître le(s) médicament(s) facturé(s) ainsi que le nombre de boîtes délivrées pour chacun. La table des prestations et les tables affinées sont jointes au moyen de 9 clés de jointure techniques permettant l'identification unique d'une prestation. Chaque prestation peut ensuite être rattachée à un unique patient au moyen du référentiel des consommants via une jointure sur BEN_NIR_ANO.

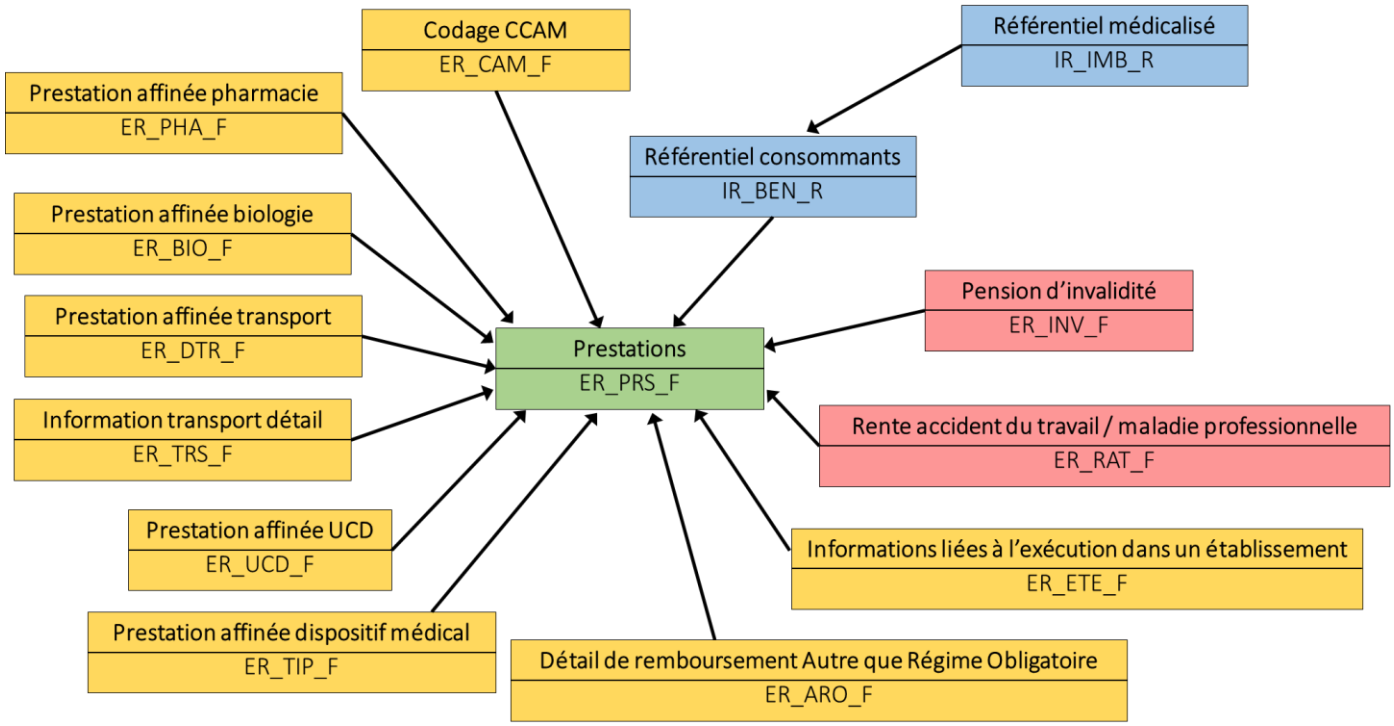


Figure 13-Modèle conceptuel du DCIR (143)

Durant la thèse, les tables d'intérêt principales ont été les tables *ER_PRS_F* (prestations médicales), *ER_PHA_F* (table affinée pharmacie), *ER_BIO_F* (table affinée biologie), *ER_CAM_F* (table affinée des actes CCAM (Classification Commune des Actes Médicaux)), *ER_TIP_F* (table affinée des dispositifs médicaux), et *ER_UCD_F* (table affinée de la pharmacie hospitalière codée et rétrocédée et médicaments en sus du GHS (Groupe Homogène de Séjour) explicité ci-dessous).

II.1.3.3. Structure du PMSI

La structure des bases du PMSI est semblable à celle du DCIR, à la différence qu'il existe deux structures indépendantes similaires, correspondant d'un côté aux séjours hospitaliers, ayant pour centre la table des séjours, et de l'autre aux ACE, avec pour table centrale la table des prestations ACE.

Concernant le codage des séjours dans la base, il diffère entre les deux PMSI considérés dues aux particularités du PMSI SSR, car les patients peuvent y recevoir des soins pouvant s'étaler sur plusieurs mois voire années. Ainsi, l'unité de codage dans le MCO est le séjour du patient, c'est-à-dire qu'une ligne dans la table principale du PMSI MCO correspond à un séjour unique. Dans le SSR, l'unité de codage est la semaine, un séjour pouvant s'étendre sur plusieurs années. Ainsi dans la table principale du SSR, un séjour est codé sur plusieurs lignes, chacune d'entre elles correspondant à une semaine

calendaire d'hospitalisation. Ce découpage temporel différent implique des traitements différents dans les bases pour les analyser ensuite conjointement.

Structure du PMSI MCO

La table des séjours comprend des informations sur l'identité du séjour avec le numéro du résumé de sortie anonyme (RSA) et le numéro FINESS (Fichier National des Établissements Sanitaires et Sociaux) de l'établissement hospitalier. Sont aussi renseignées entre autres les dates d'entrée et de sortie du séjour (*EXE_SOI_DTD* et *EXE_SOI_DTF* respectivement), le diagnostic principal (DP) et si besoin le diagnostic relié (DR) de ce dernier et le groupe homogène de malades (GHM) codant pour le coût du séjour et permettant de savoir la raison médicale majeure de l'hospitalisation. Les principales tables pouvant être jointes à la table des séjours concernent les actes CCAM, les diagnostics associés significatifs (DAS), les informations sur les unités médicales (UM), ou encore les médicaments délivrés en sus du GHS, étant une sous-division du GHM. Ces dernières sont mises en rapport grâce à un triplet unique de clés de jointure : le numéro du RSA, l'année de sortie et le numéro FINESS de l'établissement dans lequel le séjour a eu lieu.

La table des prestations ACE contient la date d'exécution des soins (*EXE_SOI_DTD*), la spécialité du praticien (*SPE_COD*) ainsi que la cotation de l'acte effectué (*ACT_COD*). Quant aux tables pouvant être jointes à la table des prestations ACE, elles portent principalement sur les actes CCAM, les actes de biologie, les médicaments délivrés en sus ainsi que les DMI (Dispositifs médicaux implantables). Ces dernières sont jointes grâce au numéro de séquence et à l'année d'exécution de la prestation et au numéro FINESS de l'établissement de santé.

Structure du PMSI SSR

Le PMSI SSR a une structure identique à celle du MCO, la principale différence provenant du fait que les séjours sont découpés en semaine et qu'un patient peut ne venir que pour certaines journées de la semaine, ce qui est spécifique aux domaines de la réadaptation. Ainsi, la table des séjours comprend des informations sur l'identité du séjour avec le numéro du résumé hebdomadaire de sortie anonymisé (RHA) et le numéro FINESS de l'établissement hospitalier. Concernant la présence d'un patient pour des soins SSR, sont présentes les dates d'entrée et de sortie du séjour global (*EXE_SOI_DTD* et *EXE_SOI_DTF* respectivement), une variable permettant de savoir le délai depuis l'entrée du séjour global et donc de repérer la semaine considérée (*RHS_ANT_SEJ_ENT*), et deux variables codant pour les jours de présence d'un patient le week-end (*JP_WE*) et hors week-end (*JP_HWE*). Les informations portant sur les raisons des soins sont contenues dans les variables *MOR_PRP* (manifestation morbide principale) et *ETL_AFF* (affection étiologique), la première codant pour le

trouble à l'origine des soins et la seconde pour la maladie à l'origine de ce trouble (dans notre cas la SEP). La finalité de la prise en charge (*FP_PEC*), correspondant au groupe de diagnostic des actes de réadaptation, est également disponible ainsi que le groupe médico-économique, équivalent du GHM du MCO dans le SSR.

Les principales tables pouvant être jointes à la table des séjours concernent les actes CCAM, les DAS, les actes CSARR (Catalogue Spécifique des Actes de Rééducation et Réadaptation) ou encore les médicaments délivrés en sus du GHJ (Groupe Homogène Journalier). Ces dernières sont mises en rapport grâce au triplet unique (RHA, numéro FINESS, année d'exécution).
Les ACE n'ont pas pu être considérés dans ce travail faute de temps.

Jointure aux données du DCIR
Afin de faire le lien avec les données du DCIR, les données du PMSI sont rattachées à un patient grâce au référentiel des consommants. Cependant, la seule clé de jointure disponible dans le PMSI est le numéro d'identité anonyme SNIIRAM, BEN_NIR_PSA, et non le couple (BEN_NIR_PSA, rang gémellaire), ne permettant pas une assignation unique de BEN_NIR_PSA et donc d'une identité. Pour garantir qu'un acte de soins (séjour hospitalier ou ACE) reste propre à un patient, si la jointure conduisait à un doublon de l'acte à cause de plusieurs identités possibles, l'acte était exclu de l'analyse conformément aux recommandations.

II.1.3.4. Jointure à des bases externes

Le SNDS recense une grande quantité de données de natures différentes, tels que des diagnostics, des actes de biologie ou des informations sur les spécialités pharmaceutiques. Le plus souvent ces informations sont codifiées grâce à des nomenclatures nationales ou internationales qu'il convient alors de joindre aux données pour pouvoir en exploiter leur plein potentiel et obtenir par exemple le libellé des actes. La majorité des variables des bases sont propres à l'Assurance Maladie. Par conséquent, des tables de correspondance sont mises à disposition par cette dernière afin de pouvoir comprendre leur signification, comme la codification de la nature des prestations, ou de la spécialité du praticien via la norme NOEMIE (Norme Ouverte d'Échanges entre la Maladie et les Intervenants Extérieurs) (144) par exemple.

D'autres bases thématiques externes au SNDS car employées dans le domaine médical par les professionnels de santé sont aussi utilisées (Table 2). Ainsi, les diagnostics des ALD recensés dans la table *IR_IMB_R* sont issus de la CIM-10 (Classification Internationale des Maladies – 10ème version) (145). La NGAP (Nomenclature Générale des Actes Professionnels) (146) sert au codage des actes et des majorations des spécialistes mais aussi du domaine paramédical. Les actes et gestes médicaux

(imagerie, échographie, sutures, …) de la table *ER_CAM_F* sont codés grâce à la CCAM (147), les actes de biologie rassemblés dans la table *ER_BIO_F* grâce à la NABM (Nomenclature des Actes de Biologie Médicale) (148) et les matériels techniques et dispositifs implantables contenus dans la table *ER_TIP*_F à la LPP (Liste des Produits et Prestations remboursables) (149). Les spécialités médicamenteuses n'étant codées que selon leur code CIP (Code Identifiant de Présentation) dans la table *ER_PHA_F*, la table *IR_PHA_R* peut être utilisée afin d'identifier la classe ATC (Classification Anatomique, Thérapeutique et Chimique) ou le nombre de comprimés par boîte par exemple.

Il en est de même dans les tables du PMSI. En effet, les différents diagnostics sont également codés au moyen de la CIM-10 (145). Les établissements de santé sont quant à eux identifiés via leur code FINESS (150) disponibles dans une base en ligne.

Table 2-Synthèse des jointures entre tables du SNDS et bases de connaissances externes

Table du SNDS	Source de données SNDS	Contenu de la table	Base externe
ER_TIP_F	DCIR	Prestations affinées dispositif médical	*LPP*
ER_CAM_F	DCIR	Prestations affinées actes médicaux	*CCAM*
ER_BIO_F	DCIR	Prestations affinées actes biologiques	*NABM*
ER_PHA_F	DCIR	Prestations affinées pharmacie	*IR_PHA_R*
IR_IMB_R	DCIR	Référentiel bénéficiaire médicalisé	*CIM-10*
MCO_C & SSR_C	PMSI	Chaînage MCO & SSR	*FINESS*
MCO_B, MCO_D & SSR_B, SSR_D	PMSI	Séjours, Diagnostics associés MCO & SSR	*CIM-10*

II.2. Population d'étude

II.2.1. Algorithme de sélection

Le design choisi pour la thèse peut s'apparenter à celui d'une cohorte rétrospective. En effet, les personnes avec une SEP ont été identifiées au moyen d'un algorithme défini par l'Assurance Maladie sur une période d'inclusion fixée à l'image d'un recrutement dans une cohorte. Les patients décédant durant la période d'étude sont conservés, tout comme les patients entrant durant la période d'inclusion, constituant ainsi une cohorte dynamique. Les données de consommation de soins ont été extraites pour la période d'étude considérée selon les données à disposition.

L'algorithme d'identification proposé par l'Assurance Maladie (40) est formé par quatre critères :

- L'existence d'une ALD active pour la SEP (ALD 25),
- La présence d'au moins une délivrance en ville ou dans un établissement (ATU incluse) d'un médicament traceur de la SEP, c'est-à-dire un traitement délivré uniquement dans le cadre de la maladie (Annexe B),
- La présence d'au moins un séjour hospitalier dans le PMSI MCO ayant pour DP, DR ou DAS le code CIM-10 *G35* correspondant à la SEP, y compris dans les données concernant les unités médicales,
- La présence d'au moins un versement d'une pension d'invalidité pour SEP.

Une personne est considérée comme ayant la SEP si elle présente au moins un des critères, le coordinateur logique entre tous les critères étant le « ou ». Dans notre cas, nous avons fait le choix de ne garder que les trois premiers critères (ALD, médicament traceur et hospitalisation). En effet, Lefeuvre et al. dans des communications dans deux congrès en 2016 ont indiqué que le critère portant sur l'invalidité ne permettait pas d'identifier de nouvelles personnes (151,152), d'où le choix de l'occulter.

Deux sources de données ayant été utilisées dans le cadre de cette thèse (EGB puis SNDS), deux populations d'étude ont été constituées sur deux périodes d'inclusion différentes.

II.2.2. Identification dans l'EGB

Concernant l'étude des parcours de soins, les bénéficiaires de l'EGB atteints de SEP ont été identifiés le 22 juin 2015 sur la période d'inclusion allant du 1er janvier 2007 au 31 décembre 2012 selon les trois critères définis précédemment. Les traitements traceurs considérés étaient les interférons-β, l'acétate de glatiramère, le fingolimod et le natalizumab. Les prestations exécutées ainsi que les séjours hospitaliers relatifs à ces derniers sur la période d'étude s'étalant du 1er janvier 2007 au 31 décembre 2013 ont par la suite été extraits à partir de la date de la première identification du bénéficiaire par un des critères. Une troncature à gauche des données du PMSI a été réalisée selon la date de première exécution de la première prestation de chacun des bénéficiaires, afin d'avoir des données complètes de consommation de soins sur une période comparable entre les deux sources de données. L'extraction de prestations a été conduite jusqu'à fin 2013 afin de s'assurer qu'une majorité des remboursements aient été effectués vis-à-vis de la durée maximale d'acceptation des feuilles de soins, 27 mois pour rappel. Les bénéficiaires ne présentant aucune consommation de soins (prestation ou séjours hospitaliers) sur l'ensemble de la période d'étude ont été exclus (n=3). Pour cette population première population, un total de 1 000 bénéficiaires SEP a été obtenu (Table 3).

Concernant l'étude de faisabilité du suivi du PGR, les bénéficiaires identifiés étaient ceux pour lesquels une délivrance de fingolimod était constatée sur la période s'étalant du 1er janvier 2011 au 31 décembre 2014. Au total, 70 patients ont été repérés dans l'EGB, pour lesquels les consommations de soins (prestations et séjours hospitaliers) ont été extraites du 1er janvier 2007 au 31 mars 2015 (Table 3).

Une population a également été construite pour la phase de développement de l'algorithme permettant de quantifier le niveau de handicap moteur dans les bases médico-administratives. Il s'agissait des patients issus de la population construite pour le recours aux soins qui était identifiés en 2007, correspondant à 650 patients, afin de s'assurer d'avoir une durée de suivi nécessaire pour pouvoir développer l'algorithme. Pour ces 650 bénéficiaires, les prestations ont été considérées du 1er janvier 2007 au 31 mai 2013 (Table 3).

Table 3-Synthèse des populations sources construites dans l'EGB et des données de consommation correspondantes

Population d'étude	Période d'identification de la SEP et algorithme utilisé	Données de consommation
Recours aux soins *(N= 1000)*	1er janvier 2007 au 31 décembre 2012 via algorithme CNAMTS	1er janvier 2007 au 31 décembre 2013
Etude de faisabilité PGR *(N=70)*	Identification de la SEP du 1er janvier 2011 au 31 décembre 2014 via délivrance de fingolimod	1er janvier 2007 au 31 mars 2015
Algorithme handicap *(N=650)*	Identification de la SEP du 1er janvier 2007 au 31 décembre 2007 via algorithme CNAMTS	1er janvier 2007 au 31 mai 2013

II.2.3. Identification dans le SNDS

II.2.3.1. Algorithme

Les patients du SNDS avec la SEP ont été identifiés par la DEMEX (service en charge des extractions à la CNAMTS) en octobre 2017 sur la période d'inclusion allant du 1er janvier 2010 au 31 décembre 2015 selon les mêmes critères que précédemment. Les traitements traceurs considérés étaient identiques à ceux utilisés pour l'EGB en ajoutant les traitements mis sur le marché entre 2013 et 2015, à savoir le teriflunomide, le diméthyl fumarate et le peginterféron. Les prestations exécutées ainsi que les séjours hospitaliers des quatre PMSI relatifs à ces derniers sur la période d'étude s'étalant du 1er janvier 2007 au 31 décembre 2015 ont par la suite été extraits (plus de 100 gigaoctets de données). La

période d'étude pour le SNDS s'étendait du 1er janvier 2010 au 31 décembre 2015, les données antérieures permettant de calculer des paramètres au démarrage de la période d'étude.

II.2.3.2. Identification des patients uniques

Un travail non nécessaire pour la population de l'EGB a été effectué sur celle issue du SNDS afin de permettre l'identification unique des personnes ayant une SEP. En effet, comme évoqué précédemment un patient peut avoir plusieurs identités (numéros enquêtés) dans le SNDS, il faut donc lui attribuer son numéro unique (BEN_NIR_ANO) à partir du fichier cohorte (Figure 14).

- Sur les 134 277 couples (NUM_ENQ, Rang gémellaire) uniques identifiés, seuls 123 640 (92.1%) avaient une correspondance avec un numéro unique ANO, conduisant à 114 418 BEN_NIR_ANO,
- Parmi ces derniers, 15 NIR ANO ont été identifiés comme correspondant à plusieurs couples (NUM_ENQ, Rang gémellaire) discordants sur les principales caractéristiques démographiques, soit l'année de naissance et le sexe. Ainsi, un même NIR ANO correspondait par exemple à un père et à sa fille, entraînant deux sexes et deux années de naissance possibles pour un même individu. Ces 15 BEN_NIR_ANO ont donc été exclus de la population d'étude,
- Par ailleurs, 34 BEN_NIR_ANO présentaient des dates de décès incohérentes,
- En croisant avec les informations de consommation de soins disponibles dans la table *ER_PRS_F*, 2 BEN_NIR_ANO ont été exclus car présentant des prestations de remboursement postérieures à leur date de décès,
- Après ces ajustements au niveau de l'unicité des patients, 114 401 BEN_NIR_ANO et donc patients uniques ont été identifiés dans les données fournies par l'Assurance Maladie.

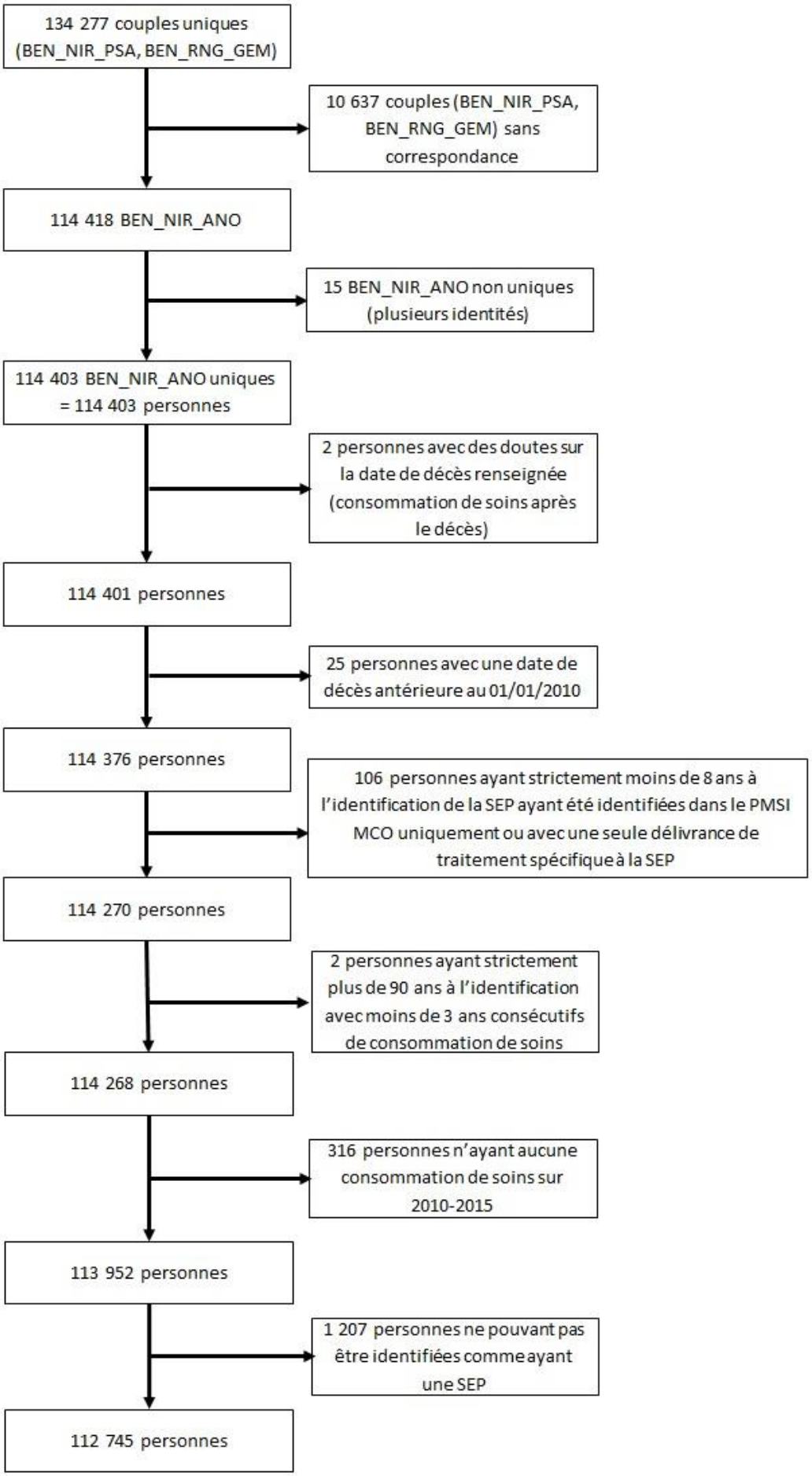


Figure 14-Flow-chart de la constitution de la population d'étude dans le SNDS

II.2.3.3. Consolidation de la population d'étude

Afin de consolider la population d'étude et ainsi éviter des faux positifs, c'est-à-dire des patients identifiés comme ayant une SEP à tort, plusieurs tests ont été effectués. Ces derniers avaient pour but de garantir la qualité des données, notamment sur l'identification de la population d'étude servant de base à toutes les analyses (Figure 14).

- Malgré une période d'identification commençant au début de l'année 2010, 25 BEN_NIR_ANO correspondant à des patients décédés avant le 1er janvier 2010 ont été identifiés et donc exclus.
- L'algorithme d'identification de l'Assurance Maladie a été appliqué une seconde fois sur les personnes restantes (n=114 376) par nos soins afin de déterminer la source et la date d'identification de chaque personne. Un quatrième critère portant sur la présence d'au moins une hospitalisation dans le PMSI SSR avec un diagnostic de SEP en morbidité principale, en affection étiologique, ou en DAS, a été pris en compte afin de retrouver le maximum de personnes. D'après les discussions avec plusieurs neurologues suivant des cas de SEP pédiatriques (démarrage de la maladie avant 18 ans), toutes les personnes identifiées entre 8 et 18 ans ont été incluses. Concernant les personnes ayant strictement moins de 8 ans à l'identification de la SEP, elles ont été incluses sauf si elles avaient été identifiées dans le PMSI MCO uniquement ou avec une seule délivrance de traitement spécifique à la SEP, ces modes d'identification pouvant indiquer un parent ayant la SEP et non l'enfant lui-même. Au total, 106 personnes ont été exclues de la population d'étude sur ce dernier critère.
- De plus, les personnes identifiés à plus de 90 ans et ayant moins de 3 ans de suivi consécutifs ont été exclus (n=2), ainsi que les personnes ne présentant aucune consommation de soins sur la période d'étude 2010-2015 (n=316).
- Sur les 113 952 personnes restantes, 1 207 n'ont pas pu être ré-identifiées comme ayant une SEP par un des quatre critères de l'algorithme employé, elles ont donc été exclues, portant la taille finale de la population d'étude sur le SNDS à 112 745 personnes.

II.3. Paramètres étudiés

II.3.1. Pré-traitement des données du SNDS

Les différentes bases constituantes du SNDS n'étant pas conçues à l'origine pour être utilisées dans le cadre d'études épidémiologiques, un traitement préliminaire a dû être effectué sur les données. Ce travail a conduit à poser plusieurs hypothèses de travail et à appliquer les recommandations formulées par la CNAMTS sur la partie data management. L'ensemble de ce travail a été effectué au travers du logiciel R (v.3.4.3) (153) sur un serveur sécurisé.

II.3.1.1. Recommandations de l'Assurance Maladie avant traitement

Les spécificités de facturation de certains établissements de santé font que la CNAMTS a formulé des consignes quant au traitement des informations du SNDS. Dans certains produits fournis par la CNAMTS, ces directives ont été appliquées et l'utilisateur n'a pas à en tenir compte, comme c'est le cas dans l'EGB. Cependant, dans le cadre d'une extraction physique des données, elles n'ont pas été prises en compte puisqu'il s'agit alors d'une extraction brute des données, charge alors à l'utilisateur de le faire.

Il existe une redondance dans la facturation des établissements de santé publics et ESPIC concernant les ACE et les séjours pour ceux qui ne sont pas passés en facturation directe à l'Assurance Maladie. Ainsi, les informations pour ces actes de soins sont présentes à la fois dans le PMSI et dans le DCIR, où elles sont transmises pour information provoquant un risque de double compte. Un premier traitement des données a donc conduit à supprimer de la table des prestations du DCIR, l'ensemble des prestations transmises pour information, soit DPN_QLF=71 (n=3 103 970 (1,6%) sur 2010-2015).

Dans le PMSI, les établissements sont repérés par leur numéro FINESS unique. Deux FINESS peuvent être distingués : le FINESS juridique correspondant au numéro de l'entité juridique, c'est-à-dire la personne morale détenant les droits et autorisations, et le FINESS géographique correspondant à un établissement rattaché à une entité juridique. Ainsi, un établissement de santé, identifié par un numéro FINESS juridique, est constitué d'un ou de plusieurs établissements dits géographiques, identifiés chacun par un numéro FINESS géographique. Par exemple, dans le cadre du CHRU (Centre Hospitalier Régional et Universitaire) de Rennes, un FINESS juridique est attribué à l'entité juridique, le CHRU, et un FINESS géographique est attribué à chacun des établissements le constituant, soit le site de l'Hôtel Dieu, le site de Pontchaillou, le site de la Tauvrais et l'Hôpital Sud. Dans le PMSI, les séjours sont codés par le numéro géographique dans le cadre des établissements de santé privés et par le numéro juridique pour les établissements publics et ESPIC (154). Par conséquent, il ne peut pas y avoir de redondance dans le codage des séjours et ACE excepté pour les groupements d'hôpitaux de Paris (APHP-Assistance Publique des Hôpitaux de Paris), de Marseille (APHM – Assistance Publique des Hôpitaux de Marseille) et de Lyon (HCL – Hospices Civils de Lyon). En effet, dans le cadre de ces trois regroupements, les séjours et ACE sont transmis en doublons dans le PMSI MCO. Afin d'éviter des redondances dans l'estimation de la consommation de soins, les séjours (n=157 705 (12,4%) sur 2010-2015) et ACE (n=474 007 (11,5%) sur 2010-2015) correspondant aux FINESS géographiques de l'APHP, de l'APHM et des HCL ont été exclus pour l'ensemble des données à disposition (2007 à 2015), comme préconisé par la CNAMTS. Par ailleurs, les séjours ayant des GHM en erreur et les prestations inter-établissement ont également été supprimés (n=31 988 (2,9%) sur 2010-2015).

II.3.1.2. Gestion des quantités délivrées négatives

Comme évoqué précédemment, certaines prestations peuvent nécessiter des régularisations à cause d'erreurs de remboursement, ce qui entraîne des prestations avec des valeurs négatives. Plusieurs variables sont alors impactées, notamment celle quantifiant le nombre de boîtes délivrés pour une spécialité médicamenteuse donnée (*PHA_ACT_QSN*) dans la table affinée des prestations pharmaceutiques (*ER_PHA_F*) ainsi que *TIP_ACT_QSN* correspondant à la quantité signée affinée d'une prestation en rapport avec un acte issu de la LPP dans la table affinée *ER_TIP_F*.

Afin de prendre en compte ces régularisations, un travail de data management a été effectué en fonction de la situation rencontrée parmi les trois possibles. Concernant la table affinée des prestations pharmaceutiques, dès qu'une régularisation (c'est-à-dire une valeur de *PHA_ACT_QSN* négative) était identifiée, un filtre était effectué sur les lignes de la table correspondant au même numéro ANO (*BEN_NIR_ANO*), à la même spécialité médicamenteuse (*PHA_PRS_C13*, *PHA_PRS_IDE*) et à la même date d'exécution (*EXE_SOI_DTD*) plus ou moins 3 jours. Cette dernière hypothèse a été choisie en se basant sur les données à disposition. Une décision a ensuite été prise en fonction du nombre de lignes obtenues avec le filtre :

(i) Si seule la ligne avec la valeur de *PHA_ACT_QSN* négative était retenue, alors cette ligne était gardée en corrigeant la valeur de *PHA_ACT*_QSN en 1 et en changeant le montant remboursé en positif. En effet, il a été considéré que la présence d'une ligne de remboursement (même négative) était liée à une délivrance d'un médicament, et donc le nombre de boîtes délivrées a été fixé à 1 boîte.

(ii) Si deux lignes étaient retenues (la ligne négative et la positive correspondante), alors la ligne avec la valeur de quantité négative était supprimée de la base et la ligne positive correspondante étant conservée (la boîte ayant été réellement délivrée d'où la facturation à l'Assurance Maladie). Il est apparu comme important le fait de considérer la délivrance alors qu'elle aurait été annulée avec la ligne ayant une quantité signée négativement.

(iii) Si trois lignes ou plus étaient retenues, alors une nouvelle ligne était créée afin de remplacer celles-ci. Cette nouvelle ligne traduisait les mêmes informations, en ayant comme quantité délivrée la somme des valeurs des *PHA_ACT_QSN* des lignes supprimées. Les lignes ajoutées ayant des valeurs de *PHA_ACT_QSN* négatives ou nulles ont enfin été supprimées.

Un raisonnement analogue a été appliqué sur la table *ER_TIP_F* concernant la variable *TIP_ACT_QSN* en ne considérant aucun jour de variation dans la date d'exécution de la prestation (*EXE_SOI_DTD*), contrairement aux trois jours de battement autorisés pour les prestations pharmaceutiques car constaté dans la base.

II.3.1.3. Gestion des séjours non uniques

La jointure entre les données du PMSI, quelle que soit la base considérée, et le référentiel des consommants permettant de faire le lien entre le parcours ambulatoire et le parcours hospitalier s'effectue sur le numéro unique d'un patient *BEN_NIR_ANO*. Une seule des deux clés de jointure concernant le patient (*BEN_NIR_ANO* et *BEN_RNG_GEM*) est donc utilisée, entraînant des problèmes d'identification unique d'un couple séjour-patient et donc possiblement un séjour correspondant à plusieurs patients. Les séjours ne permettant pas unicité de la base ont donc été supprimés, soit 433 et 58 séjours dans les PMSI MCO et SSR sur 2007-2015, respectivement, et 880 ACE dans le PMSI MCO sur 2010-2015.

Un contrôle des données du PMSI SSR a par ailleurs mis en évidence un dédoublement des séjours hospitaliers pour les années 2014 et 2015 principalement. Après analyse, il semblerait que cette erreur provienne d'un double codage des séjours selon deux versions différentes du RHA. Par conséquent, seuls les séjours où l'année de sortie correspondait à l'année du PMSI considéré ont été gardés.

II.3.2. Paramètres d'ajustement

II.3.2.1. Lieu de résidence

Une des variables sensibles du SNDS est la commune de résidence d'un bénéficiaire, constituant le niveau le plus fin de localisation disponible de ce dernier et permettant de lui assigner des variables socio-économiques contextuelles. Cette variable, appelée *BEN_RES_COM*, dans la base est couplée avec le département de résidence, *BEN_RES_DPT*. Ces deux informations peuvent être récupérées dans différentes tables, que sont le référentiel des bénéficiaires (*IR_BEN_R*), la table des prestations (*ER_PRS_R*) et la table des séjours hospitaliers MCO entre autres. Le référentiel des bénéficiaires recense les dernières valeurs connues d'un bénéficiaire. Le codage des départements est identique entre les différentes tables, mais il diffère concernant celui de la commune. Ainsi, dans *IR_BEN_R* et *ER_PRS_R*, la commune est codée selon les codes proposés par l'Insee (Institut national de la statistique et des études économiques) alors qu'elle est codée selon les codes géographiques PMSI dans le MCO. Le code Insee permet une identification unique d'une commune. A l'opposé, le code géographique PMSI correspond le plus fréquemment au code postal d'une commune ou lorsque la taille de celle-ci n'est pas assez importante (moins de 1000 habitants) à un regroupement de communes (155). Il n'y a donc pas unicité de la commune dans le cadre des codes géographiques PMSI. Néanmoins, un code PMSI peut être transformé en code Insee en passant d'abord par sa transformation en code postal.

De nombreuses valeurs manquantes et inexactitudes, notamment pour les résidents des DROM (Départements et Régions d'Outre-Mer) et de la Corse avec l'impossibilité de différencier Corse du

Sud de la Haute-Corse, ont pu être repérées dans notre extraction. Un arbre de décision a alors été défini afin d'identifier le mieux possible le dernier département et la dernière commune de résidence connus d'un bénéficiaire :

(i) Dans le cas où la commune et le département de résidence étaient renseignés de manière précise sans données manquantes dans le référentiel des consommants, alors ils étaient gardés puisque ce sont normalement les informations les plus récentes.
(ii) Dans le cas contraire, la base des prestations a été utilisée pour retrouver le dernier couple commune/département de résidence où les deux valeurs étaient renseignées.
(iii) Si aucun couple ne pouvait être retrouvé, alors une recherche dans le PMSI MCO a été effectuée pour retrouver le dernier couple commune/département sans valeurs manquantes.
(iv) Dans le cas où aucune valeur ne pouvait être imputée, des valeurs manquantes étaient ajoutées pour le bénéficiaire pour les deux paramètres.

Dans le cas d'une sélection via les données du PMSI MCO, une hypothèse supplémentaire a été effectuée du fait du codage des communes en codes géographiques PMSI. Dans le cas des petites communes, le code PMSI correspondant à plusieurs codes postaux, le premier code postal de la liste a alors été choisi en faisant l'hypothèse de la faible variation des caractéristiques socio-économiques entre plusieurs villages situés à proximité les uns des autres. Au total, 37 patients (0,03%) sur l'ensemble de la population (n=112 745) ont eu cette approximation. Pour les résidents de la Corse et des DROM, les codes Insee des communes de résidence ont été récupérés en effectuant une jointure sur les noms des communes avec la base de correspondance disponible via la plate-forme ouverte des données publiques françaises (156).

Une étude de la prévalence de la SEP au 31 décembre 2015 a été menée grâce à l'identification des communes de résidence. Une carte de la prévalence par département après standardisation selon l'âge, le sexe et le département de résidence sur la population française en 2015 en utilisant les données de recensement de l'Insee (157) a également réalisée grâce au logiciel Philcarto (158).

II.3.2.2. Variables socio-économiques contextuelles

L'identification de la commune de résidence d'un bénéficiaire permet de lui attribuer des paramètres socio-économiques et d'accès aux soins contextuels liés à son environnement de vie. Deux paramètres principaux ont été considérés : l'accessibilité potentielle localisée (APL) (159–161) et l'indice de défavorisation sociale (FDep) (162). Dans les cas où aucune commune de résidence n'a pu être identifiée ou que les indicateurs n'étaient pas estimés pour une commune donnée, une valeur manquante était attribuée aux deux paramètres pour le bénéficiaire.

L'APL est un indicateur construit par la DREES (Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques) et l'Irdes (Institut de recherche et documentation en économie de la santé) à partir des données du SNIIRAM et de l'Insee et permet de quantifier l'adéquation entre l'offre et la demande de soins d'une commune en prenant en compte le temps d'accès au professionnel de santé. Ce paramètre a été défini pour la première fois pour les médecins généralistes libéraux puis a été étendu pour d'autres professionnels de santé comme les masseurs-kinésithérapeutes, les infirmiers, ou encore les pharmacies. Au moment de ce travail, les APL des années 2010, 2013 et 2015 construites sur la base du distancier de l'Insee Metric (161) étaient disponibles. Pour les années 2010 et 2013, une APL pour un professionnel de santé donné de 8.0 correspond à une accessibilité de la population de cette commune de 8.0 équivalents temps plein du professionnel considéré pour 100 000 habitants. En 2015, l'interprétation de l'APL a été modifiée. Une APL de 8.0 correspond alors à l'interprétation suivante : « les habitants de la commune ont accès en moyenne à 8.0 consultations ou visites avec le professionnel considéré par an et par habitant ». Les APL pour 2015 apparaissent donc comme étant plus faibles que celles des années précédentes. Seules les APL de l'année 2010 ont été utilisées par la suite car correspondant au début de la période d'étude.

Le FDep a quant à lui été défini par l'Inserm à partir des données de recensement de l'Insee. Défini au niveau de l'Iris (îlots regroupés pour l'information statistique), la plus petite unité géographique française, et ensuite agrégé à l'échelle communale, il permet de quantifier le niveau socio-économique de la zone. Il correspond à la première composante d'une analyse en composantes principales (ACP) basée sur quatre variables issues du recensement : deux variables comptant positivement dans la défavorisation sociale (le pourcentage d'ouvriers dans la population active et le taux de chômage) et deux ayant une dimension négative (le revenu médian par unité de consommation dans le ménage et le pourcentage de bacheliers dans la population de plus de 15 ans) (162). Il a été calculé une première fois avec les données de 1999, puis mis à jour sur les données de 2009 et 2013. Dans cette thèse, seuls les FDep de l'année 2009 ont été considérés. Les valeurs continues correspondant au premier axe de l'ACP ont été utilisés sous la forme de quintiles au niveau populationnel comme défini dans la littérature, le premier quintile correspondant aux communes les plus favorisées et le cinquième aux plus défavorisées (162).

II.3.2.3. Variables socio-économiques individuelles

Le FDep étant une variable contextuelle liée à la commune de résidence d'un bénéficiaire, elle peut ne pas refléter le réel statut socio-économique de ce dernier. Afin de mieux approcher la réalité, des variables individuelles ont été estimées à partir des données d'affiliation des bénéficiaires. Ainsi les statuts des bénéficiaires vis-à-vis de la CMU (Couverture Maladie Universelle) et de l'AME (Aide Médicale de l'Etat) ont été identifiés.

La CMU, devenue depuis 2016 la Protection Universelle Maladie (PUMA)[10], assure aux personnes éligibles un accès gratuit à l'Assurance Maladie sous conditions de ressources (163). Cet accès leur permet de bénéficier du remboursement de la part obligatoire d'un acte de soins. Les personnes bénéficiant de la CMU sont alors affiliées à des petits régimes spécifiques rattachés au régime général. Afin de repérer ces personnes, un algorithme développé par la CNAMTS basé sur les régimes de rattachement lors des différentes prestations a été implémenté. Pour chaque patient, les prestations avec une date d'exécution renseignée (*EXE_SOI_DTD*) pour lesquelles le grand régime de rattachement (*RGM_GRG_COD*) correspondait au régime général et le petit régime de rattachement (*RGM_COD*) faisait partie de la liste présenté dans la Table 4 ont été sélectionnées. Deux périodes ont été considérées : la période d'inclusion (2010-2015) et l'ensemble de la période d'étude (2007-2015). Pour chaque période de temps, si un patient présentait au moins une prestation avec un rattachement à la CMU alors il était identifié comme bénéficiant de ce dispositif sur la période considérée.

L'AME permet aux étrangers en situation irrégulière résidant sur le territoire français sous conditions de ressources de bénéficier d'une prise en charge jusqu'à 100% des dépenses de santé et de bénéficier de l'avance de frais (163). De manière analogue à la CMU, un algorithme à trois critères développé par la CNAMTS sur les prestations a été utilisé pour identifier les patients bénéficiant de l'AME. Pour chaque patient, les prestations avec une date d'exécution renseignée (*EXE_SOI_DTD*) sélectionnées étaient celles pour lesquelles :

- le grand régime de rattachement (*RGM_GRG_COD*) correspondait au régime général et le petit régime de rattachement (*RGM_COD*) faisait partie de la liste présenté dans la Table 4,
- ou la catégorie d'organisme complémentaire CMU (*BEN_CMU_CAT*) correspondait à l'AME,
- ou le petit régime de rattachement (*RGM_COD*) correspondait au régime anonymisation et que l'organisme complémentaire CMU était égal à 75500017.

Les périodes considérées étaient les mêmes que pour la CMU. De même, pour chaque période de temps, si un patient présentait au moins une prestation avec un rattachement à l'AME alors il était identifié comme bénéficiant de ce dispositif sur la période considérée.

[10] Article 59 de la loi n°2015-1702 de financement de la sécurité sociale pour 2016 publiée au Journal officiel du 22 décembre 2015.

Table 4-Liste des petits régimes sélectionnés selon le dispositif d'accès aux soins considéré

Dispositif considéré	Petit régime de rattachement *RGM_COD*	Libellé
CMU	801	Critère de résidence avec cotisation
CMU	802	Critère de résidence sans cotisation
CMU	803	Critère de résidence sans cotisation, affiliation provisoire 3 mois renouvelables
CMU	804	Critère de résidence avec cotisation + risque accident du travail (personne exerçant une activité insuffisante pour ouvrir des droits au régime général)
CMU	806	Critère de résidence sans cotisation assuré RSA Socle et RSA socle majoré
CMU	808	Ressortissants communautaires inactifs non cotisants
CMU	809	Ressortissants communautaires inactifs cotisants
CMU	833	Critère de résidence ex-yougoslave majeur
AME	95	Aide médicale gratuite 100% Etat
AME	96	Aide médicale hospitalière gratuite 100% Etat

L'identification du statut CMU-C (Couverture Maladie Universelle Complémentaire) est possible dans le SNDS au moyen d'une variable calculée par l'Assurance Maladie (*BEN_CMU_TOP*). La CMU-C est une complémentaire santé gratuite qui prend en charge la part non couverte par les régimes d'Assurance Maladie obligatoire et permet une meilleure approche du niveau socio-économique d'un bénéficiaire que la CMU, le plafond de conditions de ressources étant plus restrictif que pour cette dernière. Cependant, le codage des remboursements en cas d'ALD concomitante privilégie l'ALD et non la CMU-C, entraînant alors une sous-estimation du nombre de personnes bénéficiant de cette complémentaire, sous-estimation qui peut s'avérer très importante dans le cas d'une maladie chronique telle que la SEP. De plus, la considération de la CMU-C nécessite d'exclure les bénéficiaires de plus de 60 ans qui bénéficient alors de l'ASPA (Allocation de Solidarité aux Personnes Agées). Pour ces deux raisons, la CMU-C n'a pas été considérée dans ce travail.

II.3.2.4. Régime de rattachement

En complément du statut vis-à-vis de la CMU, le régime de rattachement aux dernières nouvelles a été identifié, car étant le seul à disposition dans le référentiel des bénéficiaires *IR_BEN_R*. Pour ce faire, les trois premiers caractères de l'organisme d'affiliation du bénéficiaire ont été isolés permettant

d'obtenir son grand régime de rattachement. Ce paramètre a été catégorisé selon les grands régimes d'Assurance Maladie français, à savoir le régime général, les sections locales mutualistes, la MSA, le RSI et les autres régimes regroupés ensemble. Par ailleurs, les bénéficiaires de la CMU étant rattachés au régime général, la modalité concernant ce dernier a été scindée en deux catégories : régime général hors CMU et bénéficiaires de la CMU.

II.3.2.5. Affection de Longue Durée

Une des rares variables cliniques concernant le patient à disposition dans le SNDS est le statut d'affection de longue durée contenu dans la table *ER_IMB_R*. Cette table, contrairement au référentiel des consommants où les nouvelles données écrasent les précédentes, fonctionne sous le format d'un historique des demandes ALD d'un patient, c'est-à-dire qu'une même ALD pour un patient donné est codée sur plusieurs lignes, une ligne par demande de renouvellement.

Plusieurs filtres ont été appliqués sur les données afin d'obtenir le statut présence/absence de chacun des patients vis-à-vis d'une des 30 affections de la liste des ALD. Premièrement, les lignes avec un numéro d'ALD (*IMB_ALD_NUM*) égal à 98 (n=6 259 (1,9%)), correspondant au transcodage, ont été supprimées conformément au communiqué du 16 décembre 2016 de la CNAMTS. Ensuite, seules les lignes avec un motif d'exonération (*IMB_ETM_NAT*) égal à 41 ont été conservées (n=247 561 (77,8%)), permettant de ne retenir que les lignes en rapport avec les 30 ALD de la liste. Un contrôle de qualité des données a révélé des erreurs dans le codage des dates de début des ALD (*IMB_ALD_DTD*) pour certaines ALD. En effet, des ALD débutaient après 2017 (n=44) majoritairement à cause de fautes de frappe. Ces erreurs ont été corrigées dans la base en faisant le rapprochement avec les autres données sur les ALD considérées à disposition.

Afin d'isoler pour un patient donné ses informations sur l'ALD SEP, un filtre supplémentaire a été appliqué sur la variable *MED_MTF_COD* codant pour le motif médical de l'exonération selon le code CIM-10 de la SEP, *G35* pour rappel. Une indicatrice qualifiant la présence d'une ALD SEP sur la période d'étude 2010-2015 a été construite à partir des variables dates de début et de fin de l'ALD (*IMB_ALD_DTD* et *IMB_ALD_DTF* respectivement). La date de début de l'ALD a également été conservée pour calculer la durée depuis la mise en ALD. Sur un raisonnement similaire, une indicatrice codant pour la présence d'au moins une autre ALD différente d'une ALD SEP a été construite dans l'EGB uniquement.

II.3.2.6. Indice de comorbidités de Charlson

L'identification d'ALD autres qu'une ALD SEP peut permettre d'avoir une idée du nombre de comorbidités de ce dernier. Cependant, cela entraîne l'hypothèse que toutes les affections ont un poids similaire sur la santé du patient. Afin de prendre en compte ces différences, plusieurs indices ont été

définis dans la littérature dont l'indice de comorbidités de Charlson (164), un des plus populaires. Défini en 1987 (164), il s'agit d'un indice basé sur 19 comorbidités et servant à la prédiction de la mortalité à 1 an grâce à un poids attribué à chacune d'entre elles (Table 5). Cet indice à l'origine basé sur des données cliniques a ensuite été adapté aux bases médico-administratives dans la CIM-9 (Classification Internationale des Maladies – 9ème version) (165) puis la CIM-10 (166,167). La version que nous avons utilisée dans ce travail est celle définie par Bannay et al. basée sur les données du SNIIRAM (168). Afin de construire cet index dans notre population du SNDS uniquement, 17 comorbidités ont été identifiées dans les 12 mois précédant la date d'inclusion des patients (Table 5). En cas d'un manque de recul du à la disponibilité des données, une valeur manquante était imputée. Les pathologies ont été identifiées à partir de quatre sources d'information :

- les diagnostics hospitaliers principaux, reliés et associés des RSA et des RUM dans le PMSI MCO,
- les diagnostics des motifs d'exonération des ALD (*MED_MTF_COD*),
- les prestations pharmaceutiques en ville,
- les actes CCAM effectués à l'hôpital dans le PMSI MCO durant un séjour hospitalier.

Chaque pathologie identifiée conduisait à un poids, qui une fois tous sommés donnaient l'indice de Charlson. On peut noter l'importance des cancers, notamment métastatiques, fortement liés avec la mortalité, alors que plusieurs pathologies ont un poids de 0 car n'ayant plus d'influence sur cette dernière de nos jours. L'indice de Charlson ajusté sur l'âge a également été calculé en augmentant le score de Charlson de 1 unité pour chaque dizaine d'âge à partir de 50 ans (+1 pour un patient entre 50 et 59 ans, +2 entre 60 et 69 ans,…) (168).

Table 5-Pathologies considérées dans le calcul de l'indice de comorbidités de Charlson et poids associés selon l'algorithme considérée

Pathologie	Poids associé (Charlson (164))	Poids associé (Quan (167))	Poids associé (Bannay (168))
Infarctus du myocarde	1	0	0
Insuffisance cardiaque congestive	1	2	2
Pathologie vasculaire périphérique	1	0	1
Accident vasculaire cérébral	1	0	1
Démence	1	2	2
Pathologie pulmonaire chronique	1	1	1
Connectivite (atteinte diffuse, inflammatoire et chronique du tissu conjonctif)	1	1	0
Pathologie ulcéreuse peptique	1	0	0
Pathologie hépatique modérée	1	2	2
Diabète non compliqué	1	0	0
Diabète compliqué avec atteinte des organes	2	1	0
Insuffisance rénale modérée ou sévère	2	1	1
Hémiplégie	2	2	2
Tumeur (incluant leucémie et lymphome)	2	2	2
Pathologie hépatique modérée ou sévère	3	4	3
Tumeur solide métastatique	6	6	11
VIH avec ou sans SIDA	6	4	1

III. Descriptif de la population SEP en France

III.1. Population SEP dans l'EGB

Les trois populations identifiées dans l'EGB sont présentées ci-dessous. Les modes d'identification de la population sont étudiés seulement pour la population globale des 1 000 patients.

III.1.1. Mode d'identification dans l'EGB

Pour rappel, l'algorithme d'identification des cas de SEP dans l'EGB était basé sur trois critères : la présence d'une ALD SEP, ou la délivrance d'un traitement spécifique à la SEP, ou un diagnostic principal, relié ou associé SEP dans le PMSI MCO sur la période 2007-2012. Parmi les 1 000 patients identifiés dans l'EGB, 82,9% ont été identifiés grâce au critère ALD, 72,6% grâce au PMSI MCO et 54,9% grâce aux traitements spécifiques (Figure 15). Au total, 381 personnes (38,1%) ont été identifiées via les trois critères, 342 (34,2%) avec deux critères et 277 (27,7%) avec seulement un critère, dont la majorité via le PMSI (n=113 (11,3%)) ou le statut ALD (n=152 (15,2%)).

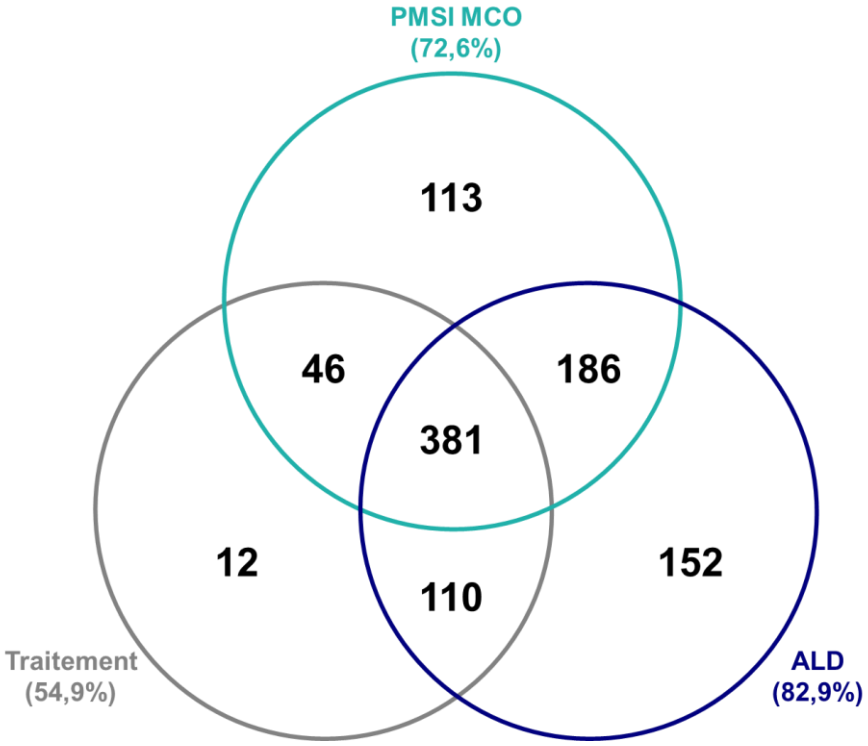


Figure 15-Diagramme de Venn du mode d'identification des personnes ayant la SEP dans l'EGB sur 2007-2012 (N=1 000)

III.1.2. Descriptif des populations de l'EGB

Les caractéristiques principales des trois populations construites dans l'EGB (population « *Recours aux soins* » (N=1 000), population « *Etude de faisabilité PGR* » (n=70) et population « *Algorithme handicap* » (n=650)) sont présentées dans la Table 6 et la Figure 16. Pour rappel, les individus des populations « *Etude de faisabilité PGR* » et « *Algorithme handicap* » sont inclus dans la population totale « *Recours aux soins* ».

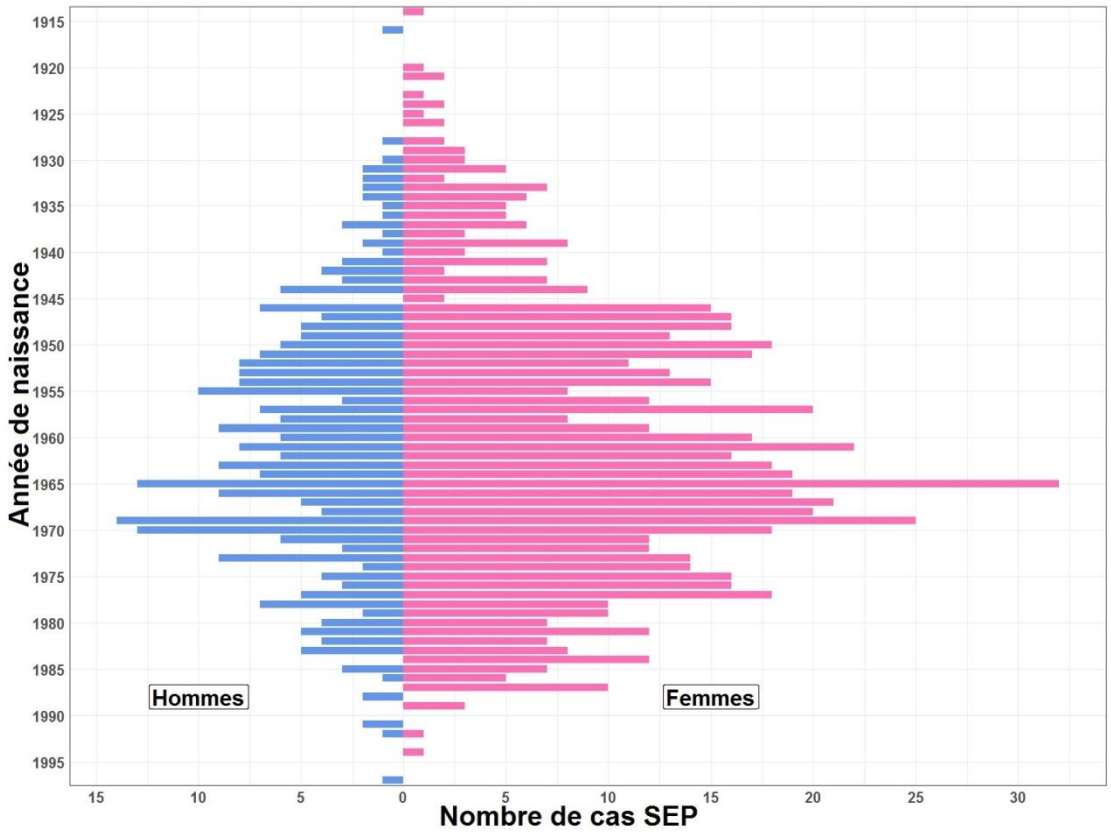


Figure 16-Pyramide des âges de la population globale identifiée dans l'EGB (N=1 000)

Le but principal de la thèse étant l'étude de la population totale en France ayant la SEP à partir du SNDS, peu de temps a été consacré à la construction de variables spécifiques à la population de l'EGB (pour rappel, les structures et variables de jointure étant différentes entre l'EGB et le SNDS, il n'a pas été possible d'avoir recours à une réplication directe des programmes). Par ailleurs, des paramètres supplémentaires spécifiques à chaque population seront décrits dans les sections suivantes.

Parmi les 1 000 patients de la population servant à l'étude du recours aux soins, le sex-ratio était de 2,5 femmes pour 1 homme avec un âge médian de 44 ans en 2007. Au total, 71 décès ont été identifiés, soit un taux de 12,9 décès pour 1 000 personnes-années et le statut d'ALD SEP a été identifié comme « valide au cours de la période d'étude » pour 838 personnes (83,8%). Le sex-ratio F:H était de 2,3 et 2,0 dans la population utilisée pour le développement de l'algorithme estimant le

handicap moteur et celle pour l'étude de faisabilité respectivement. Dans ces deux populations 94,3% des patients étaient sous ALD SEP au cours du suivi. Dans les trois populations grâce au mode de fonctionnement des bases médico-administratives, la majorité des patients était suivi durant la quasi-totalité de la période d'étude.

Table 6-Principales caractéristiques des trois populations identifiées dans l'EGB

	Population ***Recours aux soins*** (N=1 000)	**Population** ***Etude de faisabilité PGR*** (N=70)	**Population** ***Algorithme handicap*** (N=650)
Nombre de femmes (%)	711 (71,1%)	47 (67,1%)	451 (72,5%)
Année de naissance [a]	1963 (1914-1997)	1974 (1934-1987)	1960 (1916-1991)
Nombre de patients ALD SEP [b,c] (%)	838 (83,8%)	66 (94,3%)	613 (94,3%)
Durée depuis mise en ALD SEP [a,b,d] (années)	11,6 (0,0-34,6)	8,2 (1,8-23,9)	12,3 (0,0-34,0)
Au moins une autre ALD [b,e] (%)	278 (27,8%)	8 (11,4%)	121 (18,6%)
Durée de suivi [a] (années)	6,8 (0,0-7,0)	8,2 (1,5-8,2)	6,4 (0,0-6,4)
Nombre de décès (%)	71 (7,1%)	1 (1,4%)	48 (7,4%)

[a] Médiane (minimum-maximum). [b] ALD : Affection de Longue Durée. [c] ALD valide au cours du suivi. [d] Calculé à la date aux dernières nouvelles (fin de la période de suivi considérée ou date du décès du patient). [e] ALD autre que la SEP.

III.2. Population SEP dans le SNDS

III.2.1. Mode d'identification dans le SNDS

Les 112 745 cas de SEP ont été identifiés dans le SNDS au moyen de quatre critères comme évoqué précédemment. Le diagramme présentant les différents modes d'identification de cette population est présenté dans la Figure 17. Le critère ayant permis d'identifier le plus de patients est le critère de l'ALD (85,5%), suivi de celui sur les diagnostics du PMSI MCO (71,4%). A l'opposé, le critère portant sur le PMSI SSR n'a permis de repérer que 20,5% des patients. Concernant l'identification par plusieurs sources, seules 8 127 (7,2%) des personnes identifiées l'ont été avec les quatre critères, 41 893 (37,2%) par trois critères et 32 467 (28,9%) par deux critères. Au total, 30 078 (26,7%) personnes ont été repérées via une seule source, dont plus de la moitié (17 132 – 57,0%) grâce au statut ALD uniquement. Parmi les patients identifiés via les séjours hospitaliers MCO et SSR, 7 586 (65,1%) l'étaient avec un seul séjour hospitalier présentant un diagnostic SEP, dont 4 433 (58,4%)

avaient un DP ou un DR SEP, 3 160 avaient un DAS SEP et 2 avaient pour affection étiologique la SEP en SSR.

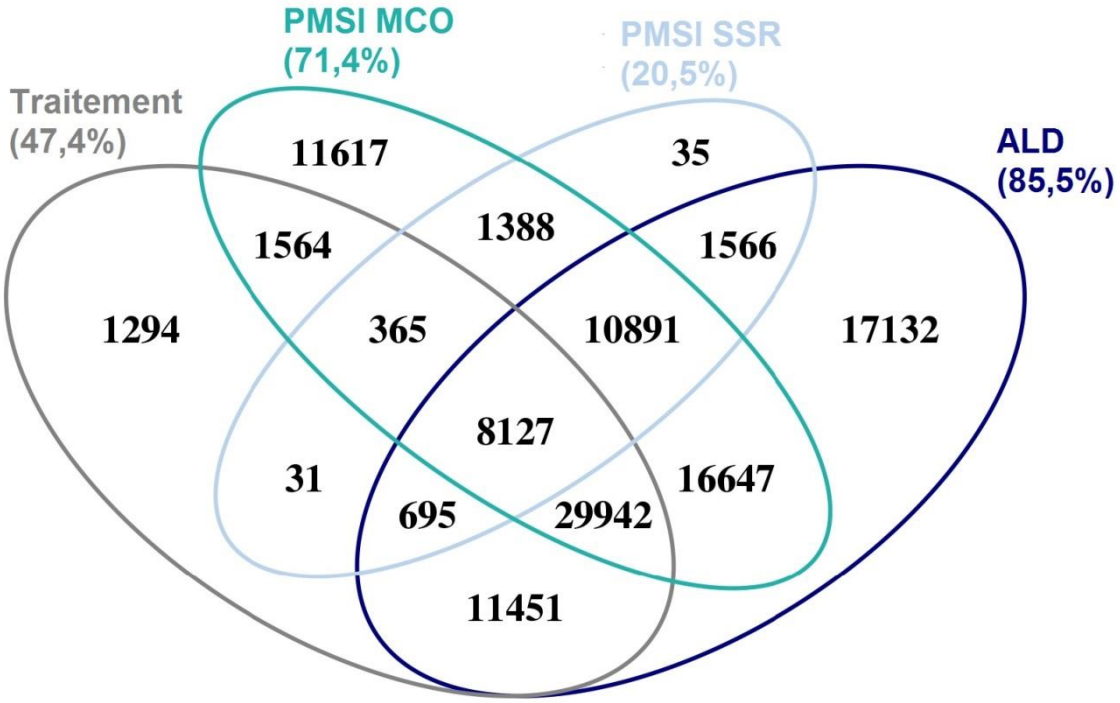


Figure 17-Diagramme de Venn du mode d'identification des personnes ayant la SEP dans le SNDS sur 2010-2015 (N=112 745)

III.2.2. Prévalence de la SEP en France

La quasi-exhaustivité du SNDS vis-à-vis de la population française a permis de réaliser une estimation de la prévalence de la SEP en France à la fin de la période d'étude, soit au 31 décembre 2015. Ce travail permet ainsi de mettre à jour les précédentes données sur le sujet datant de fin 2012.

Parmi les 112 745 individus de la population d'étude, 107 716 personnes étaient vivantes en France (France métropolitaine et DROM), soit une prévalence brute de 162,2 cas SEP pour 100 000 habitants ($IC_{95\%}$=[161,2-163,1]). En séparant les deux sexes, la prévalence brute pour les hommes était de 95,9/100 000 ($IC_{95\%}$=[94,8-97,0]) et de 224,4/100 000 ($IC_{95\%}$=[222,8-226,0]) pour les femmes. Si on ne considère que les personnes dont le lieu de résidence a été clairement identifié, le nombre de personnes atteintes s'élève alors à 107 472 soit une prévalence brute de 161,8 cas pour 100 000 ($IC_{95\%}$=[160,8-162,8]). Elle s'élève à 95,7 ($IC_{95\%}$=[94,6-96,7]) et 223,9 ($IC_{95\%}$=[222,3-225,4]) respectivement chez les hommes et les femmes.

Les résultats de la standardisation selon l'âge, le sexe et le département de résidence sur la population française en 2015 sont présentés dans la Figure 18. Le gradient croissant Sud-Ouest/Nord-Est de la prévalence de la SEP a été retrouvé, le département le plus touché étant le territoire de Belfort avec une prévalence standardisée de 259,5 ($IC_{95\%}$=[234,3-287,5]) et le moins touché en France métropolitaine étant les Deux-Sèvres avec un taux égal à 123,7 ($IC_{95\%}$=[112,8-135,6]). Les DROM avaient tous une prévalence faible de la maladie inférieure à 70 cas pour 100 000 habitants.

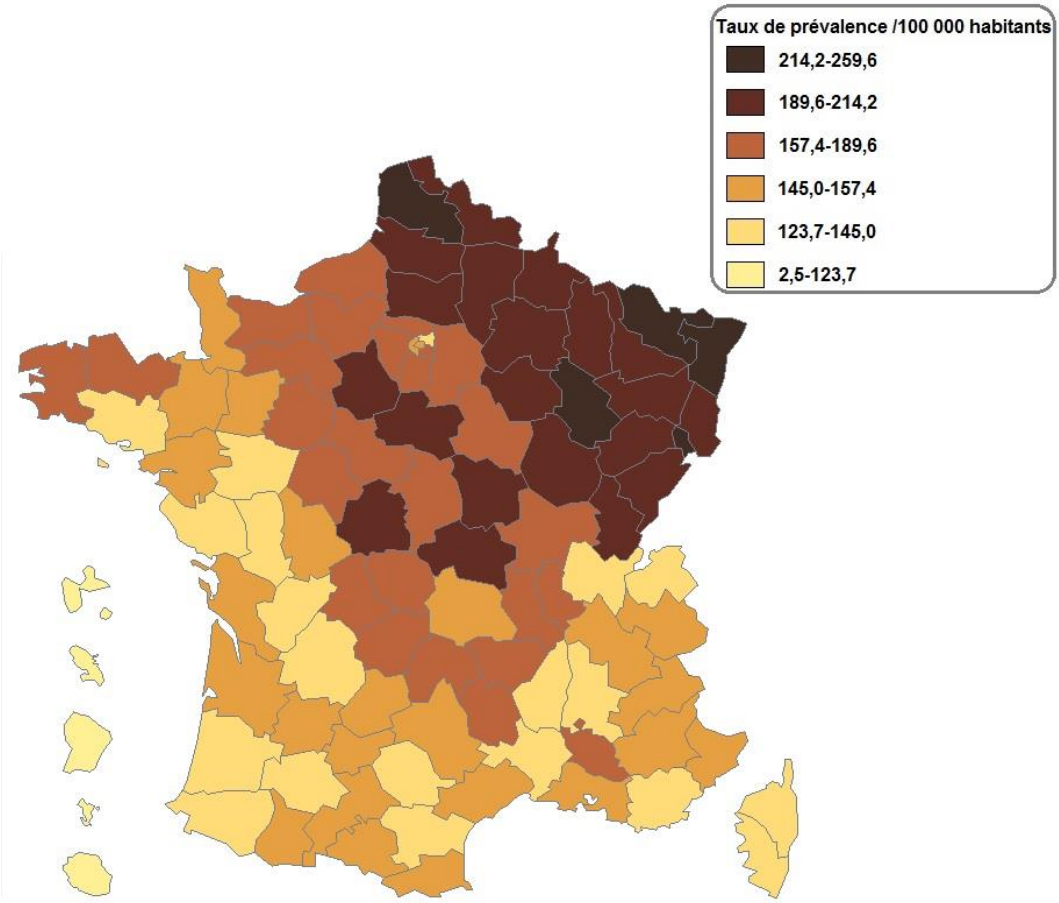


Figure 18-Carte des taux de prévalence standardisés de la SEP en France au 31 décembre 2015 (N=107 472)

III.2.3. Descriptif de la population identifiée dans le SNDS

Les caractéristiques démographiques, socio-économiques et cliniques estimées de la population identifiée dans le SNDS sont présentées dans la Table 7. Le sex-ratio de la population était de 2,4 femmes pour 1 homme avec un âge médian de 46 ans en 2010. La pyramide des âges est plus précise du fait du nombre de cas mais a une forme similaire à celle obtenue avec l'EGB (Figure 19).

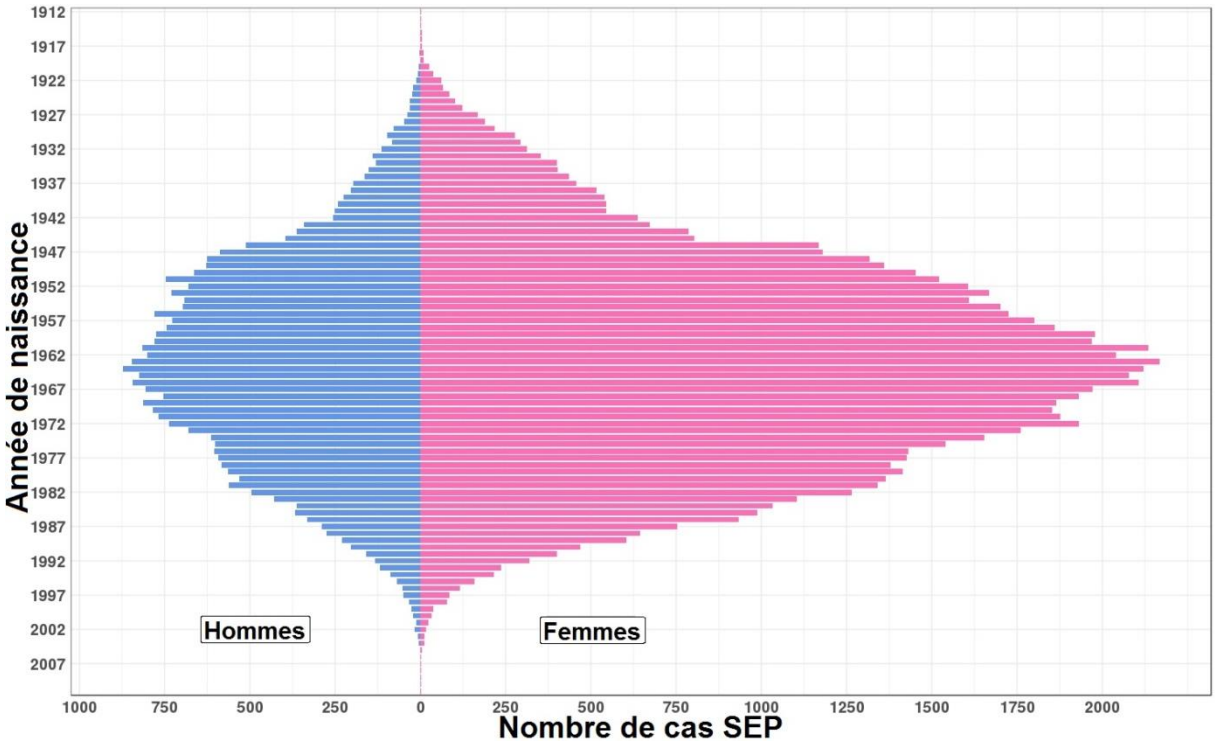


Figure 19-Pyramide des âges de la population SEP identifiée dans le SNDS (N=112 745)

Plus de la moitié des patients ont été suivis sur l'ensemble des six années de la période d'étude. Au total, 5 031 décès ont été constatés sur la période d'étude à un âge médian de 67 ans (min=7 - max=101). L'ancienneté de la maladie dans la population peut être approximée par la durée entre la mise en ALD et l'identification de la SEP dans les bases. Cette information est disponible pour 92 684 (82,2%) patients et est égale à 6,0 années.

Concernant les données cliniques estimées, 54 014 (47,9%) personnes avaient un score de comorbidités de Charlson ajusté sur l'âge supérieur ou égal à 1. Sur la période d'étude 2010-2015, 4 262 personnes (3,8%) ont bénéficié de la CMU et 154 (0,1%) de l'AME, dont 56 ayant eu recours aux deux dispositifs. Sur l'ensemble des données à disposition (2007-2015), ils étaient plus nombreux, avec respectivement 4 754 (4,2%) et 287 (0,3%) pour la CMU et l'AME.

Table 7-Descriptif de la population SEP identifiée dans le SNDS (N=112 745)

Caractéristique	Nombre de patients (%)
Nombre de femmes (%)	79 947 (70,9%)
Année de naissance [a]	1964 (1912-2011)
Age à l'identification de la SEP [a] (années)	40 (1-101)
Présence d'une ALD SEP au cours du suivi [b] (%)	96 451 (85,5%)
Durée entre la mise en ALD et l'identification de la SEP [a, b, c] (années)	6,0 (0,0-53,0)
Durée de suivi [a] (années)	6,0 (0,0-6,0)
Nom de décès sur la période d'étude (%)	5 031 (4,5%)
Régime de rattachement aux dernières nouvelles (%)	
Régime général hors CMU [d, e]	90 541 (80,3%)
Sections locales mutualistes	8 126 (7,2%)
Régime agricole	4 152 (3,7%)
Régime Social des Indépendants	4 253 (3,8%)
Bénéficiaire de la CMU [d, e]	4 262 (3,8%)
Autres régimes	1 411 (1,3%)
Score de Charlson ajusté sur l'âge [f] (%)	
0	49 735 (44,1%)
1-2	37 980 (33,7%)
3-4	12 501 (11,1%)
≥5	3 533 (3,1%)
Manquant	8 996 (8,0%)
FDep de la commune de résidence en 2009 [g] (%)	
1er quintile (communes les plus favorisées)	20 939 (18,6%)
2ème quintile	20 522 (18,2%)
3ème quintile	21 353 (18,9%)
4ème quintile	21 210 (18,8%)
5ème quintile (communes les plus défavorisées)	20 929 (18,6%)
Manquant	7 792 (6,9%)
APL de la commune de résidence en 2010 [a, h] (%)	
Médecin généraliste	70,9 (0,0-617,2)
Kinésithérapeute	67,4 (0,0-631,6)
Infirmier	67,2 (0,0-701,2)
Pharmacie	36,4 (0,0-824,5)
Manquant (%)	7 007 (6,2%)

[a] Médiane (minimum-maximum). [b] Affection de Longue Durée. [c] Calculée pour les patients ayant été identifié en premier grâce à leur statut ALD (N=92 684). [d] Couverture Maladie Universelle. [e] Bénéficiaire de la CMU sur la période d'étude 2010-2015. [f] Score de Charlson selon la définition de Bannay et al. [g] Indice de défavorisation sociale. [h] Accessibilité Potentielle Localisée.

IV. Traitements de fond dans la SEP

IV.1. Traitements de fond dans le SNDS

IV.1.1. Méthodes

IV.1.1.1. Identification des traitements

Les principaux traitements de fond de la SEP peuvent être identifiés dans le SNDS. Ainsi, l'ensemble des traitements délivrés en ville sont repérables dans la table *ER_PHA_R*. Cependant, concernant ceux délivrés à l'hôpital, seuls le natalizumab et le rituximab peuvent être identifiés dans les tables *ER_UCD*_R (pour les établissements de santé privés à but lucratif) et *FICHCOMP_MED* du PMSI MCO (pour les établissements de santé publics et les ESPIC) car étant des médicaments facturés en sus du GHS. Cependant, le rituximab est une molécule nouvellement utilisée dans la SEP et n'est pas spécifiquement utilisée pour cette indication, mais également pour le traitement du cancer et de l'arthrite rhumatoïde. Par conséquent, le rituximab pour les raisons énoncées précédemment, et la mitoxantrone, pour son impossibilité d'identification dans les bases, n'ont pas été étudiés dans ce travail. Les traitements non spécifiques à la SEP délivrés en officine (méthotrexate, mycophénolate mofétil et azathioprine) ont été considérés et rassemblés dans une catégorie « hors AMM » puisqu'ils constituent les traitements historiques de la SEP délivrés avant l'utilisation des premiers interférons-β.

Traitements délivrés en ville

Les traitements de fond délivrés dans les officines ont été recherchés dans la table des prestations pharmaceutiques, comme indiqué ci-dessus. Ils ont été identifiés au moyen de leurs codes CIP 7 et 13, (le code CIP 7 étant contenu dans le deuxième) dans les variables *PHA_PRS_IDE* et *PHA_PRS_C13*, respectivement (Annexe C). Suite aux modifications de la base apportées en pré-traitement (cf. section II.3.1.2. Gestion des), aucune délivrance négative de traitement n'était présente dans la base et seuls les remboursements liés à l'Assurance Maladie ont été considérés (*BSE_REM_SGN*=1). Pour chaque spécialité, les nombres de délivrances (nombre de remboursements distincts) et de boîtes délivrées (*PHA_ACT_QSN*) ont été calculés sur une base mensuelle. Par conséquent, si un patient avait le même mois la délivrance de plusieurs traitements de fond de la SEP, la règle arbitraire de ne considérer que le premier traitement a été prise.

Traitements délivrés à l'hôpital

Seul le natalizumab (Tysabri®) a pu être repéré dans les différentes tables. Les délivrances (et par conséquent injections) en établissements de santé publics et ESPIC ont été identifiées dans la table des médicaments en sus du GHS du MCO (*FICHCOMP_MED*). En complément, les délivrances pour les injections en établissements de santé privés à but lucratif ont été repérées dans la table du DCIR des médicaments en rétrocession (*ER_UCD_R*). Le PMSI SSR n'a pas été considéré, puisque la perfusion a lieu exclusivement en MCO.

Pour la partie concernant les hôpitaux publics et ESPIC, la table *FICHCOMP_MED* a été jointe à la table des séjours du PMSI MCO pour obtenir la date d'exécution de la prestation grâce à un couple de clé unique formé du numéro FINESS de l'établissement (*ETA_NUM*) et du RSA du séjour (*RSA_NUM*) pour une année donnée. Seuls les séjours pour lesquels le mois d'administration du médicament en sus correspondait au mois d'entrée du séjour, et pour lesquels le délai d'administration du médicament depuis l'entrée du séjour était cohérent avec la date de sortie ont été gardés (n=344 806 supprimés).

Les injections de natalizumab ont ensuite été repérées à partir du code UCD (Unité Commune de Dispensation) de la spécialité (9293940 en format court (*UCD_COD*) et 3400892933405 en format long (*UCD_UCD_COD*)), sans sélection d'un DP ou DR particulier. Une procédure similaire a été effectuée pour repérer les délivrances de natalizumab en rétrocession pour les injections dans les établissements de santé privés à but lucratif.

Nouveaux utilisateurs

Afin de compléter les informations sur l'utilisation des traitements, le nombre de nouveaux utilisateurs de chacun des traitements a été calculé sur une base annuelle. En effet, ces données permettent de mieux comprendre la dynamique de prescription des traitements et ainsi observer de potentielles variations de prise en charge au cours du temps. Dans ce travail, un initiateur a été défini comme un patient ayant une délivrance d'un traitement, sans délivrance de ce dernier dans les six mois précédents.

IV.1.1.2. Représentations graphiques

Afin d'étudier l'utilisation des traitements de la SEP, plusieurs représentations graphiques ont été utilisées, chacune permettant une meilleure compréhension de la dynamique de prise en charge.

Dans un premier temps, un chronogramme, graphique issu de l'analyse de séquences, a été construit (169). Il permet de représenter pour chaque unité de temps en abscisses, le pourcentage cumulé du nombre de patients ayant au moins une délivrance d'une spécialité thérapeutique en ordonnées. De

plus, il a l'avantage de montrer en un graphique unique les variations au cours du temps de prise en charge, les tendances de prescription, ainsi que l'impact de l'arrivée d'une nouvelle molécule sur le marché. Cependant, ce type de représentation ne permet pas d'étudier l'aspect longitudinal des séquences thérapeutiques des patients. Par conséquent, des diagrammes de type *sunburst* ont été construits. Ces derniers permettent la représentation de l'ensemble des séquences thérapeutiques de la population d'étude grâce à des cercles concentriques symbolisant chacun un nouveau traitement dans la séquence. Pour cette représentation uniquement, seules les périodes sans traitement de plus de six mois consécutifs ont été considérées dans la représentation afin de ne pas la surcharger artificiellement par des périodes vierges lors du switch entre deux traitements. Cependant, le *sunburst* ne permet pas une représentation de la dimension temporelle. C'est pourquoi dans un dernier temps, des diagrammes de Sankey ont été considérés. En plus de la représentation à des temps fixés de la proportion de patients sous une thérapeutique particulière comme le chronogramme, les flux de patients entre les différents états, donc les séquences longitudinales, sont alors figurés.

L'ensemble de ces diagrammes a été réalisé en considérant chaque traitement indépendamment les uns des autres (seuls les interférons étaient regroupés ensemble), puis les traitements regroupés selon leur mode d'administration (oral, injectable, perfusion ou « hors AMM »). Des valeurs manquantes ont été indiquées lorsque les données pour un patient étaient indisponibles ou après son décès.

IV.1.2. Résultats

IV.1.2.1. Descriptif de la population

Parmi les 112 745 patients du SNDS, 60 559 (53,7%) ont été identifiés comme ayant eu au moins une délivrance d'au moins un des traitements considérés. Si l'échantillon était restreint aux traitements spécifiques uniquement (groupe « hors AMM » exclu), 53 469 patients (47,4%) étaient conservés. Les interférons étaient la substance active la plus prescrite avec 31 849 patients (28,2%) ayant eu au moins une délivrance suivi par l'acétate de glatiramère avec 15 272 bénéficiaires (13,5%).

En considérant les voies d'administration, 43 076 (38,2%), 22 985 (20,4%) et 8 240 patients (7,3%), avaient eu respectivement au moins une délivrance d'un injectable, d'un oral ou une perfusion de natalizumab au cours de leur suivi. Concernant les traitements historiques hors AMM, 9 494 patients (8,4%) ont eu au moins une délivrance.

IV.1.2.2. Distribution des traitements de 2010 à 2015

Les différents chronogrammes permettent de visualiser l'évolution de la prévalence des traitements au cours du suivi (Figure 20). L'impact de l'arrivée des nouveaux traitements oraux sur le marché est observable à partir de 2012 pour le fingolimod, fin 2014 pour le teriflunomide et mi-2015 pour le

diméthyl fumarate. La part des traitements oraux est ainsi passée de 2,1% en janvier 2013 à 6,8% à la fin de la période d'étude. Parallèlement à cette augmentation, le nombre de patients sous injectable a diminué progressivement passant de 17,7% en janvier 2010 à 13,3% en décembre 2015 avec un pic à 20,3% en janvier 2013. A l'opposé, la prévalence du groupe « hors AMM » est restée stable autour de 2,5% tout comme les perfusions de natalizumab autour de 3,0%.

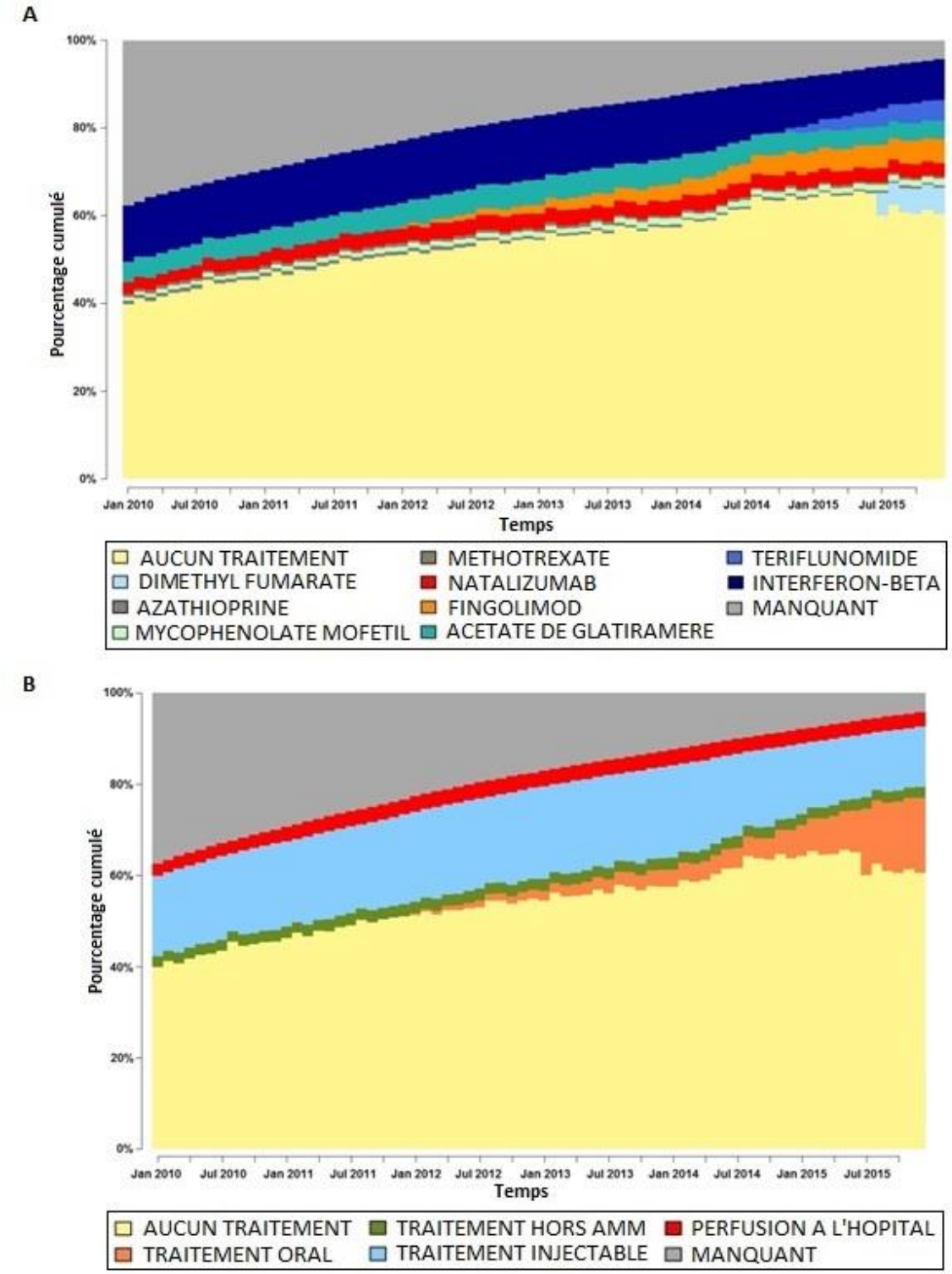


Figure 20-Chronogrammes de la répartition des traitements sur la période d'étude 2010-2015 selon (A) la substance active et (B) la voie d'administration (N=112 745)

L'étude de l'évolution du nombre d'initiateurs a montré des résultats concordants permettant néanmoins une meilleure vision de l'impact des traitements oraux (Figure 21). Ainsi, le nombre de nouveaux utilisateurs d'un injectable a été divisé par deux entre 2012 et 2015, notamment pour les interférons passant de 4 303 nouveaux utilisateurs en 2012 à 1 623 en 2015. *A contrario*, les trois traitements oraux ont connu une explosion dans le nombre d'initiations dès leur arrivée sur le marché, notamment le teriflunomide en augmentation avec près de 5 600 nouveaux utilisateurs en 2015. Le fingolimod a connu une forte prescription la première année de son introduction (2 496 nouveaux utilisateurs en 2012) puis une stagnation dans le nombre de nouveaux initiateurs autour de 2 000 par an. L'augmentation la plus marquée a été observée pour le diméthyl fumarate avec plus de 8 000 nouveaux initiateurs en seulement huit mois de commercialisation en officine.

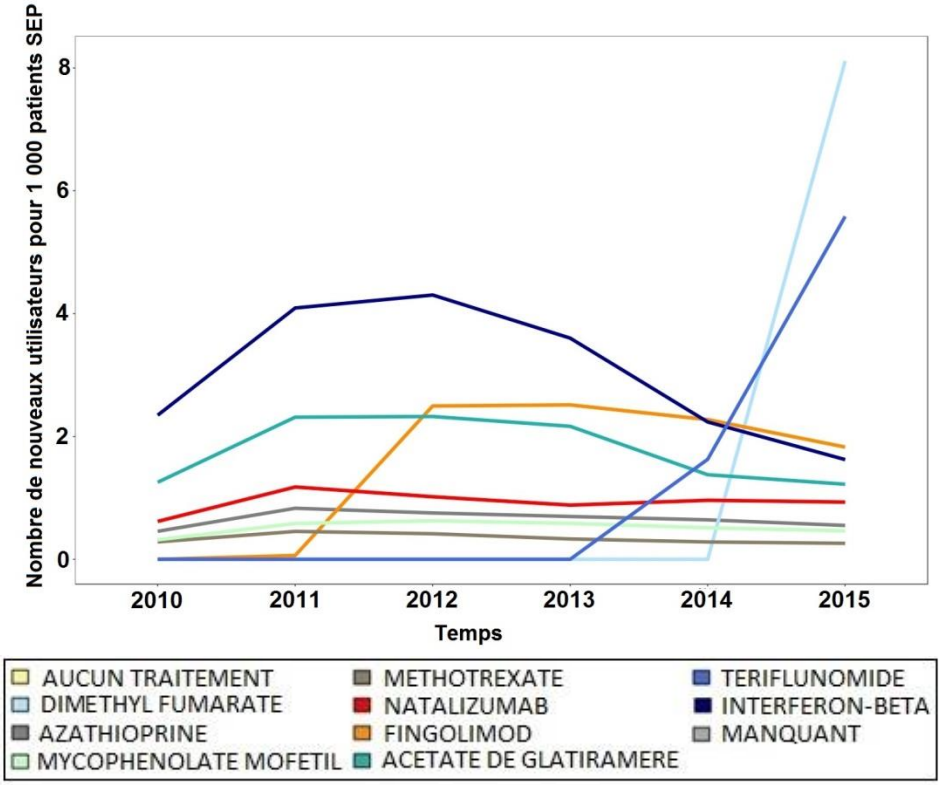


Figure 21-Evolution du nombre annuel de nouveaux utilisateurs de chaque substance active sur 2010-2015

IV.1.2.3. Séquences thérapeutiques

Parmi les patients ayant reçu au moins un des traitements considérés, 61,3% ont eu la délivrance de seulement une substance active sur l'ensemble du suivi et à l'inverse 71 (0,1%) ont reçu cinq traitements ou plus. L'ensemble des séquences thérapeutiques observées dans la base sont synthétisées dans la Figure 22. Pour rappel, seules les phases sans traitement de plus de 6 mois sont représentées afin d'alléger la représentation. La plus longue séquence thérapeutique observée était composée de 29

phases, c'est-à-dire sous-séquences d'un traitement considéré, successives correspondant à une personne suivant une bithérapie interféron-azathioprine. Le sunburst a permis de mettre en évidence la grande hétérogénéité des séquences thérapeutiques dans la SEP avec plus de 2 500 séquences différentes identifiées, la plus fréquente étant composée d'une seule phase correspondant à la prise d'un interféron (16,9%) correspondant à une certaine stabilité dans le traitement pour certains patients (Figure 22).

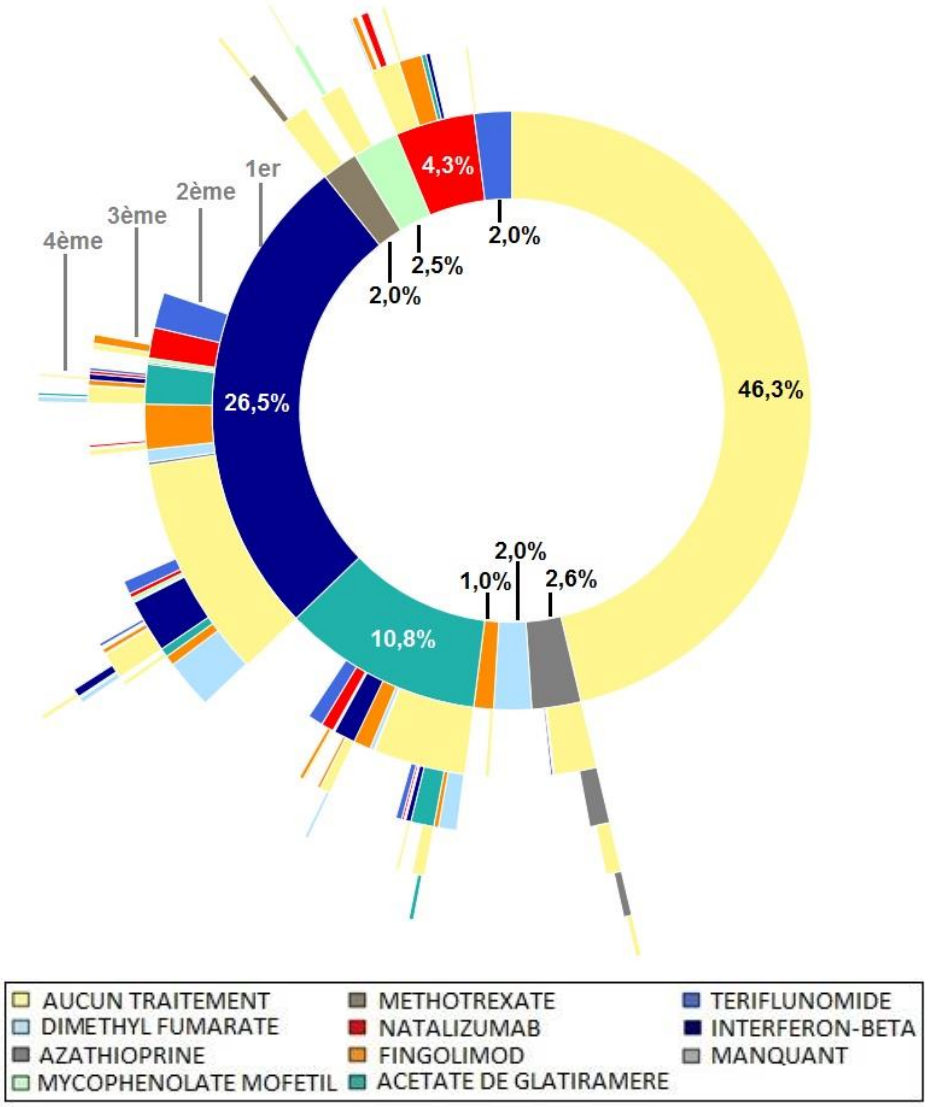


Figure 22-Sunburst des séquences thérapeutiques observées sur la période d'étude 2010-2015 (N=112 745)

Cette stabilité a été constatée dans les diagrammes de Sankey permettant la visualisation des flux de patients entre chaque traitement (Figure 23). Par ailleurs, l'étude de ces derniers a permis de retrouver

l'impact de l'introduction du diméthyl fumarate dans les séquences thérapeutiques. En effet, 78,5% des patients démarraient le diméthyl fumarate en 2015 sans présence d'un traitement antérieur.

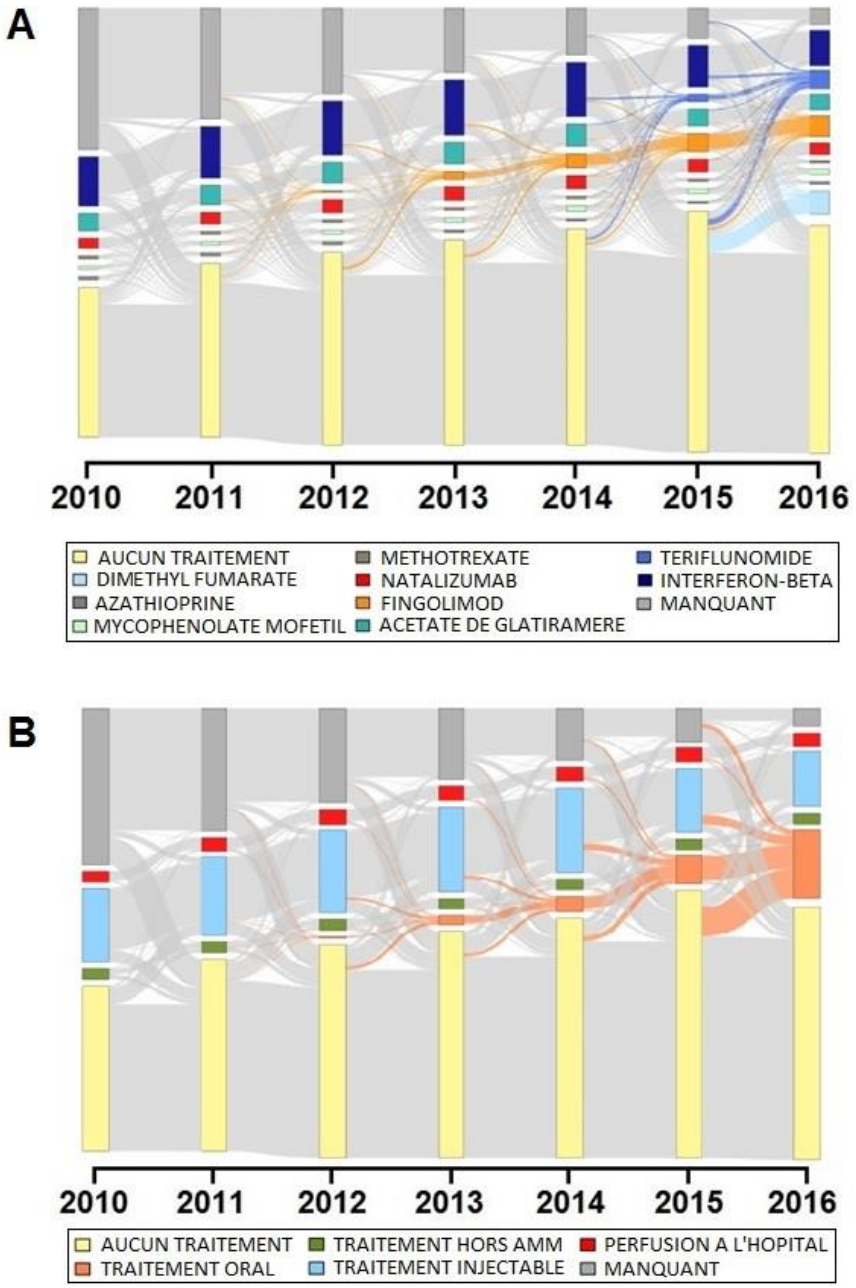


Figure 23-Diagrammes de Sankey de la répartition des traitements sur la période d'étude 2010-2015 selon (A) la substance active et (B) la voie d'administration (N=112 745)

IV.2. Suivi du plan de gestion des risques dans l'EGB

IV.2.1. Méthodes

Comme évoqué précédemment, une étude de faisabilité concernant le suivi du PGR des nouveaux traitements oraux de la SEP a été réalisée dans l'EGB uniquement au début de l'année 2016 (l'extraction SNDS n'était donc pas disponible), avec en exemple le cas du fingolimod disponible en officine depuis 2012. La base considérée a donc été celle formée des patients identifiés dans l'EGB entre 2007 et 2012 pour lesquels des consommations de soins étaient disponibles du 1er janvier 2007 au 31 mars 2015.

La prescription de fingolimod nécessite une prise en charge particulière tant en amont qu'en aval de la première délivrance effectuée à l'hôpital. Cette prise en charge spécifique due à des effets indésirables observés durant les essais cliniques est décrite dans le PGR. Ainsi, suite à une forme grave mortelle de la varicelle ayant eu lieu, la mise sous fingolimod impose que le patient soit immunisé contre le virus soit par infection préalable ou vaccination. La molécule pouvant entraîner un ralentissement du rythme cardiaque (bradycardie) dans les premières heures suivant l'instauration du traitement, une surveillance cardiaque de 6 heures a été instaurée nécessitant une première prise du traitement dans un établissement de santé. Une fois le traitement instauré, une élévation des enzymes hépatiques peut être observée, d'où un suivi régulier de leurs taux. De même, un œdème maculaire pouvant apparaître dans les premiers mois a été constaté chez 0,6% des patients traités, d'où l'instauration d'un contrôle ophtalmique au bout de 3 à 4 mois. Le PGR est synthétisé dans la Figure 24.

Figure 24-Recommandations de suivi des patients sous fingolimod d'après le PGR (Laboratoire Novartis)

La première délivrance de fingolimod n'étant pas identifiable dans l'EGB, le début du traitement a été approximé par la première délivrance du traitement en officine en ville. Concernant les actes à

effectuer en amont (bilan pré-thérapeutique), un acte d'imagerie du type IRM, justifiant la mise sous traitement doit être réalisé. Par ailleurs, dans les six mois précédant la première prescription, doivent être réalisés un bilan hépatique complet (transaminases, bilirubine, NFS (Numération de la Formule Sanguine) plaquettes) ainsi qu'un test de présence du Virus Varicelle-Zona (VZV) (170). Une fois la première délivrance effectuée, des bilans hépatiques doivent être effectués à intervalles réguliers ainsi qu'un suivi neurologique (Figure 24). Afin de vérifier le suivi de ces préconisations, les IRM et les actes de biologie ont été repérés dans l'EGB (algorithmes améliorés ensuite dans le SNDS). Les consultations avec un ophtalmologue libéral, un neurologue libéral et un médecin généraliste libéral ont également été considérées, les consultations externes étant inaccessibles au moment des analyses (cf. *VI.1.1.1. Consultations avec un spécialiste libéral ou dans un établissement de santé privé* pour la définition de l'algorithme d'identification).

IV.2.1.1. Identification des actes d'imagerie dans l'EGB

Les IRM ont été repérées dans la table *EB_CAM_F* de l'EGB concernant les actes CCAM effectués en ville et dans la table *MCO_CCAM* du PMSI MCO pour les actes effectués lors de séjours, les actes externes effectués à l'hôpital étaient inaccessibles dans l'EGB au moment des analyses. Un filtre a été effectué dans les deux cas (sur la variable *CAM_PRS_IDE* pour les actes de ville et *CDC_ACT* pour les actes hospitaliers) au moyen des codes CCAM des IRM séparés en deux catégories, IRM médullaires (colonne vertébrale et moelle épinière) et IRM cérébrales (Table 8), en prenant en compte l'injection d'un produit de contraste type gadolinium ou non. Pour chaque imagerie identifiée, la date d'exécution a été conservée afin de calculer la durée entre ce dernier et la première délivrance de fingolimod en officine.

Table 8-Codes CCAM considérés pour l'identification des IRM

Code CCAM	Catégorie	Libellé
AAQN004	IRM cérébrale	Remnographie [IRM] fonctionnelle du cerveau pour étude des fonctions motrices
AAQN900	IRM cérébrale	Remnographie [IRM] fonctionnelle du cerveau pour étude des fonctions visuelles
AAQN901	IRM cérébrale	Remnographie [IRM] fonctionnelle du cerveau pour étude des fonctions phasiques
ACQJ002	IRM cérébrale	Remnographie [IRM] du crâne et de son contenu, avec injection intraveineuse de produit de contraste
ACQN001	IRM cérébrale	Remnographie [IRM] du crâne et de son contenu, sans injection intraveineuse de produit de contraste
LHQJ001	IRM médullaire	Remnographie [IRM] d'un ou de deux segments de la colonne vertébrale et de son contenu, avec injection intraveineuse de produit de contraste
LHQJ002	IRM médullaire	Remnographie [IRM] de 3 segments de la colonne vertébrale ou plus et de son contenu, avec injection intraveineuse de produit de contraste
LHQN001	IRM médullaire	Remnographie [IRM] d'un ou de deux segments de la colonne vertébrale et de son contenu, sans injection intraveineuse de produit de contraste
LHQN002	IRM médullaire	Remnographie [IRM] de 3 segments de la colonne vertébrale ou plus et de son contenu, sans injection intraveineuse de produit de contraste
ZZQN001	IRM médullaire	Remnographie [IRM] comportant 6 séquences ou plus

IV.2.1.2. Identification des actes de biologie dans l'EGB

Seuls les actes biologiques effectués en ville ont pu être repérés dans l'EGB dans la table *EB_BIO_F* au travers de la variable *BIO_PRS_IDE*. En effet, les actes biologiques effectués lors de séjours hospitaliers n'apparaissent pas dans le PMSI. Les différents codes de la NABM regroupés selon l'examen considéré sont rassemblés dans la Table 9. De manière analogue au paragraphe précédent, la date d'exécution de chaque acte biologique a été isolée pour calculer la durée depuis le début du traitement.

Table 9-Codes NABM considérés pour l'identification des actes de biologie

Code NABM	Catégorie	Libellé
0516	Bilan hépatique	Alanine aminotransférase (ALAT, TGP) (sang)
0517	Bilan hépatique	Aspartate aminotransférase (ASAT, TGO) (sang)
0522	Bilan hépatique	Transaminases (ALAT et ASAT, TGP et TGO) (sang)
1104	NFS	Hémogramme y compris plaquettes (NFS, NFP)
1777	VZV	Varicelle-Zona : Diagnostic d'une infection récente ou d'une récurrence : IGG et IGM par EIA
1778	VZV	Varicelle-Zona : SD infection récente ou récurrence : IGG + IGM par IFI
3777	VZV	Varicelle-Zona : Diagnostic d'une infection récente ou d'une récurrence : IGG et IGM par EIA+IT
3778	VZV	Varicelle-Zona : SD infection récente ou récurrence : IGG + IGM par IFI+ITERARIF

IV.2.2. Résultats

Parmi les 70 bénéficiaires de l'EGB considérés, l'échantillon était formé de 23 hommes et 47 femmes (respectivement 32,9% et 67,1%) ayant un âge médian au début du traitement par fingolimod de 39 ans (min=25 ans ; max=79 ans). Au total, 94,3% de la population était en ALD SEP à la mise en place du traitement avec une durée médiane de 6,4 ans depuis le début de l'ALD. La durée moyenne sous fingolimod était de 1,3 ans (écart-type=0,9 ans) avec l'arrêt du traitement pour 20 patients dont 9 qui l'ont repris ultérieurement. Concernant la prescription du traitement, le fingolimod a été pris en relais d'un autre traitement chez 60 patients (85,7%) : 37 après un interféron-β, 15 l'acétate de glatiramère, 7 le natalizumab et 1 le mycophénolate mofétil.

Le suivi des recommandations est synthétisé dans la Figure 25. Parmi les 70 patients, un bilan hépatique et un dépistage du VZV en amont de la première délivrance ont été retrouvés pour respectivement 47 (67,1%) et 30 (42,9%) d'entre eux. Une IRM a été retrouvée pour seulement 1 patient (1,4%) dans les 12 mois précédant le début du traitement. Concernant le suivi, le nombre de bilans hépatiques diminue fortement au cours du temps avec toujours moins de 50% des bilans retrouvés en autorisant une fenêtre de 7 jours autour des dates anniversaire. Une consultation avec un ophtalmologue libéral entre 3 et 4 mois plus ou moins 7 jours après la première délivrance a été retrouvée pour 5 patients (7,1%).

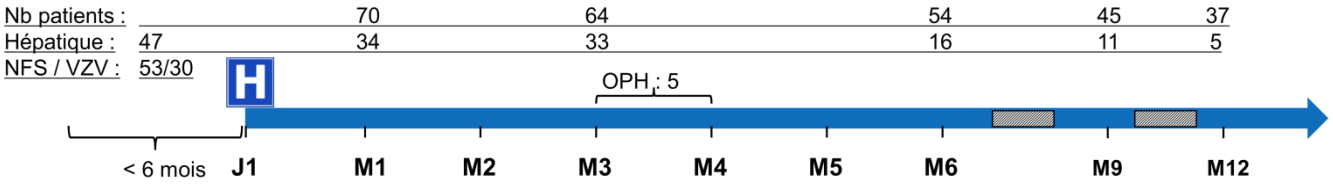


Figure 25-Synthèse des résultats concernant le suivi des recommandations dans l'EGB suite à la délivrance de fingolimod (n=70)

IV.3. Discussion

IV.3.1. Discussion des résultats

Ce premier axe de thèse s'intéressait à l'utilisation des traitements de fond dans la SEP et à la faisabilité de réaliser à partir des bases médico-administratives le suivi du PGR des nouvelles molécules sur le marché avec en exemple le cas du fingolimod. Ce travail était une étape préalable indispensable à la caractérisation de la typologie de parcours de soins construite dans l'axe principal de cette thèse. Elle a également permis de mettre en évidence l'impact de l'arrivée des traitements oraux dans la prise en charge de la maladie, et les changements inhérents dans les prescriptions. La seconde partie sur le PGR avait pour but de tester la possibilité du suivi de ce dernier dans l'EGB avant une possible extension à la base du SNDS complète.

Plus de la moitié des patients (53,7%) a eu au moins une délivrance d'un traitement de la SEP au cours du suivi. Ce chiffre s'élevait à 47,4% si on ne considérait que les traitements spécifiques (groupe « hors AMM » exclu). Ce résultat est cohérent avec les données de l'OFSEP dans lesquelles 63,8% des patients ont été identifiés comme recevant un traitement disponible dans le SNDS, groupe « hors AMM » inclus (Annexe D). La différence entre les deux bases peut s'expliquer par le probable biais de recrutement dans l'OFSEP où seuls les patients suivis dans les centres experts de la SEP sont inclus avec de fortes disparités régionales (171). Dans une étude récente publiée en 2018 à partir de données issues de onze centres experts inclus dans la cohorte de l'OFSEP (il s'agit donc d'un sous-groupe de l'ensemble des données OFSEP), Defer et al. (172) ont trouvé une proportion de 54,8% de patients traités avec un traitement de première ligne. Le fait que cette étude ait été basée sur les patients vus dans des centres experts explique sans doute cette proportion élevée, en accord avec les données de l'OFSEP, comparé à notre résultat. Par ailleurs, en ne considérant que l'année 2013 sur laquelle l'étude de coûts de Lefeuvre et al. (105) a été effectuée, 37,3% des personnes identifiées avaient bénéficié d'au moins une délivrance d'un traitement spécifique à la SEP (groupe « hors AMM » exclu), ce qui est en accord avec la proportion d'un tiers de patients traités citée dans l'article.

Concernant l'utilisation des traitements de fond, l'explosion de l'utilisation des traitements oraux a également pu être observée dans les données de l'OFSEP, notamment celle du diméthyl fumarate, grâce à des données plus récentes allant jusqu'en 2017 (Annexe D). Cette part importante prise par le diméthyl fumarate sur le marché a par ailleurs été visible plus précocement dans l'OFSEP grâce aux données sur les essais cliniques non disponibles dans le SNDS. En effet, les traitements oraux étaient fortement attendus par les patients, dus notamment au passage à une prise journalière sans injection. Cette attente a par ailleurs été amplifiée compte-tenu du délai entre les premiers résultats des essais thérapeutiques, l'obtention de l'AMM et enfin la mise à disposition dans les pharmacies. Ce délai a alors entraîné un effet de stock à l'arrivée de ces molécules causé par les patients ayant attendu ces dernières pour commencer leur traitement et ceux souhaitant changer pour des raisons de tolérance ou de convenance personnelle.

La grande variabilité des séquences thérapeutiques trouvée dans le SNDS (plus de 2 500 séquences) a été retrouvée dans les données de l'OFSEP, où plus de 2 700 séquences différentes ont pu être identifiées (Annexe D). Les résultats concernant le nombre de molécules administrées dans les séquences dans le SNDS (61,3% des patients traités avec une seule substance active) étaient cohérents avec ceux de obtenus dans l'OFSEP (60,6% des patients traités avec une seule substance).

Concernant la deuxième partie du travail sur le suivi du PGR dans les bases médico-administratives, aucune donnée publiée n'est disponible à l'heure actuelle à notre connaissance nous permettant d'effectuer une comparaison de nos résultats. Nos résultats suggèrent un faible niveau d'adhésion aux recommandations, mais il importe de souligner d'emblée les limites de l'EGB pour ce type d'évaluation, notamment dans le suivi de la réalisation des actes biologiques et la détection des IRM, malgré l'introduction d'une fenêtre de battement dans la réalisation des actes. Cependant, cette conclusion est à nuancer du fait que cette étude de faisabilité a été réalisée début 2016 alors que les connaissances à l'époque des bases de données et notamment que les ACE n'étaient pas accessibles. La prise en compte des consultations et actes externes auraient permis de mieux approcher respectivement le suivi ophtalmologique et le suivi par imagerie et ainsi ne pas (ou moins) sous-estimer le suivi du PGR. Néanmoins, même en ayant ces informations à disposition le suivi des actes biologiques en intra-hospitalier est impossible à partir des bases médico-administratives, ces informations étant stockées dans les entrepôts de données de chaque centre hospitalier. De plus, certains centres peuvent avoir mis en place une organisation particulière en interne du suivi des traitements qui sera elle aussi intraçable dans le SNDS.

IV.3.2. Forces et limites

Le principal avantage de l'utilisation du SNDS pour l'étude des traitements dans la SEP est l'exhaustivité de cette base contrairement aux données usuellement collectées dans les cohortes telles que celle de l'OFSEP. En effet, outre la quasi-exhaustivité de la population commune à l'ensemble des axes et discutée ultérieurement dans la discussion générale, le SNDS permet une exhaustivité sur la délivrance des traitements de fond. Ainsi, la totalité des principaux traitements de la SEP est repérable dans les données. Les seules molécules non identifiables, que sont la mitoxantrone, l'alemtuzumab et la cyclophosphamide, ne représentent qu'une très faible part des patients traités en France (2,8% des patients dans l'OFSEP (Annexe D)). Cette proportion est sans doute non-extrapolable à l'ensemble des patients en France par rapport à la réalité du fait que ces molécules sont uniquement délivrées en perfusion dans les établissements de santé et donc dans les centres experts, où est recrutée la majeure partie des patients de l'OFSEP (à l'exception de quelques réseaux régionaux en Alsace et du registre lorrain). De plus, une hypothèse raisonnable est que la majorité de ces patients soit traitée par une autre molécule au cours de la période de suivi consécutive de six années, ce qui entraîne qu'ils sont inclus dans le total de patients traités. Cependant, il est important de garder à l'esprit que cette identification est uniquement possible grâce au remboursement de l'ensemble des traitements considérés, entraînant donc un renseignement dans les bases médico-administratives.

Une seconde force de ce travail repose sur les hypothèses et le data management effectués. Ainsi, l'expertise développée sur les données a permis de limiter les artefacts liés aux différentes régularisations effectuées par l'Assurance Maladie et a vraisemblablement conduit à un meilleur décompte des délivrances. Ainsi, l'algorithme développé dans ce travail a conduit à une limitation de la surestimation des délivrances et du nombre de boîtes délivrées, constituant un point important dans un travail de l'étude de l'observance des traitements par exemple. Par ailleurs, l'algorithme développé dans le cadre de l'identification des délivrances de natalizumab en milieu hospitalier a reposé sur un travail de data management permettant de ne garder que les délivrances de médicaments concordantes avec les séjours hospitaliers à disposition. Ceci a ainsi permis de ne pas surestimer la délivrance de ce traitement. De plus, cette molécule étant facturée en sus du GHS sur l'ensemble du suivi, il apparaît comme étant peu probable que certaines injections soient manquantes dans les données.

En dépit de ces hypothèses, il a pu être noté une grande variabilité dans le nombre de phases dans les séquences thérapeutiques, notamment dans le cas de bi-thérapies. Il s'agit là d'une des limites du SNDS due à la quasi-exhaustivité des bases, même si cela est à relativiser en regard du nombre de séquences thérapeutiques également identifiées dans l'OFSEP. De plus, en cas de deux délivrances de deux molécules distinctes ayant été effectuées le même mois chez un même patient, seule la première délivrée a été considérée afin de ne pas introduire de double comptage des patients. Cette limite a donc

tendance à potentiellement sous-estimer le nombre mensuel de patients ayant une délivrance d'un traitement particulier et à décaler ce patient le mois d'après dans le cadre d'un switch de traitement par exemple. A l'inverse, les données de l'OFSEP recensent les dates précises de début et d'arrêt de traitement, permettant ainsi de limiter ce biais.

Concernant la partie sur le suivi suite à une initiation de fingolimod, outre les limites évoquées précédemment sur l'identification des actes de biologie et la nécessité de prendre en compte les ACE, une limite majeure réside dans l'impossibilité d'identifier la première prise du traitement à l'hôpital. En effet, cette première prise a le plus souvent lieu en hôpital de jour sans acte discriminant permettant une identification dans le PMSI. Afin de pallier à ce problème, la date de première délivrance du fingolimod a été assimilée au début du traitement, mais il serait nécessaire de faire varier cette date de un à plusieurs jours au cours d'analyses de sensibilité afin de voir l'influence sur les résultats, même s'ils ne devraient pas souffrir d'une trop grande variabilité à cause de la fenêtre de battement introduite dans l'exécution des actes.

Enfin, une hypothèse majeure commune aux deux parties de cet axe repose sur la nature des données recueillies. En effet, le SNDS étant une base médico-administrative, les informations recueillies sont uniquement des données de délivrances et non de prise du médicament par le patient. L'hypothèse principale de ce travail est donc que si un patient continue à retourner chercher son traitement à la pharmacie, ou à aller dans un établissement de santé pour ses perfusions, c'est qu'il prend effectivement son traitement. Cette hypothèse est très répandue dans le domaine de l'étude de l'utilisation des traitements (*drug utilization research*) de manière globale du fait des données à disposition. En effet, l'obtention de données sur la prise réelle d'un médicament est très difficile dans la vie réelle, nécessitant un matériel adapté notamment utilisé dans les études d'observance.

Certes, plusieurs améliorations pourraient être apportées afin d'améliorer l'identification des traitements notamment dans le cadre de switches ou de bi-thérapies. Les variations alors induites seraient probablement minimes au vu de la taille de la population d'étude et de l'usage des données produites, servant de base descriptive principalement. Une autre perspective pouvant être envisagée pour ce travail est une comparaison des pratiques de prescription au niveau territorial afin d'analyser les variations spatiales de prise en charge en France, des « écoles » de prescription liées aux centres experts ayant pu être mises en évidence par l'OFSEP.

V. Quantification du niveau de handicap moteur

Un des principaux inconvénients des bases médico-administratives telles que le SNDS est l'absence totale de données cliniques hormis les diagnostics des séjours hospitaliers. L'apparition d'un handicap, notamment un handicap moteur, est la traduction principale de la SEP. C'est pourquoi, un volet exploratoire dédié à la quantification de dernier dans les bases médico-administratives a été considéré dans cette thèse.

V.1. Méthodes

Afin de quantifier le niveau de handicap moteur d'un patient, un score EDSS « estimé », basé sur les prescriptions des aides techniques à la marche, a été construit dans la population du SNDS après une phase de développement préliminaire sur l'EGB.

V.1.1. Identification des aides techniques à la marche

Les aides techniques à la marche sont des dispositifs médicaux remboursés quand ils font l'objet d'une prescription. Par conséquent, l'ensemble des informations les concernant peut être retrouvé dans le SNDS et plus particulièrement dans la table *ER_TIP_F*, correspondant aux prestations affinées sur les dispositifs médicaux.

Cette dernière recense ainsi l'ensemble des achats, locations et entretiens relatifs aux aides techniques à la marche, chacune de ces actions fournies étant stockées dans la variable *TIP_PRS_TYP*. Les dates de début et de fin d'une location ou d'un entretien sont disponibles respectivement dans les variables *TIP_ACL_DTD* et *TIP_ACL_DTF*. Dans le cas d'un achat, la date de fin *TIP_ACL_DTF* n'étant logiquement pas renseignée, il a été considéré que le patient avait à sa disposition le matériel jusqu'à la fin de son suivi. Afin de pouvoir faire le lien avec les patients, la table a été jointe à la table principale des prestations *ER_PRS_F*. Cette jointure a permis de récupérer la date de début de location/achat en cas de date manquante au travers de la date d'exécution de la prestation *EXE_SOI_DTD*. Par ailleurs, pour les prestations de location sans date de fin *TIP_ACL_DTF* renseignée, elle a été approximée en utilisant la variable *TIP_ACT_QSN* codant pour le nombre de semaines ou de jours de location dépendant du matériel.

Les aides techniques considérées ont été les suivantes : les cannes, déambulateurs, VHP électriques et manuels, lits médicalisés, soulève-malades et les matériels anti-escarres de classe II et III (matériel pour les patients présentant un risque d'escarre moyen à élevé ou alités plus de 15h par jour). Les aides comprenant simultanément un lit médicalisé et un matériel anti-escarres de classe II ont également été identifiées. L'ensemble de ces aides a été repéré au moyen des codes LPP de chaque dispositif dans la variable *TIP_PRS_IDE* codant les matériels et entretiens de ces derniers (Annexe E). Les codes ont été identifiés sur la période d'étude 2010-2015, mais également sur la période précédant l'identification pour les patients ayant un suivi disponible afin de quantifier le niveau de handicap à l'inclusion.

Pour quelques patients, seul le code d'entretien de matériel d'aide technique (sans code d'achat antérieur) est identifiable dans les bases. Nous avons alors considéré qu'un achat du matériel à une date antérieure à celle d'identification avait été réalisé (n= 3978 patients concernés (3,5%)). Dans le cas contraire, l'entretien n'était pas considéré. Concernant les locations, une fenêtre de battement de 30 jours entre la fin de location d'un équipement et le début d'une nouvelle location du même dispositif a été utilisée pour juger de la continuité de la séquence, aucune coupure dans la séquence de location du patient n'étant alors retenue, de même entre une location et un achat. Afin de ne pas biaiser l'estimation du score avec des locations dues à des épisodes traumatiques, des fractures ou des accidents entre autres, seules les locations de plus de 4 mois consécutifs ont été considérées dans le calcul.

V.1.2. Construction du score EDSS « estimé »

Les dates de début et de fin de location et d'achat de chacune des aides techniques à la marche ont ensuite été compilées afin de créer la séquence des aides de chaque patient sur l'ensemble de la période d'étude.

Basé sur les critères moteurs de l'échelle EDSS, un algorithme mettant en correspondance les aides techniques et un score EDSS « estimé » a été défini (Table 10). Les scores « $\geq$ *8,5* » nécessitant de prendre en compte deux aides simultanément, les chevauchements entre équipements ont donc été considérés. Dans une étape de consolidation des résultats, les chevauchements entre scores EDSS estimés sur une même période ont été supprimés en gardant le score EDSS le plus élevé sauf en cas d'apparition d'un nouvel équipement.

Table 10-Attribution des scores EDSS « estimés » selon les aides techniques à la marche considérées

Aide technique à la marche	Score EDSS estimé
(Lit médicalisé & matériel anti-escarres classe II ou III) **ou** (Lit médicalisé & soulève-malade)	≥ 8,5
VHP électrique	7,5
VHP manuel	7
Déambulateur	6,5
Canne	≥ 5,5-6
Aucun équipement	≤ 5

VHP : Véhicule pour Personne Handicapée.

V.2. Résultats

V.2.1. Aides techniques à la marche

Parmi les 112 745 personnes de la population d'étude du SNDS, au moins une aide technique à la marche a été retrouvée pour 44 149 (39,2%) d'entre eux. Ces patients avaient un âge médian de 52 ans en 2010 pour un sex-ratio de 2,3 femmes pour un homme.

La répartition des patients selon les différents équipements considérés est présentée dans la Table 11. L'aide technique la plus prescrite était la canne avec près d'un quart des patients ayant eu au moins une délivrance au cours du suivi. Au total, 23 670 patients (21,0%) ont vu un VHP manuel ou électrique leur être prescrit, dont 6 060 (25,6%) ayant eu la prescription des deux types de fauteuils.

Table 11-Prescription des aides techniques à la marche sur la période d'étude 2010-2015 dans la population identifiée dans le SNDS (N=112 745)

Aide technique	Nombre de patients (%)
Canne	27 317 (24,2%)
Déambulateur	14 351 (12,7%)
VHP manuel	21 876 (19,4%)
VHP électrique	7 854 (7,0%)
Soulève-malade	5 833 (5,2%)
Lit médicalisé	11 025 (9,8%)
Matériel anti-escarres de classe II	20 946 (18,6%)
Lit médicalisé et matériel anti-escarres de classe II dans le même équipement	2 973 (2,6%)

Plusieurs aides techniques pouvant avoir été prescrites et délivrées à un même patient, ce dernier peut donc apparaître dans plusieurs lignes. VHP : Véhicule pour Handicapé Physique.

V.2.2. Score EDSS « estimé »

En considérant à présent les scores EDSS « estimés », 23 326 patients (20,7%) présentaient un score « estimé » différent de « *≤ 5* », symbolisant l'absence de handicap moteur, à leur date d'identification dans le SNDS (Table 12). Aux dernières nouvelles (31 décembre 2015 ou décès), 40 376 patients (35,8%) présentaient un score correspondant à un handicap moteur (score≥5,5). A l'inclusion, 2,1% des patients avaient un score EDSS estimé supérieur ou égal à 8,5, correspondant à un alitement quasi-permanent.

Table 12-Score EDSS « estimé » à l'inclusion et aux dernières nouvelles dans la population identifiée dans le SNDS (N=112 745)

Score EDSS estimé à l'inclusion	Nombre de patients (%)	Score EDSS estimé aux dernières nouvelles	Nombre de patients (%)
≤ 5	89 419 (79,3%)	≤ 5	72 369 (64,2%)
≥ 5,5-6	9 142 (8,1%)	≥ 5,5-6	17 417 (15,4%)
6,5	2 624 (2,3%)	6,5	6 442 (5,7%)
7	6 615 (5,9%)	7	8 449 (7,5%)
7,5	2 539 (2,3%)	7,5	3 722 (3,3%)
≥ 8,5	2 406 (2,1%)	≥ 8,5	4 346 (3,9%)

En comparant les EDSS à l'inclusion et ceux aux dernières nouvelles (en médiane 6,0 années pour rappel), une stagnation a été le plus souvent constatée (72,5%) et une augmentation du score pour 23,0% de la population d'étude (Table 13). Les 5% de la population restante ont eu une diminution de leur score EDSS « estimé » au cours du suivi. Les scénarii correspondant à cette dernière situation pourraient correspondre à une poussée au début du suivi, ou à un passage en EHPAD (Etablissement d'Hébergement des Personnes Agées Dépendantes) à la fin du suivi entraînant l'arrêt des locations des équipements.

Table 13-Répartition des patients en fonction de l'évolution du score EDSS « estimé » entre l'inclusion et les dernières nouvelles dans la population du SNDS (N=112 745)

		Score EDSS estimé aux dernières nouvelles					
		≤ 5	≥ 5,5-6	6,5	7	7,5	≥ 8,5
Score EDSS estimé à l'inclusion	≤ 5	69 804 (61,9%)	10 635 (9,4%)	3 261 (2,9%)	3 782 (3,4%)	956 (0,8%)	981 (0,9%)
	≥ 5,5-6	334 (0,3%)	6 010 (5,3%)	1 074 (1,0%)	1 014 (9,0%)	398 (0,4%)	312 (2,8%)
	6,5	281 (0,2%)	205 (0,2%)	1 067 (0,9%)	484 (0,4%)	249 (0,2%)	338 (0,3%)
	7	834 (0,8%)	426 (0,4%)	744 (0,7%)	2 700 (2,4%)	779 (0,7%)	1 132 (1,0%)
	7,5	348 (0,3%)	97 (0,1%)	188 (0,2%)	227 (0,2%)	1 134 (1,0%)	545 (0,5%)
	≥ 8,5	768 (0,7%)	44 (0,0%)	108 (0,1%)	242 (0,2%)	206 (0,2%)	1 038 (0,9%)

L'algorithme avait été raffiné lors de la phase de test sur l'EGB afin de prendre en compte ces particularités liées au suivi du patient. Aucune baisse de handicap entre le début et la fin du suivi n'était alors observée (Table 14). Cependant, ce travail n'a pas pu être implémenté pour la population du SNDS par manque de temps.

Table 14-Répartition des patients en fonction de l'évolution du score EDSS « estimé » entre l'inclusion et les dernières nouvelles dans la population due l'EGB (N=650)

		Score EDSS estimé aux dernières nouvelles					
		≤ 5	≥ 5,5-6	6,5	7	7,5	≥ 8,5
Score EDSS estimé à l'inclusion	≤ 5	381 (58,6%)	80 (12,3%)	20 (3,1%)	54 (8,3%)	12 (1,8%)	57 (8,8%)
	≥ 5,5-6	-	-	1 (0,1%)	-	-	-
	6,5	-	-	-	-	-	-
	7	-	-	-	7 (1,1%)	3 (0,5%)	13 (2,0%)
	7,5	-	-	-	-	10 (1,5%)	6 (0,9%)
	≥ 8,5	-	-	-	-	-	6 (0,9%)

V.3. Discussion

V.3.1. Discussion des résultats

Le but de ce deuxième axe de la thèse était de construire un algorithme permettant l'estimation d'un score EDSS à partir des données de remboursement des aides techniques à la marche afin de fournir une nouvelle variable permettant de caractériser les parcours de soins construits lors du dernier axe de la thèse.

Sur la période d'étude, une aide technique à la marche a été prescrite à près de 40% des patients suivis, dont plusieurs à une même personne signalant une élévation du score EDSS au cours du suivi. Ainsi cette augmentation du score EDSS « estimé » a été constatée chez 23% de la population d'étude. Cette faible part d'augmentation peut s'expliquer par le choix de comparer la mesure à l'inclusion à celle estimée aux dernières nouvelles, mais aussi par le délai nécessaire pour atteindre un score EDSS élevé. En effet, le délai médian pour l'atteinte d'un EDSS 7 à partir d'un EDSS 4 est de 12,1 ans par exemple dans les formes rémittentes et progressive d'emblée (24). Par ailleurs, une majeure partie de la population d'étude présentait un score EDSS « estimé » inférieur ou égal à 5 (79,3% à l'inclusion et 64,2% aux dernières nouvelles) pouvant s'expliquer par une maladie récente pour la majorité des patients (durée médiane approximée par la durée entre la mise en ALD et l'identification de 6,0 années). Ce résultat est en cohérence avec les données de la littérature faisant état d'une durée entre 20 et 30 ans pour le besoin d'une aide à la marche et donc un score EDSS égal à 6 (24).

Ce travail étant pionnier en France, il n'existe actuellement que peu de données permettant une comparaison des résultats obtenus à notre connaissance. En effet, la majorité des publications sur les aides techniques sont des études de coûts au niveau populationnel et celles portant sur le score EDSS sont principalement des études sur l'évolution de ce dernier et l'atteinte de certaines bornes comme l'EDSS 3 et l'EDSS 6. De plus, les données utilisées concernant la mesure du handicap résultent d'un examen clinique réalisé par le neurologue conduisant à l'attribution du score EDSS. Une seule étude basée sur l'administration de questionnaires réalisée par Lebrun-Frenay et al. en 2016 sur un échantillon de 491 patients SEP recrutés majoritairement via des associations de patients a pu être retrouvée (118). Dans cette dernière, 14,5% des patients déclaraient avoir acheté une aide à la marche et 8,4% un VHP. Ces deux chiffres sont du même ordre de grandeur que ceux avancés dans notre travail en considérant les possessions d'aides techniques des patients à l'inclusion (10,4% ayant une aide à la marche et 8,2% un VHP) ou aux dernières nouvelles (respectivement 21,1% et 10,8%) par exemple. Les fréquences plus élevées de recours aux équipements observées dans l'étude de Lebrun-Frenay et al. tiennent sans doute au recrutement de patients volontaires, comparé à notre échantillon plus représentatif de l'ensemble des patients.

V.3.2. Forces et limites

Ce travail exploratoire et innovant a eu pour but d'exploiter au maximum les données à disposition dans le SNDS afin de mesurer indirectement une dimension essentielle de l'histoire naturelle de la SEP qu'est la progression du handicap. Bien que préliminaires, nos résultats sont cohérents avec les connaissances cliniques. Les principales limites tiennent à l'incapacité de déterminer avec précision l'évolution de niveau faible de handicap, et à l'imprécision temporelle de notre mesure.

Ainsi, les accidents traumatiques non liés à la SEP peuvent conduire à la location de matériels techniques comme un VHP par exemple. C'est pourquoi seules les locations d'une durée supérieure ou égale à 4 mois ont été considérées. Cette durée a été choisie en supposant d'une manière raisonnable suite à des discussions avec des cliniciens qu'un accident traumatique tel qu'une fracture d'un membre inférieur ne conduirait qu'en de très rares cas à une immobilisation de ce membre pour une durée plus longue. Aucune restriction n'a été effectuée sur les achats d'une aide technique. En effet, il a été considéré qu'un achat représentait un besoin sur le long terme d'un patient et donc ne pouvait pas être lié à un accident traumatique. De plus, la fenêtre de battement de 30 jours autorisée entre deux locations ou la fin d'une location et l'achat d'un même équipement permet de limiter les baisses de score EDSS dues à une hospitalisation ou à un délai du au changement pour un équipement plus récent. Elle permet également de prendre en compte les possibles erreurs de codage pouvant être présentes dans la base concernant les durées de location ainsi que les dates de début et de fin d'achat/location. En outre, le travail effectué sur les données d'entretien a permis de retrouver un équipement pour 3 978 patients. Les équipements identifiés avec les codes d'entretien (VHP manuel et électrique, lit médicalisé, matériel anti-escarres) correspondant à des scores EDSS « estimés » élevés, la prise en compte de ces derniers renforce la validité de l'algorithme.

Par ailleurs, la récupération des dates manquantes de début et de fin de location permet de mieux délimiter dans le temps chacun des équipements et ainsi de mieux identifier les chevauchements. Le score EDSS « estimé » supérieur ou égal à 8,5 a été défini sur la base de l'utilisation simultanée de deux équipements, à savoir un lit médicalisé et soit un matériel anti-escarres de classe II ou III soit un soulève-malade. Cette simultanéité dans les prescriptions a été choisie afin de ne pas surestimer le score EDSS d'un patient en ne considérant que l'achat ou la location d'un lit médicalisé qui pourrait avoir lieu bien avant l'atteinte de l'EDSS 8,5. Le soulève-malade caractérisant l'incapacité du patient à effectuer seul les transferts vers et hors de son lit, son ajout dans l'algorithme est apparu comme pertinent pour le niveau $\geq 8,5$ de l'échelle EDSS (Annexe A). Un raisonnement analogue pour le matériel anti-escarres de classes II et III a été effectué, ce dernier étant conseillé pour les patients restant alités plus de 15h par jour ou présentant un risque d'escarre moyen à élevé du fait de leur immobilisation.

Néanmoins, l'algorithme développé présente certaines limites dues aux données à disposition mais également au caractère exploratoire de son développement. En effet, le premier score EDSS « estimé » était le score de 5,5-6, caractérisant le besoin d'une aide unilatérale à la marche, le plus souvent une canne. Cependant, la canne est un équipement qu'un patient peut être susceptible d'acheter seul sans une prescription ou qu'il peut récupérer auprès d'une autre personne de son entourage. Ce constat, apparu suite à des discussions avec des spécialistes MPR, en charge de la prescription des aides techniques, peut alors conduire à une sous-estimation du score EDSS pour certains patients, dont la possession d'une canne est impossible à tracer dans les bases. Toutefois, cette sous-estimation est quasi-inexistante pour les autres équipements considérés, car étant plus difficiles à se procurer sans ordonnance. A l'opposé, aucune aide technique à disposition dans la base ne permet de discriminer un score EDSS « estimé » inférieur à 5,5-6, puisque le premier équipement (la canne pour rappel) correspond à un score EDSS 6. Par ailleurs, les forfaits de livraison des différentes aides techniques n'ont pas été considérés dans les codes recherchés, car il a été supposé que chaque forfait de livraison était accompagné d'une location ou d'un achat du matériel considéré.

Une baisse du score EDSS « estimé » a été constatée chez 5% des patients. Comme évoqué précédemment, cette baisse peut être liée à une première mesure à l'inclusion lors d'une poussée, mais également d'un arrêt des locations suite à une hospitalisation ou à un placement en EHPAD. Pour limiter ce biais, plusieurs hypothèses basées sur l'arrêt de locations d'équipements importants (déambulateur, VHP, lit médicalisé) après une période sans, avaient été développées dans l'EGB mais non implémentées dans le SNDS. Toutefois, ces hypothèses nécessitent une expertise clinique pour validation avant leur implémentation sur la base globale.

Enfin, ce travail constitue seulement une première étape qui appelle une étape de validation sur la base de données cliniques. En effet, une telle étape nécessite le chaînage des données du SNDS avec une base de données cliniques. Cette dernière devra alors disposer d'informations nécessaires sur les patients et leur consommation de soins pour permettre une bonne jointure avec le SNDS, mais aussi des données de mesure du score EDSS validées par un neurologue. En France, deux bases de données de ce type ont pu être identifiées : le LORSEP en Lorraine et la cohorte EDMUS de l'OFSEP. Un chaînage avec le SNDS est en cours pour la première et est planifié pour la seconde. Une fois cette étape réalisée, l'algorithme pourra être testé, validé et amélioré sur l'une de ces bases, voire les deux pour éviter un biais lié à la représentativité des bases.

Cette mesure du niveau de handicap moteur dans le SNDS constitue donc un travail exploratoire nécessitant une validation et plusieurs améliorations, non réalisées dans le cadre de cette thèse faute d'accès aux données. Néanmoins, cet axe de travail nous a permis d'estimer le niveau de handicap moteur des patients SEP, servant de paramètre d'ajustement pour l'axe s'intéressant aux parcours de

soins. L’algorithme, tel qu’il a été défini, est suffisant, car il n’a pas pour vocation de permettre de distinguer des variations de 0,5 point du score EDSS.

VI. Parcours de soins

VI.1. Méthodes

VI.1.1. Recours aux soins

Le parcours de soins d'un patient étant défini comme l'enchaînement des différents actes de soins, le recours aux soins de chaque patient a été calculé. Les actes considérés ont été :

- les consultations avec des spécialistes libéraux ou exerçant dans un établissement de santé privé,
- les consultations externes réalisées dans les établissements de santé publics et les ESPIC,
- ainsi que les séjours hospitaliers dans le domaine du MCO et du SSR.

N'ayant pas besoin d'une précision des données au jour près, et l'analyse statistique ne permettant pas alors de traiter une telle volumétrie de données, les différents éléments constitutifs du parcours de soins ont été agrégés au niveau mensuel. L'ensemble des indicateurs de recours aux soins n'ont pas été tous calculés dans les différentes populations d'étude en fonction des données et du temps à disposition pour réaliser les études.

Les données du SNDS comportent des informations sur les remboursements effectués par l'Assurance Maladie pour la part obligatoire mais également des informations issus de certaines complémentaires santé (part de remboursement pour l'Alsace-Moselle, etc.). Une prestation pouvant mener à un remboursement de la part obligatoire et à un second de la part complémentaire, des doublons peuvent apparaître. Par conséquent, nous ne nous sommes concentrés que sur les remboursements et les régularisations inhérentes à la part obligatoire de l'Assurance Maladie, c'est-à-dire les lignes de prestations pour lesquelles *BSE_REM_SGN* était égale à 1 (remboursement) ou 2 (régularisation).

VI.1.1.1. Consultations avec un spécialiste libéral ou dans un établissement de santé privé

Deux formes de consultations existent : les consultations au cabinet du spécialiste, dans le cas où le patient se déplace voir ce dernier, et les consultations au domicile du patient, le médecin se déplaçant. Dans la NGAP, les premières sont codées avec le terme « consultations », alors que les deuxièmes correspondent à des « visites ». Dans cette thèse, les deux types de consultations ont été analysés conjointement, d'où l'utilisation du terme générique « consultations » par la suite, conduisant à la construction d'un indicateur mensuel unique de recours aux soins

Le nombre mensuel de consultations a été calculé à partir de la variable *PRS_ACT_NBR*, quantifiant le nombre d'actes effectués lors de la prestation comme préconisé par la CNAMTS. Pour ce faire, les codes de nature de prestation (*PRS_NAT_REF*) permettant de repérer les consultations ont été identifiés dans la table des valeurs correspondante mise à disposition par la CNAMTS (Table 15) (173). Ces codes ont ensuite été utilisés en tant que filtre sur la table des prestations *ER_PRS_F*. Les majorations n'ont pas été prises en compte car étant reliées à un acte de base correspondant à une consultation, elles auraient entraîné un dédoublement des consultations.

Table 15-Codes de nature de prestation considérés pour l'identification des consultations

Code *PRS_NAT_REF*	**Code NGAP**	**Libellé**
1111	C	Consultation au cabinet par le médecin généraliste, le chirurgien-dentiste omnipraticien ou la sage-femme
1112	CS	Consultation au cabinet par le médecin spécialiste qualifié, le médecin spécialiste qualifié en médecine générale ou le chirurgien-dentiste spécialiste qualifié
1113	CNP	Consultation au cabinet par le médecin neuropsychiatre qualifié, psychiatre qualifié ou neurologue qualifié
1114	CSC	Consultation spécifique au cabinet par un médecin spécialiste en pathologie cardiovasculaire ou en cardiologie et médecine des affections vasculaires
1115		Consultation bilan
1117	C2	Consultation des spécialistes cotée C2
1118	C2,5	Consultation des psychiatres cotée C2,5
1211	V	Visite au domicile du malade par le médecin généraliste, le chirurgien-dentiste omnipraticien ou la sage-femme
1212	VS	Visite au domicile du malade par le médecin spécialiste qualifié, le médecin spécialiste qualifié en médecine générale ou le chirurgien-dentiste spécialiste qualifié
1213	VNP	Visite au domicile du malade par le médecin neuropsychiatre qualifié, psychiatre qualifié ou neurologue qualifié
1214	VL	Visite longue et complexe réalisée au domicile du patient atteint de maladie neurodégénérative par le médecin traitant
1221	VU	Visite à domicile urgente
1222	VU/MU	Visite à domicile urgente avec majoration d'urgence

Un second filtre a ensuite été appliqué sur le praticien exécutant en fonction de la spécialité considérée. Les spécialités choisies a priori étaient celles étant le plus susceptibles d'intervenir dans la prise en charge de la SEP, à savoir le médecin généraliste, le neurologue, le médecin MPR et l'ophtalmologue pour le suivi du PGR (Table 16) (144). Une analyse de la base a montré que certaines consultations réservées à certaines spécialités, par exemple la consultation spécifique cardiologie, pouvaient être utilisées par d'autres spécialités que le cardiologue. Par conséquent, l'ensemble des consultations ont été assignées à la spécialité, quel que soit le code NGAP. Si une valeur mensuelle du nombre de consultations était négative due aux régularisations de remboursement, elle était alors forcée à 0.

Table 16-Codes des spécialités considérées pour la consommation de soins

Code de la spécialité	Libellé
1, 22, 23	Médecin généraliste
15	Ophtalmologue
31	Médecin MPR
32	Neurologue

VI.1.1.2. Consultations externes

Un travail similaire a été effectué pour calculer un indicateur du recours aux soins via les consultations externes réalisées dans les établissements de santé publics et les ESPIC. La base utilisée a été celle relatives aux facturations des ACE disponible dans le PMSI MCO. Les consultations ont été extraites au moyen d'un filtre portant sur les codes NGAP des actes réalisés (variable *ACT_COD*), à savoir C, CS, CNP et CSC (146). Les spécialités considérées ont été les mêmes que celles repérées dans le cadre des consultations dans le DCIR (Table 16).

VI.1.1.3. Ensemble des consultations

Afin de décrire au mieux la consommation de soins des différents patients, des paramètres de consommation globale ont été définis. Ainsi, il a été construit une variable quantifiant la somme mensuelle de l'ensemble des consultations réalisées en ville, en établissements de santé privés, publics et ESPIC toutes spécialités confondues (l'ensemble des spécialités disponibles dans le SNDS a été considéré). Cette variable a été utilisée dans un but descriptif et n'a pas servi à la construction du parcours de soins d'un patient.

VI.1.1.4. Actes infirmiers et de kinésithérapie

Contrairement aux consultations avec des professionnels de santé médicaux, les soins paramédicaux avec des infirmiers et des kinésithérapeutes ne sont pas des soins ponctuels et ne sont donc pas

renseignés de la même manière dans le SNDS. Ainsi, ces actes ayant des codes de nature de prestation dédiés (Table 17) sont le plus souvent effectués en série, c'est-à-dire en plusieurs séances successives, et peuvent donc n'être rentrés qu'une fois l'ensemble des actes effectués. Dans ce cas, la date de début des soins *EXE_SOI_DTD* est différente de la date de fin des soins *EXE_SOI_DTF*, et le nombre d'actes effectués durant ce laps de temps est spécifié dans *PRS_ACT_QTE*. Cependant, il existe une majorité de prestations ponctuelles où alors les deux dates coïncident, où *PRS_ACT_QTE* quantifie alors le nombre d'actes effectués durant la prestation.

Afin de conserver le niveau mensuel, les prestations en séries sur plusieurs mois ont été séparées en prestations mensuelles. Une nouvelle variable mensuelle correspondant au nombre d'actes effectués sur la période totale divisé par le nombre de mois sur lesquels la prestation s'étale a été créée et assignée au niveau mensuel, en supposant que le nombre d'actes était lissé sur l'ensemble de la période.

Table 17-Codes de nature de prestation considérés pour la sélection des actes infirmiers et de kinésithérapie

Code *PRS_NAT_REF*	**Code NGAP**	**Libellé**
3111	AMI	Acte médico-infirmier
3112	AIS	Acte infirmier de soins
3121	AMC	Actes pratiqués par le masseur-kinésithérapeute dans une structure de soins ou un établissement, y compris lorsque le malade y a élu domicile
3122	AMK	Actes pratiqués par le masseur-kinésithérapeute au cabinet ou au domicile du malade, à l'exception des actes effectués dans un cabinet installé au sein d'un établissement d'hospitalisation privé au profit d'un malade hospitalisé

VI.1.1.5. Séjours hospitaliers

Deux paramètres ont été calculés pour quantifier le passage des personnes ayant une SEP dans les établissements de santé publics et privés pour des séjours en MCO : le nombre de séjours et le nombre de jours d'hospitalisation. Un séjour hospitalier pouvant durer plus d'un mois, et l'unité de temps choisie étant le mois, les séjours s'étalant sur plusieurs mois ont été séparés en séjours mensuels (ce découpage n'a pas inféré dans le calcul du nombre global de séjours hospitaliers pour un patient). De plus, le nombre mensuel de jours d'hospitalisation a été majoré par le nombre de jours par mois. Une

distinction a été effectuée entre les séjours liés à la SEP (DP ou DR égal à *G35*) et ceux n'étant pas liés à la maladie.

Concernant le PMSI SSR, une indicatrice permettant de savoir si un patient avait eu au moins un séjour en soins de suite et de réadaptation sur la période d'étude 2010-2015 a été construite.

VI.1.2. L'analyse de séquences

VI.1.2.1. Contexte

Afin de décrire précisément les parcours de soins et de les comparer entre plusieurs groupes, une méthodologie originale en santé a été utilisée : l'analyse de séquences. Cette dernière correspond à un ensemble de techniques développées dans le domaine des sciences sociales afin d'étudier de manière exploratoire et descriptive des données longitudinales, principalement concernant le parcours de vie. Introduites à partir des années 1980 (174), ces méthodes sont à présent largement répandues dans le domaine des *lifecourse data* et ont bénéficié de nombreuses avancées méthodologiques au cours du temps permettant ainsi une plus large utilisation (175–181). Elles sont principalement utilisées pour l'étude des trajectoires professionnelles (182–184), de la situation maritale (180,185) et pour l'étude des comportements dans la population, comme l'impact d'un diagnostic sur les parcours de vie par exemple (186). A l'inverse, peu de littérature fait état de son utilisation dans le domaine de la santé (124,128). Au démarrage de ce travail de thèse, une seule étude réalisée par une enseignante-chercheure de l'EHESP sur les consommations de soins des femmes enceintes était référencée à notre connaissance (124). Cependant, cette publication ne détaillait pas la méthode employée, ainsi que ses limites et ses avantages.

Suite aux récents développements, l'analyse de séquences originellement utilisée pour l'étude d'une seule variable, peut à présent être utilisée pour l'étude synchronisée de plusieurs variables à travers une analyse multidimensionnelle (multichannel sequence analysis) (178,179,187). Elle peut également constituer une première étape et servir de base à une analyse de survie ou à une régression logistique (188,189). Enfin, à l'origine simplement descriptive, elle permet à présent la recherche de facteurs explicatifs au moyen d'une analyse discriminante semblable à une ANOVA (Analyse de la variance), appelée *discrepancy analysis* (175,180), et la réalisation d'arbres de régression (*regression trees*) (190,191), synthétisant les résultats. D'un point de vue purement esthétique, l'analyse de séquences présente des atouts non négligeables dans la présentation de résultats de manière graphique (192).

L'ensemble des résultats présentés ultérieurement sur l'analyse de séquences ont été obtenus à l'aide des librairies *TraMineR* et *WeightedCluster* disponibles sur le logiciel R (153) et développés respectivement par Gabadinho et al (169) et Studer (190). Par ailleurs, l'application de l'analyse de

séquences dans le cadre des parcours de soins a fait l'objet d'une publication intitulée « Use of state sequence analysis for care pathway analysis: the example of multiple sclerosis » dans *Statistical Methods in Medical Research* en 2018 (Annexe F).

VI.1.2.2. Termes spécifiques à l'analyse de séquences

Les méthodes issues du domaine de l'analyse de séquences ont recours à un vocabulaire spécifique. Afin de permettre une meilleure compréhension, cette première section est dédiée à la définition de certains termes employés ci-après. Comme évoqué précédemment, l'analyse de séquences se concentre sur l'évolution d'une variable au cours d'une période de temps donnée. Cette variable peut être quantitative ou qualitative. La valeur prise par la variable à un instant t donné est appelée « état ». Un état peut donc correspondre à un nombre ou à une situation (marié, célibataire par exemple). L'ensemble des états uniques, et donc des valeurs possibles de la variable, est appelé « alphabet ». Une « séquence » est alors définie comme étant la succession d'états au cours d'une période donnée.

Un parcours de soins étant, quant à lui, défini comme la succession des actes et consultations d'un patient auprès des professionnels de santé, il peut être assimilé à une séquence. Dans ce cas, un état peut représenter une consultation, un acte médical et l'alphabet est alors formé de l'ensemble des consultations et actes possibles. Dans ce travail, à cause de la complexité et de la quantité de données à disposition, nous avons fait le choix de définir les états comme étant un nombre de recours à un spécialiste ou de jours dans un établissement de santé défini.

VI.1.2.3. Construction du parcours de soins

La première étape de l'analyse de séquences est la construction du parcours de soins et donc l'identification des états constitutifs de la séquence de soins à travers plusieurs questions préliminaires (Figure 26).

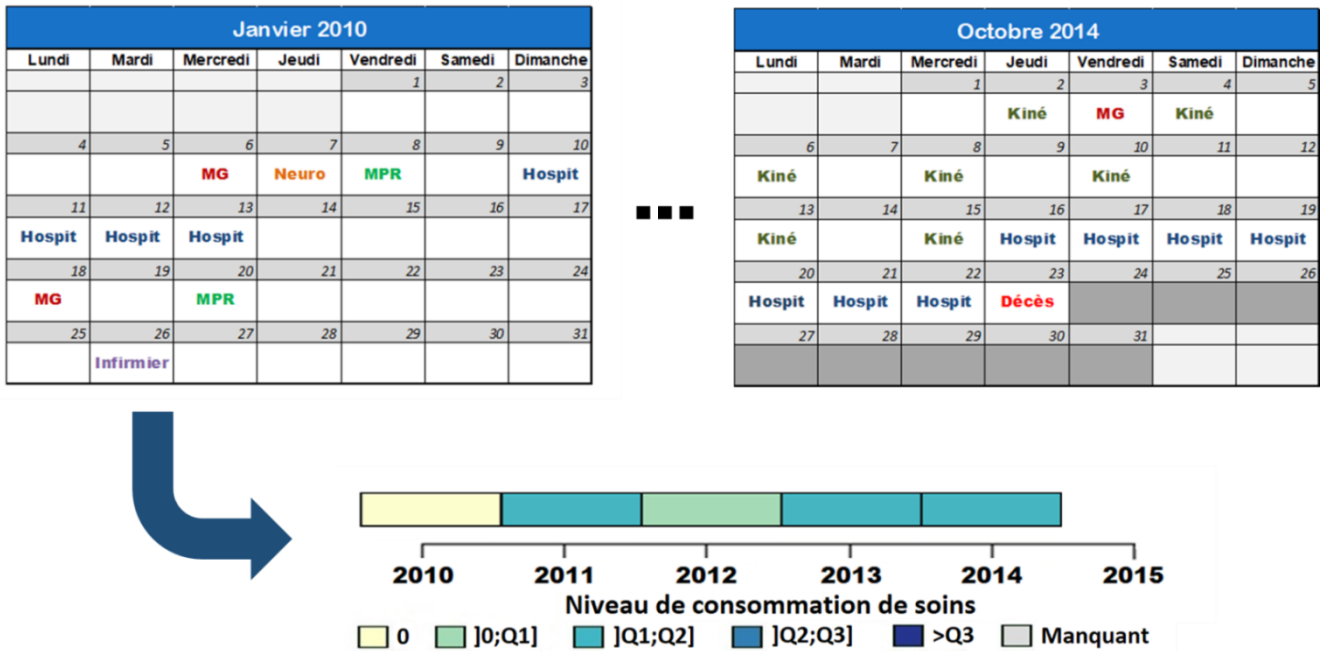


Figure 26-Représentation schématique de la construction d'une séquence correspondant au parcours de soins d'un individu pour l'analyse de séquences

Choix de l'unité temporelle

La première concerne la période d'étude et l'unité temporelle de la séquence. En effet, la période permet de définir le début et la fin de la séquence pour chaque personne et a donc une influence sur la longueur finale de la séquence. L'unité temporelle, quant à elle, conditionne la valeur des états, notamment dans notre cas d'étude du recours aux soins. L'analyse de séquences laissant une grande flexibilité à l'utilisateur, l'unité temporelle peut très bien aller de l'heure, comme dans l'étude des emplois du temps (193), à une durée plus importante comme le mois (184), le trimestre (124) ou encore l'année (194). La combinaison de la période d'étude et de l'unité temporelle déterminant le nombre d'états, elle permet de déterminer les longueurs minimale et maximale des séquences. Une revue des pratiques effectuée sur des séquences ayant pour spécificité de longs épisodes dans un même état, conduite par Dlouhy et al. (134), a recommandé d'utiliser des séquences avec au minimum 25 états. Elle a également indiqué que les résultats ne changent qu'à la marge lorsque la longueur des séquences est gonflée artificiellement par la diminution de l'unité temporelle. Cependant, ces recommandations ne concernent pas des séquences marquées par une forte variabilité intra-séquence, comme cela peut être le cas dans les parcours de soins.

Gestion des valeurs manquantes

Par ailleurs, il est possible que des valeurs manquantes apparaissent dans les parcours de soins, que ce soit au début, au milieu ou à la fin de la séquence. Ces dernières peuvent être liées à la base de données elle-même mais aussi à la disponibilité de cette dernière. L'analyse de séquences supporte assez bien les valeurs manquantes si elles ne dépassent pas 30% de la longueur totale de la séquence (134). Ainsi, plusieurs méthodes d'imputation ont été développées par Halpin pour des valeurs manquantes en début ou en fin de séquence (195), et celles créant des trous dans le suivi (196). Une autre solution pour le traitement des valeurs manquantes débutant les séquences est la suppression de ces dernières, ce qui revient à aligner à gauche les séquences. Si cela n'a pas de sens vis-à-vis de la question de recherche, il est alors conseillé de séparer les séquences en plusieurs groupes selon leur longueur (séquences courtes *vs* séquences longues, ou séparer les séquences selon l'année d'inclusion des patients par exemple). Il est toutefois possible de conserver ces valeurs manquantes initiales et de réaliser l'ensemble de la procédure d'analyse. Cependant, les séquences commençant par l'état « valeur manquante » seront similaires ce qui entraînera leur rapprochement dans un même groupe lors de la classification. Le groupe alors obtenu correspondra à l'ensemble des patients ayant eu une entrée retardée dans l'étude et non à une trajectoire-type, ce qui en fera un groupe sans intérêt clinique au moment de l'interprétation.

Définition de l'alphabet

Une fois qu'une réponse a été apportée à ces premières questions, arrive la définition de l'alphabet, c'est-à-dire la liste finie des différents états constitutifs des séquences. Il s'agit d'un élément-clé de la construction du parcours puisqu'il contribue à la complexité ou non des séquences. Ainsi, plus la longueur de l'alphabet sera importante, plus le nombre d'états sera élevé et donc plus l'interprétation finale des groupes obtenus pourra être complexe (197). A l'inverse, un alphabet court entraînera des séquences plus simples et donc des groupes finaux pouvant être triviaux. Cependant, aucune recommandation n'existe à notre connaissance sur le nombre optimal d'états, qui est laissé à la discrétion de l'utilisateur. En outre, un des avantages de l'analyse de séquences est que tant des variables quantitatives que qualitatives peuvent être considérées dans l'alphabet. Il faut néanmoins qu'elles soient toutes de la même nature afin de garantir une homogénéité de l'alphabet, c'est-à-dire tous les états sont quantitatifs ou qualitatifs.

Gestion d'un arrêt prématuré du parcours

Enfin, la dernière phase de cette première étape de construction est la gestion de la fin des séquences. En effet, dans le cas où tous les patients ne sont pas suivis sur la même durée, certains peuvent avoir des séquences plus courtes du fait de leur décès ou d'une sortie d'étude. Trois choix peuvent alors être envisagés : l'arrêt de la séquence, la création d'un état spécifique ou l'utilisation de valeurs manquantes. Dans le premier cas, le patient verra sa séquence tronquée à droite, c'est-à-dire que cette dernière sera plus courte que les autres, créant par conséquent des différences de longueurs entre séquences. La création d'un état spécifique, tel que « Décès » ou « Perdu de vue », peut être opportun s'il a un sens au vu de la question de recherche. Cependant, un risque de plus grande similitude entre les séquences disposant de cet état est fortement probable, à cause de la répétition de ce dernier à la fin de la séquence. Il en est de même avec l'utilisation de l'état « Valeur manquante ». Cela constitue alors une situation analogue à celle évoquée avec les valeurs manquantes en début de séquence et la création d'un groupe dédié.

VI.1.2.4. Mesure de la dissimilarité entre les parcours

Une fois les parcours créés, la principale étape de l'analyse de séquences est la comparaison de ces derniers deux à deux, nécessitant la définition d'une mesure de dissimilarité. Cette dernière selon ses caractéristiques de construction peut alors simplement être le reflet d'une dissimilarité ou une distance si elle en remplit les axiomes (198). De nombreuses mesures de similarité ou dissimilarité utilisées à l'origine dans différents domaines d'application et adaptées à l'analyse de séquences sont décrites dans la littérature (135) : les distances de Levenshtein I et II (199) et l'optimal matching (174) issues de la bio-informatique, la distance du khi-deux (200) plus commune en statistique ou encore des

mesures développées en sciences sociales comme le nombre de sous-séquences communes (*Number of Matching Subsequences (NMS))* ou la plus longue sous-séquence commune (201,202).

Concept de l'optimal matching

Dans ce travail, nous nous sommes concentrés sur l'optimal matching étant la méthode la plus communément utilisée en analyse de séquences (135,203) et celle présentant la plus grande flexibilité permettant une bonne adaptabilité à la question de recherche. A l'origine développée en biologie pour la comparaison des séquences ADN au cours des années 1980 (204), l'optimal matching, ou « méthode d'appariement optimal » (205), a été introduit au cours des années 1990 dans les sciences sociales par Andrew Abbott (174,206) puis son utilisation s'est largement répandue (203). Il s'agit d'une mesure de dissimilarité d'édition, c'est-à-dire qu'elle repose sur l'assignation d'une valeur de dissimilarité dépendante du nombre d'opérations nécessaires pour rendre deux séquences strictement identiques (135). Ainsi, si on considère deux séquences formées de trois états chacune ne différant que selon leur dernier état. Le fait de changer le dernier état de la première séquence en celui de la seconde séquence suffirait à les rendre parfaitement identiques. Dans ce cas, les deux séquences auraient une faible valeur de dissimilarité. Pour chaque opération additionnelle, la mesure de dissimilarité augmenterait alors.

Opérations élémentaires

L'optimal matching utilise trois opérations élémentaires : l'insertion, la délétion et la substitution (Figure 27). L'insertion correspond à l'ajout d'un état dans une séquence, alors qu'à l'inverse la délétion correspond à la suppression d'un état. Chacune de ces deux opérations peut avoir lieu n'importe où dans la séquence (au début, à la fin ou entre deux états au cœur de cette dernière). Comme l'insertion et la délétion repose toutes les deux sur le concept de distorsion du temps (l'ajout ou la suppression d'un état modifie la séquence temporelle réelle des évènements) (207), elles ont été rassemblées dans le même concept d'indel, formé de la contraction des deux noms. A l'opposé, la substitution respecte la dimension temporelle de la séquence mais modifie la séquence des états, en remplaçant un état par un autre état différent au même instant t (207).

Mesure de dissimilarité

Ces trois opérations servent alors de base au calcul de la mesure de dissimilarité entre deux séquences. Ainsi à chaque opération (indel et substitution) est assigné un coût spécifique pouvant être constant ou pouvant varier en fonction des états. La dissimilarité entre deux séquences correspond alors à la somme des coûts de chacune des opérations élémentaires nécessaires à la transformation d'une séquence en l'autre. Comme plusieurs combinaisons d'opérations peuvent permettre la transformation, l'algorithme d'optimal matching conserve uniquement la dissimilarité minimale. Cette mesure dépend donc fortement à la fois des coûts assignés et du nombre d'opérations à effectuer (135). Afin de tester

l'ensemble de ces combinaisons et ainsi calculer la dissimilarité entre les séquences, l'algorithme de Needleman-Wunsch (208) est implémenté dans la majorité des logiciels statistiques usuels.

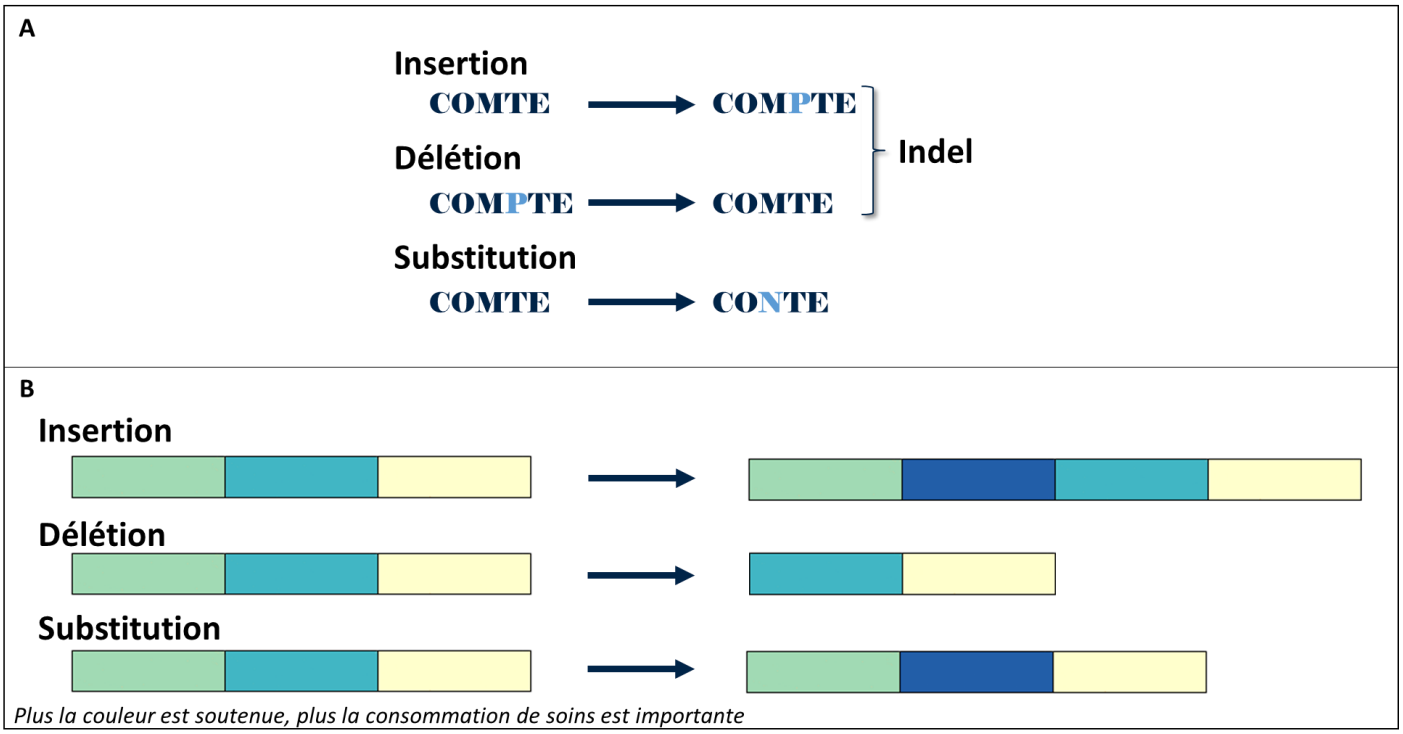


Figure 27-Opérations élémentaires de l'analyse de séquences présentées avec (A) un exemple avec des mots et (B) avec des parcours de soins assimilés à des séquences

Définition des coûts

L'étape majeure de l'optimal matching est la définition des coûts (209), puisqu'étant à l'origine de la mesure de dissimilarité, ils ont un impact direct dans les résultats de la classification (207).

Coûts de substitution

Les premiers à être définis sont les coûts de substitution, qui peuvent être définis à partir de données théoriques ou de manière empirique directement dans la base. Ces coûts prennent la forme d'une matrice symétrique contenant une valeur pour chaque substitution d'un état par un autre et des zéros sur la diagonale, la substitution d'un état par lui-même ne changeant pas une séquence. Dans le cadre de l'utilisation de coûts théoriques, l'ensemble des coûts de la matrice doivent respecter l'inégalité triangulaire afin de garantir la validité de la dissimilarité calculée (135). En effet, en l'absence de cette hypothèse, la mesure renvoyée par l'algorithme de Needleman-Wunsch (208) pourrait être faussée ainsi que la classification (198). Si aucune donnée théorique ne permet de définir de tels coûts, les coûts de substitution peuvent être déterminés empiriquement en se basant sur les taux de transition

observés dans l'ensemble des séquences (169,197). Dans ce cas, le coût de la substitution de l'état i, noté s_i, par l'état j, noté s_j, et inversement, $C_{i\to j/j\to i}$ est défini par

$$C_{i\to j/j\to i} = 2 - p(s_i|s_j) - p(s_j|s_i)$$

Où $p(s_i|s_j)$ et $p(s_j|s_i)$ sont respectivement les probabilités d'observer la transition s_j vers s_i et s_i vers s_j dans la base des séquences (169). Dans le cas de présence d'un état codant pour une valeur manquante, le coût est communément fixé à 2 quand les coûts de substitution basés sur les transitions sont utilisés. De même que pour des coûts théoriques, ces derniers doivent respecter l'inégalité triangulaire afin de permettre l'utilisation de l'algorithme de Needleman-Wunsch.

Coût d'un indel

Dans un second temps, le coût d'un indel, qui peut être fixe ou lié à chacun des états de l'alphabet (135), doit être déterminé. C'est le choix de ce coût réalisé par l'utilisateur qui rend l'optimal matching flexible à chaque question de recherche. En effet, le coût de l'indel conditionne la signification de la mesure de dissimilarité obtenue par optimal matching (Figure 28). Ainsi, si on se place dans le cas extrême d'un coût très faible pour les indels comparé à celui des substitutions, l'utilisation des indels sera alors fortement favorisée dans la transformation des séquences. Les indels entraînant une distorsion du temps, deux séquences seront similaires si elles proposent le même enchaînement d'états. La mesure de dissimilarité ainsi estimée se rapprochera de la distance de Levenshtein II, correspondant à la notion de plus longue sous-séquence commune (207). En revanche, dans le cas extrême inverse, où les indels auraient un coût très élevé comparé aux coûts de substitution, ces dernières seront favorisées au détriment des indels. Les substitutions entraînant une distorsion dans la séquence des états, deux séquences seront d'autant plus proches qu'elles disposent d'un grand nombre d'états identiques aux mêmes instants. La mesure de dissimilarité sera similaire alors à la distance de Hamming, correspondant au nombre d'états identiques (210). Par conséquent en choisissant un coût d'indel plus ou moins élevé en comparaison des coûts de substitution, l'utilisateur peut choisir la mesure de dissimilarité la plus adaptée et ainsi construire la matrice de dissimilarités entre les séquences.

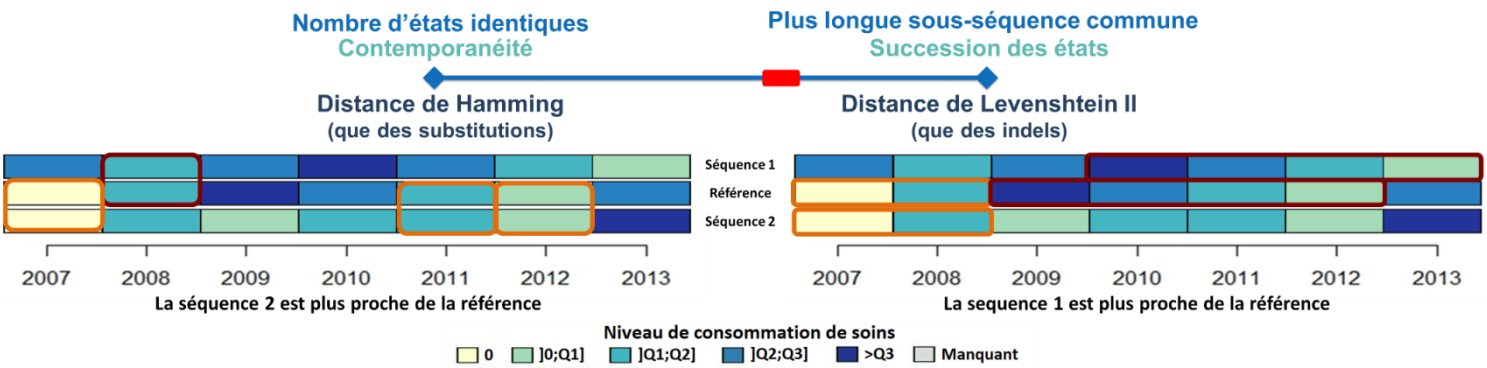


Figure 28-Impact de la mesure de dissimilarité sur la comparaison des séquences (adapté de (207))

Normalisation de la dissimilarité

Comme évoqué précédemment des différences de longueurs dans les séquences peuvent exister dans le cas de séquences tronquées à droite. Cette différence peut être intéressante à prendre en compte, car influant potentiellement sur la valeur de la mesure de dissimilarité (169). Ainsi, comme expliqué par Gabadinho et al., une dissimilarité de 10 entre deux séquences de 100 états chacune est en réalité moins importante qu'une dissimilarité de 10 entre deux séquences de 5 états chacune (169). Afin de prendre en compte cet aspect et avoir une valeur de dissimilarité ne dépendant pas de la longueur des séquences, cette dernière peut être normalisée. Plusieurs formules de normalisation de la dissimilarité entre deux séquences peuvent être utilisées comme celle définie par Abbott consistant à diviser par la longueur de la plus longue séquence des deux ou par la somme des deux longueurs (169,198). Ces deux normalisations ne garantissent pas la conservation de l'inégalité triangulaire contrairement à celle proposée par Yujian and Bo garantissant l'inégalité après normalisation (211). Cependant, de récents développements en sciences sociales ont critiqué l'utilisation de la normalisation en analyse de séquences (212).

VI.1.2.5. Classification des parcours

La matrice de dissimilarités ainsi construite sert ensuite de base pour la classification des séquences afin de créer la typologie.

Choix de l'algorithme de classification

Deux méthodes principales peuvent être utilisées afin de réaliser le clustering : les méthodes de classification ascendante hiérarchique (CAH) et les algorithmes de k-medoids.

D'après la revue de Dlouhy et al., la CAH selon le critère de Ward est la plus utilisée en analyse de séquences (134). L'optimal matching résultant d'un processus de minimisation, il ne s'agit pas d'une distance euclidienne (198), mais l'inégalité triangulaire permet de garantir les axiomes de la distance. La CAH selon le critère de Ward ne peut donc théoriquement pas être utilisée avec l'optimal matching puisqu'elle repose sur la somme des écarts à la moyenne au carré non généralisable aux distances non euclidiennes. Néanmoins, les travaux de Batagelj (213) et Studer et al. (175) ont montré que la CAH selon le critère de Ward peut être vue comme une procédure visant à minimiser une pseudo-variance d'un point de vue plus général, ce qui tendrait à autoriser son utilisation avec une distance non-euclidienne. De plus, la classification repose sur des heuristiques, c'est-à-dire des méthodes d'apprentissage suboptimales qui ne permettent pas de trouver la meilleure partition à chaque fois, et non des méthodes absolues. Batagelj a ainsi montré que la méthode de Ward n'est pas une mauvaise heuristique, même lorsque la distance est non-euclidienne (213). En addition, Dlouhy et al. en utilisant des simulations ont également abondé en ce sens en concluant que la CAH selon le critère de Ward

permettait d'obtenir la meilleure partition qu'importe la longueur de la séquence en comparaison des autres méthodes adaptées aux distances non-euclidiennes (134).

L'algorithme de k-medoids préconisé dans l'analyse de séquences est l'algorithme PAM (Partitioning Around Medoids) (190). Les résultats obtenus avec cet algorithme sont moins bons qu'avec ceux obtenus avec une CAH selon le critère de Ward (134). Néanmoins, il peut être intéressant d'appliquer cet algorithme suite à une CAH afin d'obtenir potentiellement une meilleure qualité de la partition (190).

Nombre optimal de clusters

Le nombre optimal de clusters peut être choisi théoriquement ou de façon empirique. Dans le cas d'une CAH selon le critère de Ward, la chute de l'inertie, le critère communément utilisé, peut être utilisé pour déterminer le nombre de clusters. Cependant, cette chute pouvant être difficile à observer dans le cas de clustering sur des données de grandes dimensions, des critères de qualité de la partition peuvent être utilisés. Plusieurs critères existent dans la littérature et peuvent être utilisés (190) (Table 18). Quatre critères sont principalement utilisés en analyse de séquences : l'ASWw (weighted Average Silhouette Width) (214), le HC (Hubert's C) (215), le PBC (Point Biserial Correlation) (216,217) et le HG (Hubert's Gamma) (218). Ces critères conduisant dans la majorité du temps au même choix du nombre de clusters, ils sont utilisés en combinaison de deux ou trois critères.

Table 18-Critère de mesure de la qualité d'une partition (adapté de (190))

Critère de qualité	Abrev.	Etendue	Objectif	Interprétation
Point Biserial Correlation	PBC	[-1 ; 1]	Maximisation	Mesure de la capacité de la classification à reproduire les distances
Hubert's Gamma	HG	[-1 ; 1]	Maximisation	Mesure de la capacité de la classification à reproduire les distances (ordre de grandeur)
Hubert's C	HC	[0 ; 1]	Minimisation	Ecart entre la partition obtenue et la meilleure partition qu'il serait théoriquement possible d'obtenir avec ce nombre de groupes et ces distances
Weigthted Average Silhouette Width	ASWw	[-1 ; 1]	Maximisation	Cohérence des assignations. Une cohérence élevée indique des distances inter-groupes élevées et une forte homogénéité intragroupe.

Abrev. : Abréviation

VI.1.2.6. Représentation des séquences

Un des avantages de l'analyse de séquences est l'ensemble des représentations graphiques à disposition afin de décrire et communiquer les résultats de la classification. Les deux représentations les plus utilisées sont l'index plot, aussi appelé tapis, et le chronogramme (169) (Figure 29). L'index plot correspond à la superposition de chacune des séquences de la base, ce qui permet de voir d'un seul coup d'œil l'ensemble des parcours de soins. Dans ce diagramme, l'axe *x* représente le temps selon l'unité temporelle choisie et l'axe *y* les identifiants de chacune des séquences, qui n'est pas usuellement figuré (169). L'index plot permet de représenter l'aspect longitudinal des parcours, à l'opposé le chronogramme permet de synthétiser l'information, mais l'aspect longitudinal est alors perdu. En effet, comme dans l'index plot l'axe des abscisses du chronogramme représente le temps alors que sur l'axe des ordonnées est figuré le pourcentage cumulé de patients dans les différents états.

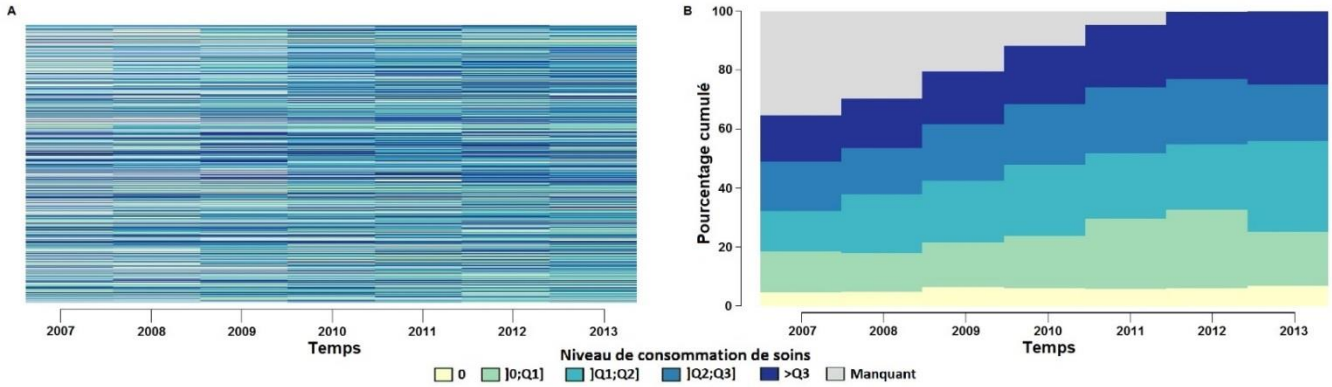


Figure 29-Représentation des séquences sous forme (A) d'un index-plot et (B) d'un chronogramme

VI.1.4. L'analyse multichannel sequence analysis (MCSA)

Un des principaux inconvénients de l'analyse de séquences unidimensionnelle comme présentée ci-dessus est l'impossibilité de prendre en compte plusieurs séquences distinctes pour un même patient, d'où la création d'une variable synthétisant la consommation de soins. Or, le recours aux soins d'un patient n'est pas uniforme auprès de l'ensemble des professionnels de santé et rend donc intéressant le fait d'étudier les différences de recours entre professionnels. Afin de résoudre ce problème, au cours des dernières années, le concept de *multichannel sequence analysis (MCSA)*, ou analyse séquentielle multidimensionnelle (178), a été développé par Pollock (179) et Gauthier et al. (178,187).

IV.1.4.1. Définition

La MCSA est une extension de l'analyse de séquences unidimensionnelle présentée précédemment. La principale différence entre les deux méthodes réside dans le nombre de séquences prises en compte par individu. Ainsi, dans l'analyse de séquences unidimensionnelle seule une séquence peut être

considérée pour un individu, ce qui en fait une analyse unidimensionnelle. A l'opposé, dans la MCSA plusieurs séquences et donc plusieurs dimensions (*channels* en anglais) peuvent être prises en compte simultanément pour un même individu, d'où le terme de multidimensionnel ou *multichannel* (187). Ainsi, en MCSA, chaque séquence traduit une dimension du parcours global d'un individu. En transposant cette définition au domaine des parcours de soins, on comprend alors qu'un parcours de soins se décompose en plusieurs recours à différents professionnels, comme le recours aux médecins généralistes et le recours aux neurologues. L'ensemble de ces recours pris en compte simultanément constitue alors le parcours de soins global d'un individu, d'où l'intérêt de la méthode.

La MCSA suit les mêmes étapes que l'analyse de séquence unidimensionnelle, c'est-à-dire la construction des séquences, le calcul de la dissimilarité entre ces dernières grâce à l'optimal matching et enfin leur classification. La dernière étape de clustering est identique à celle présentée précédemment et utilise les mêmes méthodes. Les deux premières étapes ne changeant pas fondamentalement sur le fond nécessitent tout de même quelques précisions.

IV.1.4.2. Construction des séquences

La principale différence entre MCSA et analyse de séquences unidimensionnelle réside dans le nombre de séquences construites pour un même individu. Chacune de ces séquences correspond à une dimension du parcours de soins (recours à un médecin généraliste, à un neurologue, séjours hospitaliers, …) et possède son propre alphabet d'états (Figure 30). Elles sont toutes synchronisées et bâties sur la même unité temporelle, c'est-à-dire qu'elles commencent et finissent au même instant, et possèdent le même nombre d'états pour un même individu. Ainsi, le 4ème état de la séquence « recours aux médecins généralistes » et le 4ème état de la séquence « recours aux neurologues » correspondent au même instant t pour un individu donné (187). Une fois chacune des dimensions séquencées, elles sont considérées simultanément pour le calcul de la dissimilarité entre deux individus en utilisant la méthode d'optimal matching.

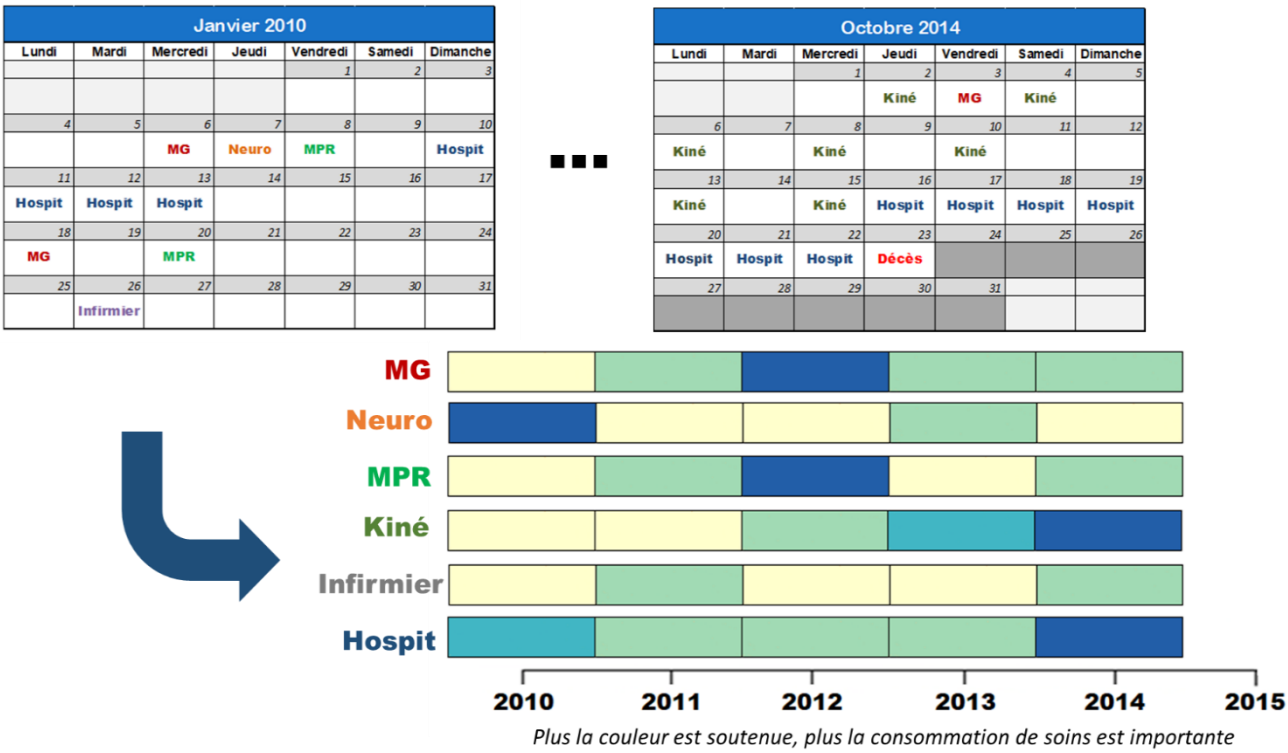


Figure 30-Représentation schématique de la construction d'une séquence correspondant au parcours de soins d'un individu pour la MCSA

IV.1.4.3. Calcul de la dissimilarité

Chaque dimension de la séquence globale d'un individu possédant son propre alphabet, les coûts de substitution et d'indels sont propres à chacune d'entre elles (178,179,187). Par conséquent, il y a autant de matrices de coûts de substitution et de valeurs d'indels qu'il y a de dimensions dans la séquence. La mesure de dissimilarité entre deux séquences est alors calculée en appliquant la méthode d'optimal matching décrite précédemment. La valeur de dissimilarité finale entres les deux séquences est alors calculée sur la combinaison des trajectoires en utilisant les coûts dérivés de chaque dimension (169,179). La matrice de dissimilarités ainsi obtenue peut être utilisée pour identifier la typologie finale grâce à une méthode de classification.

VI.1.5. Applications

VI.1.5.1. Application de l'analyse de séquences unidimensionnelle à l'EGB

La procédure d'analyse de séquences a été appliquée dans un premier temps sur la population de 1 000 patients identifiés dans l'EGB (population « *Recours aux soins* ») sur la période d'étude 2007-2013, soit sept ans. Une unité temporelle placée au jour est apparue comme étant inutile, puisque la longueur des séquences aurait été trop importante pour permettre d'obtenir des groupes homogènes, à cause de la forte hétérogénéité des parcours de soins. Le mois et le trimestre ont également été envisagés comme unité. Cependant, le recours à certains spécialistes étudiés étant peu fréquent, notamment les

neurologues, ces deux unités n'ont pas été retenues. Par ailleurs, les critères d'inclusion ont permis d'identifier à la fois des cas prévalents et des cas incidents de SEP. Ces patients n'étaient pas tous au même stade de la maladie à l'inclusion, et aucune variable à disposition ne permettait une identification claire de la durée depuis le diagnostic (la durée depuis la mise en ALD n'étant qu'un proxy de la durée de la maladie). Par conséquent, le choix s'est finalement porté dans un premier temps sur une unité de temps correspondant à une année calendaire afin de ne pas introduire un biais en utilisant comme unité temporelle l'âge du patient. Les parcours de soins étaient alors composés de sept variables annuelles synthétisant la consommation de soins mises bout à bout et constituant ainsi des séquences d'une longueur maximale de sept états.

Pour rappel, la période d'inclusion de la population d'étude s'est étalée de 2007 à 2012 et l'EGB est une cohorte dynamique. Par conséquent, plusieurs patients identifiés ne disposaient pas de données de consommation de soins pour la période précédant leur inclusion, créant ainsi des valeurs manquantes au début des parcours de soins et donc des séquences. L'unité de temps choisie, étant l'année calendaire, elle ne permettait pas un alignement à gauche des séquences. Afin d'éviter la création d'un cluster dédié à ces entrées retardées dans la cohorte, l'analyse s'est concentrée sur les patients identifiés en 2007, soit 648 patients.

Une variable caractérisant la consommation de soins individuelle a ensuite été créée. Elle correspondait à la somme du nombre annuel de consultations en libéral uniquement (les ACE étant indisponibles dans l'analyse portant sur l'EGB pour rappel) avec un des spécialistes considérés (médecin généraliste, neurologue, spécialiste MPR) et des hospitalisations liées à la SEP en MCO. De plus, une variable caractérisant le recours aux neurologues hospitaliers avait été calculée dans l'EGB en se basant sur la prescription de traitements de fond pour les patients ne disposant d'aucun recours à un neurologue libéral au cours de leur suivi. Ainsi, une consultation annuelle était imputée au démarrage du traitement et tous les ans à la date anniversaire si le traitement était toujours en cours ou lors d'un changement de traitement. Cette variable ainsi construite a été ajoutée à la variable globale de recours aux soins annuel. Cette dernière suivant une loi de Poisson avec large dispersion, elle a été catégorisée en cinq états suivant les quartiles (]0-Q1], [Q1-Q2], [Q2-Q3] et > Q3) plus un groupe supplémentaire caractérisant l'absence de consommation de soins (soit un recours annuel de 0 consultation ou séjour hospitalier lié à la SEP). Aucune valeur manquante n'était présente dans les séquences du fait de l'exhaustivité des données. Lorsqu'un patient décédait au cours du suivi, sa séquence était tronquée à droite. En effet, la création d'un état « décès » aurait conduit au rassemblement de ces patients au sein d'un même groupe.

Comme aucune donnée théorique n'était à disposition permettant de fixer les coûts de substitution *a priori*, ces derniers ont été évalués empiriquement en se basant sur les taux de transitions observés

dans la base. Dans le but de comparer les parcours de soins selon l'évolution du recours sur le long-terme et ne pas se concentrer uniquement sur des évènements discordants mais plutôt sur le parcours dans son ensemble, un coût unique pour les indels a été défini afin d'approcher la distance de Levenshtein II.

Une CAH selon le critère de Ward a enfin été réalisée pour créer la typologie (l'algorithme PAM n'améliorant pas la qualité de la partition). Le nombre optimal de groupes a été défini en utilisant la combinaison de deux critères de qualité de la partition : l'ASWw et le HC. Une fois la typologie formée, elle a été caractérisée à partir des données individuelles à disposition et des données de consommation de soins.

VI.1.5.2. Application de la MCSA à l'EGB

La MCSA a été appliquée sur la population des 1 000 patients identifiés dans l'EGB (population « *Recours aux soins* ») sur la période d'étude 2007-2013. Suite à l'application de l'analyse de séquences unidimensionnelle sur cette population, plusieurs améliorations ont été apportées dans la sélection de la population d'étude grâce au travail méthodologique et à l'expertise acquise sur la méthode. Ainsi, afin d'avoir des données de consommation de soins sur l'ensemble du suivi et ainsi limiter le biais lié à l'entrée retardée des individus dans la cohorte, seuls les patients SEP identifiés en janvier 2007 dans le régime général ont été gardés (n=543). En effet, seules les informations sur les bénéficiaires affiliés au régime général étaient disponibles avant 2010 dans l'EGB. La période d'inclusion a par ailleurs été restreinte au mois de janvier 2007 pour limiter la sous-estimation introduite par une identification au cours de l'année 2007 et l'indisponibilité des données de recours aux soins antérieures à cette date d'identification. Dans le but d'améliorer la résolution des parcours de soins due à la considération des recours aux professionnels indépendamment les uns des autres, l'unité temporelle choisie a été le semestre calendaire (la première unité de temps s'étend donc du 1er janvier 2007 au 30 juin 2007, puis du 1er juillet 2007 au 31 décembre 2007, etc.). Les parcours de soins étaient donc d'une longueur maximale de 14 états. Les séquences ont été tronquées à droite en cas de décès et aucun alignement à gauche n'a été performé.

Les dimensions du parcours de soins considérées ont été comme précédemment le recours aux médecins généralistes libéraux, aux neurologues (libéraux et hospitaliers imputés comme dans le cas de l'analyse de séquences unidimensionnelle sur l'EGB), et le nombre de séjours hospitaliers liés à la SEP en PMSI MCO. Ont été ajoutés comme dimensions : le recours aux spécialistes MPR libéraux, aux infirmiers et aux kinésithérapeutes. Concernant les recours aux médecins généralistes, aux neurologues et aux MPR, ces nombres correspondaient au nombre de consultations et visites effectuées. Quant au recours aux infirmiers et aux kinésithérapeutes, il était approximé dans cette partie au nombre de jours où un acte avait été effectué. Au total six dimensions ont donc constitué la

séquence d'un individu. Compte-tenu de la grande variabilité dans les valeurs de chacune des variables quantifiant le recours à un spécialiste, les états de chaque dimension ont été catégorisés selon la distribution avec l'ajout d'un état correspondant à l'absence de recours durant le semestre considéré.

Les coûts de substitution étaient basés sur les taux de transition observés au sein des données. Le coût des indels a été fixé à 1 pour l'ensemble des dimensions. Une CAH selon le critère de Ward a enfin été réalisée pour créer la typologie (l'algorithme PAM n'améliorant pas la qualité de la partition). Le nombre optimal de groupes a été défini en utilisant la combinaison de deux critères de qualité de la partition : l'ASWw et le HC, comme précédemment. Une fois la typologie formée, elle a été caractérisée à partir des données individuelles à disposition et des données de consommation de soins.

VI.1.5.3. Application de la MCSA au SNDS

Selon une procédure analogue, la MCSA a ensuite été appliquée aux données du SNDS. Afin de ne pas avoir à aligner les séquences à gauche ou à séparer en deux groupes les individus, seuls les patients SEP ayant pu être identifiés comme ayant la SEP avant le 1er janvier 2010 et ayant au moins huit semestres de consommation de soins consécutifs (les individus décédant durant le huitième semestre inclus), afin de se limiter à un maximum de 33,3% de valeurs manquantes par séquence, ont été considérés (n=64 373). La période d'étude s'étendait de janvier 2010 à décembre 2015. L'unité temporelle étant le semestre calendaire, les parcours de soins étaient donc d'une longueur maximale de 14 états. Les séquences ont été tronquées à droite en cas de décès d'un individu.

Les dimensions du parcours de soins considérées ont été le recours aux médecins généralistes libéraux et aux neurologues libéraux et hospitaliers, aux infirmiers libéraux et aux kinésithérapeutes libéraux, ainsi que le nombre de jours d'hospitalisation liés à la SEP en PMSI MCO. Concernant les recours aux médecins généralistes et aux neurologues, ces nombres correspondaient au nombre de consultations et visites effectuées. Quant au recours aux infirmiers et aux kinésithérapeutes, il était approximé dans cette partie au nombre d'actes effectués au cours du semestre et non plus au nombre de jours où un acte avait été effectué comme dans l'EGB. Le recours aux spécialistes MPR libéraux a été supprimé car seulement un faible recours avait été constaté au global, la majorité des recours à ces spécialistes ayant lieu dans les ACE du PMSI SSR. Enfin, ce n'était plus le nombre d'hospitalisations par semestre qui était considéré mais le nombre de jours d'hospitalisation étant plus proche de la nature des autres dimensions. Au total cinq dimensions ont donc constitué la séquence d'un individu. Compte-tenu de la grande variabilité dans les valeurs de chacune des variables quantifiant le recours à un spécialiste, les états de chaque dimension ont été catégorisés selon la distribution avec l'ajout d'un état correspondant à l'absence de recours durant le semestre considéré.

Les coûts de substitution étaient basés sur les taux de transition observés au sein des données. Le coût des indels a été fixé à 1 pour l'ensemble des dimensions. Afin de prendre en compte les différences de longueur entre les séquences, la normalisation de Yujian et Bo a été appliquée sur les distances calculées par optimal matching. Une CAH selon le critère de Ward a enfin été réalisée pour créer la typologie (l'algorithme PAM n'améliorant pas la qualité de la partition). Le nombre optimal de groupes a été défini en utilisant la combinaison de deux critères de qualité de la partition : l'ASWw et le HC, comme précédemment. Une fois la typologie formée, elle a été caractérisée à partir des données individuelles à disposition et des données de consommation de soins.

VI.2. Résultats

VI.2.1. Analyse de séquences unidimensionnelle sur l'EGB

Parmi les 648 patients identifiés au cours de l'année 2007, le sex-ratio F:H était de 2,7 et l'âge médian en 2007 était de 46 ans (Table 19). Ces patients étaient plus âgés que ceux identifiés en 2008 ou plus tard (p<0,001). De plus, ils étaient plus nombreux en ALD à la fin du suivi (p<0,001) et depuis plus longtemps (p<0,001). Parmi ces patients, 511 (78,9%) avaient une ALD SEP au début du suivi, avec une durée médiane de 6,9 ans (étendue : 0,0-29,2 ans). Au total, 55 décès (8,5%) ont eu lieu durant le suivi, à un âge médian de 63,5 ans. Au cours de la période d'étude de sept ans, la quasi-totalité des patients (98,6%) ont eu recours au moins une fois à un médecin généraliste et 362 patients (55,9%) à un neurologue libéral. Le nombre médian de consultations avec un médecin généraliste par personne-année était de 8,0 consultations (7,1 pour les hommes et 8,3 pour les femmes) alors qu'il était de 1,1 consultation avec un neurologue (libéral ou hospitalier imputé) (1,0 pour les hommes et 1,1 pour les femmes).

Table 19-Caractéristiques des populations pour l'étude du recours aux soins grâce à l'analyse de séquences unidimensionnelle identifiées dans l'EGB en 2007 (N=648) et après (N=352)

	Patients identifiés en 2007 n=648	**Patients identifiés en 2008 et après n=352**	**p-value [a]**
Nombre de femmes (%)	473 (73,0%)	238 (67,6%)	0,086
Année de naissance [b]	1961 (1914-1991)	1966 (1924-1997)	<0,001
Nombre de patients ALD SEP [c] (%)	589 (90,9%)	249 (70,7%)	<0,001
Durée depuis la mise en ALD [b,c,d] (années)	12,7 (0,0-34,6)	6,5 (0,2-33,8)	<0,001
Au moins une autre ALD [c,e] (%)	186 (28,7%)	92 (26,1%)	0,429
Durée de suivi [b] (années)	6,9 (0,0-7,0)	3,5 (0,0-6,0)	<0,001
Nombre de décès (%)	55 (8,5%)	16 (4,5%)	0,029
Au moins une délivrance d'un traitement de fond (%)	368 (56,8%)	171 (48,6%)	0,015
Part du suivi sous traitement [b,f] (%)	61,6 (3,3-100,0)	66,0 (7,6-100,0)	0,218
Consommation de soins			
Consultations avec un généraliste [b,g]	43,0 (0,0-350,0)	17,0 (0,0-266,0)	<0,001
Consultations avec un généraliste par personne-année	8,0	6,5	
Consultations avec un neurologue [b,h]	6,0 (0,0-48,0)	2,0 (0,0-19,0)	<0,001
Consultations avec un neurologue par personne-année	1,1	1,0	
Au moins une hospitalisation liée à la SEP en MCO [i] (%)	307 (47,3%)	166 (47,2%)	1,000
Nombre d'hospitalisations liées à la SEP en MCO par personne-année	0,24	0,31	

[a] P-value résultant de la comparaison des patients identifiés en 2007 et de ceux identifiés en 2008 ou après selon le test du Khi-deux, exact de Fisher ou de Mann-Whitney. [b] Médiane (minimum-maximum). [c] ALD: Affection de Longue Durée. [d] Calculée aux dernières nouvelles (31/12/2013 ou date de décès). [e] ALD autre que la SEP. [f] Pour les patients ayant au moins une prescription de traitement de fond. [g] Nombre total de consultations sur l'ensemble du suivi. [h] Nombre total de consultations avec un neurologue libéral et imputées avec un neurologue hospitalier au cours du suivi. [i] Nombre total d'hospitalisations liées à la SEP au cours du suivi (hors injection de natalizumab à l'hôpital).

Les coûts de substitution ont été estimés en se basant sur les transitions observées dans l'ensemble de la population formée des 1 000 patients (Table 20), afin de bénéficier du maximum d'informations concernant la consommation de soins à disposition. Le coût des indels a été fixé à 0,9

Table 20-Matrice des coûts de substitution pour l'analyse de séquences unidimensionnelle appliquée sur les patients SEP identifiés en 2007 dans l'EGB basés sur les transitions observées dans les données (N=1 000)

Coût	→ 0	→]0;Q_1]	→]Q_1;Q_2]	→]Q_2;Q_3]	→ >Q_3
0 →	0	1,70	1,88	1,88	1,93
]0;Q_1] →	1,70	0	1,44	1,79	1,93
]Q_1;Q_2] →	1,88	1,44	0	1,44	1,83
]Q_2;Q_3] →	1,88	1,79	1,44	0	1,54
>Q_3 →	1,93	1,93	1,83	1,54	0

L'optimal matching et la CAH ont conduit à une partition en cinq clusters (Figure 31), correspondant à une valeur d'ASWw de 0,238 et à une valeur de HC de 0,102. La partition peut être considérée comme étant de faible qualité. Parmi les cinq clusters, les groupes 4 et 5 étaient prédominants avec respectivement 180 (27,8%) et 188 (29,0%) patients (Table 21). Basé sur les durées moyennes dans chaque état, ces deux clusters correspondaient aux patients ayant une forte consommation de soins. Le troisième groupe comprenait 128 individus (19,8%) ayant un recours aux soins intermédiaire. A l'opposé, les patients du groupe 2 (n=124 (19,1%)) avaient un faible recours, et les 28 individus (4,3%) du cluster 1 n'avaient pas de consommation durant la moitié du suivi.

Table 21-Durée moyenne dans chacun des états en année (part de la durée totale) de chaque état selon la classification obtenue pour les patients identifiés en 2007 dans l'EGB (n=648)

Groupe\Etat	n (%)	0	[0–Q1]	[Q1–Q2]	[Q2–Q3]	> Q3
1	28	**3,46**	1,18	0,14	0,64	0,93
	(4,3%)	**(49,4%)**	(16,9%)	(2,0%)	(9,1%)	(13,3%)
2	124	0,31	**4,06**	1,86	0,55	0,19
	(19,1%)	(4,4%)	**(58,0%)**	(26,6%)	(7,9%)	(2,7%)
3	128	0,09	1,39	**4,07**	1,20	0,16
	(19,8%)	(1,3%)	(19,9%)	**(58,1%)**	(17,1%)	(2,3%)
4	180	0,13	0,57	1,76	**2,95**	0,95
	(27,8%)	(1,9%)	(8,1%)	(25,1%)	**(42,1%)**	(13,6%)
5	188	0,10	0,24	0,45	1,37	**4,71**
	(29,0%)	(1,4%)	(3,4%)	(6,4%)	(19,6%)	**(67,3%)**

L'état prédominant de chaque groupe est indiqué en gras.

La consommation de soins annuelle a été classifiée en 5 groupes basés sur les quartiles de la distribution : aucune consommation (0), moins que le premier quartile (]0 ; Q_1]), entre le premier quartile et la médiane (]Q_1 ; Q_2]), entre la médiane et le troisième quartile (]Q_2 ; Q_3]) et supérieur au troisième quartile (>Q_3).

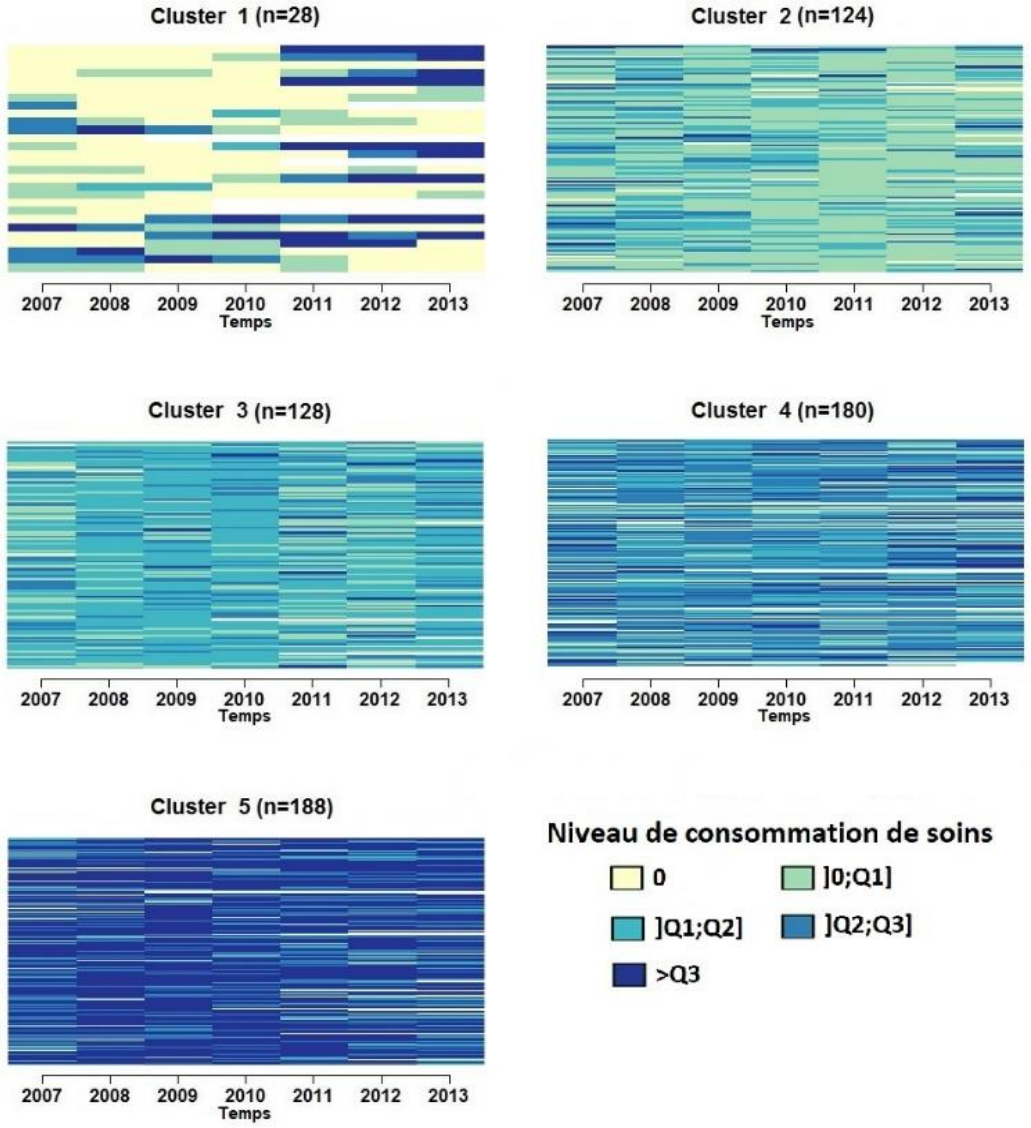


Figure 31-Index plots des groupes obtenus après application de l'analyse de séquences unidimensionnelle sur la population des individus identifiés en 2007 dans l'EGB avec un coût d'indel de 0,9 et des coûts de substitution basés sur les transitions (n=648)

En regardant plus précisément les caractéristiques de consommation de soins et démographiques de chacun des groups de la partition, il est apparu que le recours aux soins différait d'un groupe à l'autre avec un gradient dans la consommation (Table 22). Les nombres de consultations par personne-année avec un médecin généraliste et un neurologue (libéral ou hospitalier imputé) augmentaient respectivement de 2,9 à 14,5 et de 0,2 à 1,3 parmi les groupes. La durée de suivi est la plus faible dans le groupe 1 (p<0,001), probablement lié au haut nombre de décès. De plus, le nombre de patients utilisant au moins un traitement de fond était le plus bas des cinq groupes (p=0,001). Les patients du groupe 2 étaient significativement plus jeunes (p<0,001) et étaient ceux ayant été le plus traités

(p<0,001), avec une durée médiane de suivi sous traitement de près de 75%. De même, le cluster 3 incluait une large proportion de patients traités pendant 76% du suivi. Le groupe 4 était le groupe avec les patients les plus âgés (p<0,001). Le cluster 5 était formé de patients âgés qui avaient significativement plus de comorbidités, quantifiées par le nombre d'ALD autre que pour la SEP, que les autres (p<0,001). Ces patients qui avaient plus de séjours hospitaliers liés à la SEP, étaient majoritairement traités bien que moins longtemps que dans les groupes 2 et 4.

Table 22-Caractéristiques des clusters de la partition en cinq groupes obtenue suite à l'analyse de séquences unidimensionnelle appliquée sur les patients SEP identifiés en 2007 dans l'EGB (n=648)

	Cluster					
	1	**2**	**3**	**4**	**5**	**p-value [a]**
	n=28	n=124	n=128	n=180	n=188	
Nombre de femmes (%)	23 (82,1%)	72 (58,1%)	92 (71,9%)	142 (78,9%)	144 (76,6%)	**<0,001**
Année de naissance [b]	1958,5 (1930-1989)	1965,5 (1929-1991)	1960 (1923-1988)	1957,5 (1914-1986)	1959 (1916-1987)	**<0,001**
Nombre de patients ALD SEP [c] (%)	24 (85,7%)	110 (88,7%)	120 (93,8%)	167 (92,8%)	168 (89,4%)	0,373
Durée depuis la mise en ALD [b,c,d] (années)	12,4 (1,0-26,2)	12,6 (1,0-28,6)	13,0 (3,0-34,6)	12,1 (0,6-34,0)	12,8 (0,0-34,3)	0,629
Au moins une autre ALD [c,e] (%)	9 (32,1%)	19 (15,3%)	32 (25,0%)	48 (26,7%)	78 (41,5%)	**<0,001**
Durée de suivi [b] (années)	6,5 (0,0-7,0)	6,9 (0,0-7,0)	6,9 (0,0-7,0)	7,0 (0,0-7,0)	7,0 (0,0-7,0)	**<0,001**
Nombre de décès (%)	6 (21,4%)	4 (3,2%)	4 (3,1%)	26 (14,4%)	15 (8,0%)	**<0,001**
Au moins une délivrance d'un traitement de fond (%)	6 (21,4%)	78 (62,9%)	74 (57,8%)	95 (52,8%)	115 (61,2%)	**0,001**
Part du suivi sous traitement [b,f] (%)	14,3 (3,6-59,0)	75,5 (3,3-100,0)	76,5 (4,4-100,0)	60,2 (3,3-100,0)	51,8 (4,4-100,0)	**<0,001**
Consommation de soins						
Consultations avec un généraliste [b,g]	19,0 (0,0-73,0)	18,5 (0,0-51,0)	32,0 (10,0-54,0)	48,0 (0,0-96,0)	89,0 (31,0-350,0)	**<0,001**
Consultations avec un généraliste par personne-année	4,1	2,9	4,7	7,5	14,5	
Consultations avec un neurologue [b,h]	0,0 (0,0-8,0)	6,0 (0,0-23,0)	7,0 (0,0-35,0)	5,0 (0,0-38,0)	7,0 (0,0-48,0)	**<0,001**

	Cluster					
	1	**2**	**3**	**4**	**5**	**p-value [a]**
	n=28	n=124	n=128	n=180	n=188	
Consultations avec un neurologue par personne-année	0,2	0,9	1,0	1,1	1,3	
Au moins une hospitalisation liée à la SEP en MCO [i] (%)	8 (28,6%)	61 (49,2%)	44 (34,3%)	84 (46,7%)	110 (58,5%)	**<0,001**
Nombre d'hospitalisations liées à la SEP en MCO par personne-année	0,14	0,23	0,09	0,25	0,35	

[a] P-value résultant de la comparaison des différents groupes selon le test du Khi-deux, exact de Fisher ou de Kruskal-Wallis. [b] Médiane (minimum-maximum). [c] ALD: Affection de Longue Durée. [d] Calculée aux dernières nouvelles (31/12/2013 ou date de décès). [e] ALD autre que la SEP. [f] Pour les patients ayant au moins une prescription de traitement de fond. [g] Nombre total de consultations sur l'ensemble du suivi. [h] Nombre total de consultations avec un neurologue libéral et imputées avec un neurologue hospitalier au cours du suivi. [i] Nombre total d'hospitalisations liées à la SEP au cours du suivi (hors injection de natalizumab à l'hôpital).

VI.2.2. MCSA sur l'EGB

Parmi les 543 patients identifiés en janvier 2007, le sex-ratio F:H était de 2,8 et l'âge médian en 2007 était de 48 ans (Table 23). Un score EDSS « estimé » strictement supérieur à 5 a été retrouvé pour 241 patients (44,4%) au cours du suivi. Au total, 51 décès (9,4%) ont eu lieu durant le suivi, à un âge médian de 64,0 ans. Au cours de la période d'étude s'étalant de 2007 à 2013, la quasi-totalité des patients (98,5%) ont eu recours au moins une fois à un médecin généraliste et 377 patients (69,4%) à un neurologue libéral. Le nombre médian de consultations avec un médecin généraliste par personne-année était de 6,5 consultations (5,6 pour les hommes et 6,9 pour les femmes) alors qu'il était de 0,6 consultation avec un neurologue (libéral ou hospitalier imputé) (0,4 pour les hommes et 0,6 pour les femmes).

Table 23-Caractéristiques de la population identifiée en janvier 2007 dans l'EGB pour l'étude du recours aux soins grâce à la MCSA (N=543)

	Patients identifiés en janvier 2007 n=543
Nombre de femmes (%)	400 (73,7%)
Année de naissance [a]	1959 (1916-1988)
Nombre de patients ALD SEP [b] (%)	509 (93,7%)
Durée depuis la mise en ALD [a, b, c] (années)	13,3 (0,2-34,6)
Au moins une autre ALD [b, d] (%)	152 (28,0%)
Durée de suivi [a] (années)	7,0 (0,0-7,0)
Nombre de décès (%)	51 (9,4%)
Au moins une délivrance d'un traitement de fond (%)	306 (56,4%)
Au moins une délivrance de natalizumab (%)	79 (14,5%)
Part du suivi sous traitement [a, e] (%)	62,1 (3,3-100,0)
Nombre de patients avec score EDSS « estimé » supérieur à 5 (EDSS>5) [f] (%)	241 (44,4%)
Consommation de soins	
Recours aux généralistes [g]	6,54 (0,0-45,91)
Recours aux neurologues [g, h]	0,57 (0,0-19,12)
Recours aux spécialistes MPR [g, i]	0,00 (0,00-3,29)
Au moins une hospitalisation liée à la SEP en MCO (%)	292 (53,8%)
Hospitalisations liées à la SEP en MCO [g, j]	0,57 (0,14-182,62)
Recours aux kinésithérapeutes [g, k]	5,72 (0,00-248,46)
Recours aux infirmiers [g, k]	3,74 (0,00-365,24)

[a] Médiane (minimum-maximum). [b] ALD: Affection de Longue Durée. [c] Calculée aux dernières nouvelles (31/12/2013 ou date de décès). [d] ALD autre que la SEP. [e] Pour les patients ayant au moins une prescription de traitement de fond. [f] Score EDSS « estimé » basé sur l'algorithme défini dans la section V. [g] Recours médian par personne-année (minimum-maximum) [h] Neurologue libéral et neurologue hospitalier imputé. [i] MPR : Médecine Physique et de Réadaptation. [j] Calculé pour les patients ayant au moins une hospitalisation liée à la SEP au cours du suivi. [k] Recours quantifié en nombre jours avec au moins un acte paramédical effectué.

Les bornes choisies pour la catégorisation du recours aux soins par dimension sont indiquées dans la Table 24.

Table 24-Bornes de la catégorisation des dimensions considérées dans la MCSA sur la population identifiée dans l'EGB en janvier 2007

Dimension	Bornes
Consultations avec un généraliste	0, [1 ; 3], [4 ; 6], ≥7
Consultations avec un neurologue (libéral ou hospitalier imputé)	0, 1, ≥2
Consultations avec un spécialiste MPR	0, 1, ≥2
Nombre de jours avec au moins un acte de kinésithérapie	0, [1 ; 24], [25 ; 48], ≥49
Nombre de jours avec au moins un acte infirmier	0, [1 ; 6], [7 ; 48], ≥49
Nombre d'hospitalisations liées à la SEP dans le PMSI MCO	0, 1, ≥2

MPR : Médecine Physique et de Réadaptation

Les coûts de substitution, une matrice par dimension de la séquence, ont été estimés en se basant sur les transitions observées dans les 543 séquences (Table 25). Le coût de l'ensemble des indels a été fixé à 1.

Table 25-Matrices des coûts de substitution pour la MCSA sur la population identifiée dans l'EGB en janvier 2007 basés sur les transitions observées dans les données (N=543)

Consultations avec un généraliste

	→ 0	**→ [1 ; 3]**	**→ [4 ; 6]**	**→ ≥7**	**→ Manquant**
0 →	0	1,45	1,89	1,97	2,00
[1 ; 3] →	1,45	0	1,45	1,85	2,00
[4 ; 6] →	1,89	1,45	0	1,50	2,00
≥7 →	1,97	1,85	1,50	0	2,00
Manquant →	2,00	2,00	2,00	2,00	0

Consultations avec un neurologue (libéral ou hospitalier imputé)

	→ 0	**→ 1**	**→ ≥2**	**→ Manquant**
0 →	0	1,54	1,78	2,00
1 →	1,54	0	1,48	2,00
≥2 →	1,78	1,48	0	2,00
Manquant →	2,00	2,00	2,00	0

Table 25-Matrices des coûts de substitution pour la MCSA sur la population identifiée dans l'EGB en janvier 2007 basés sur les transitions observées dans les données (N=543) (continued)

Consultations avec un spécialiste MPR					
	→ 0	**→ 1**	**→ ≥2**	**→ Manquant**	
0 →	0	1,32	1,37	2,00	
1 →	1,32	0	1,66	2,00	
≥2 →	1,37	1,66	0	2,00	
Manquant →	2,00	2,00	2,00	0	
Nombre de jours avec au moins un acte de kinésithérapie					
	→ 0	**→ [1 ; 24]**	**→ [25 ; 48]**	**→ ≥49**	**→ Manquant**
0 →	0	1,65	1,93	1,98	2,00
[1 ; 24] →	1,65	0	1,62	1,93	2,00
[25 ; 48] →	1,93	1,62	0	1,68	2,00
≥49 →	1,98	1,93	1,68	0	2,00
Manquant →	2,00	2,00	2,00	2,00	0
Nombre de jours avec au moins un acte infirmier					
	→ 0	**→ [1 ; 6]**	**→ [7 ; 48]**	**→ ≥49**	**→ Manquant**
0 →	0	1,46	1,80	1,97	2,00
[1 ; 6] →	1,46	0	1,72	1,95	2,00
[7 ; 48] →	1,80	1,72	0	1,87	2,00
≥49 →	1,97	1,95	1,87	0	2,00
Manquant →	2,00	2,00	2,00	2,00	0
Nombre d'hospitalisations liées à la SEP dans le PMSI MCO					
	→ 0	**→ 1**	**→ ≥2**	**→ Manquant**	
0 →	0	1,24	1,52	2,00	
1 →	1,24	0	1,75	2,00	
≥2 →	1,52	1,75	0	2,00	
Manquant →	2,00	2,00	2,00	0	

MPR : Médecine Physique et de Réadaptation

L'optimal matching et la CAH ont conduit à une partition en cinq clusters (Figure 32). La qualité de la partition était faible avec une valeur d'ASWw de 0,176 et une valeur de HC de 0,157.

En analysant les caractéristiques démographiques, cliniques estimées et de recours aux soins des individus des différents groupes, chacun de ces derniers a pu être qualifié (Table 26). Ainsi, le groupe 1 prédominant (n=271 (49,9%)), composé des individus les plus jeunes (p<0,001) et ayant une ALD

récente débutant en médiane en 2001, a été qualifié comme correspondant à une « SEP peu active ». En effet, il correspondait à des individus ayant un faible recours aux soins au cours du suivi avec 6,3 consultations avec un médecin généraliste par personne-année et 0,8 consultations par personne-année avec un neurologue libéral. Le second groupe était formé de 47 patients (8,7%) ayant un fort recours aux soins paramédicaux avec une médiane de 113,9 jours par personne-année avec un acte de kinésithérapie et 128,5 avec un acte infirmier. De plus, 89,4% des patients de ce groupe présentaient un score EDSS « estimé » strictement supérieur à 5 au cours du suivi correspondant à un phénotype de SEP progressive, d'où sa qualification en groupe « SEP avancée ». Composé de 124 individus (22,8%), le cluster 3 a été qualifié de « SEP active ». En effet, ces patients étaient ceux qui étaient le plus souvent traités à l'hôpital (21,0%) et 65,3% d'entre eux présentaient un handicap moteur (score EDSS « estimé » >5) au cours du suivi. En addition, ils avaient un nombre important d'hospitalisations liées à la SEP (hors injection de natalizumab) par rapport aux autres groupes avec 0,93 hospitalisations par personne-année et un recours aux soins de kinésithérapie élevé, pouvant s'expliquer par la prise en charge d'un handicap ponctuel. Le groupe 4 a été caractérisé de « SEP avancée avec des comorbidités ». Une des principales caractéristiques de ce groupe est le décès des 40 individus (7,4%) le constituant au cours du suivi. Ces derniers étaient plus âgés avec un âge médian de 62 ans en 2007 et 42,5% d'entre eux présentaient une autre ALD que celle pour la SEP. De plus, ils étaient très peu traités par un traitement de fond (20,0%) et avaient un fort recours aux médecins généralistes avec en médiane 10,51 consultations par personne-année. Enfin, le dernier groupe était formé de 61 patients (11,2%) qui avaient dans la quasi-totalité bénéficié d'au moins une délivrance d'un traitement de fond (88,5%) et qui étaient traités sur presque l'ensemble de leur suivi (82,4 % du suivi sous traitement en médiane). De plus, ils avaient un fort recours aux neurologues libéraux avec en médiane 2,8 consultations par personne-année. C'est pourquoi, ce groupe a été considéré comme correspondant aux « SEP sous contrôle ».

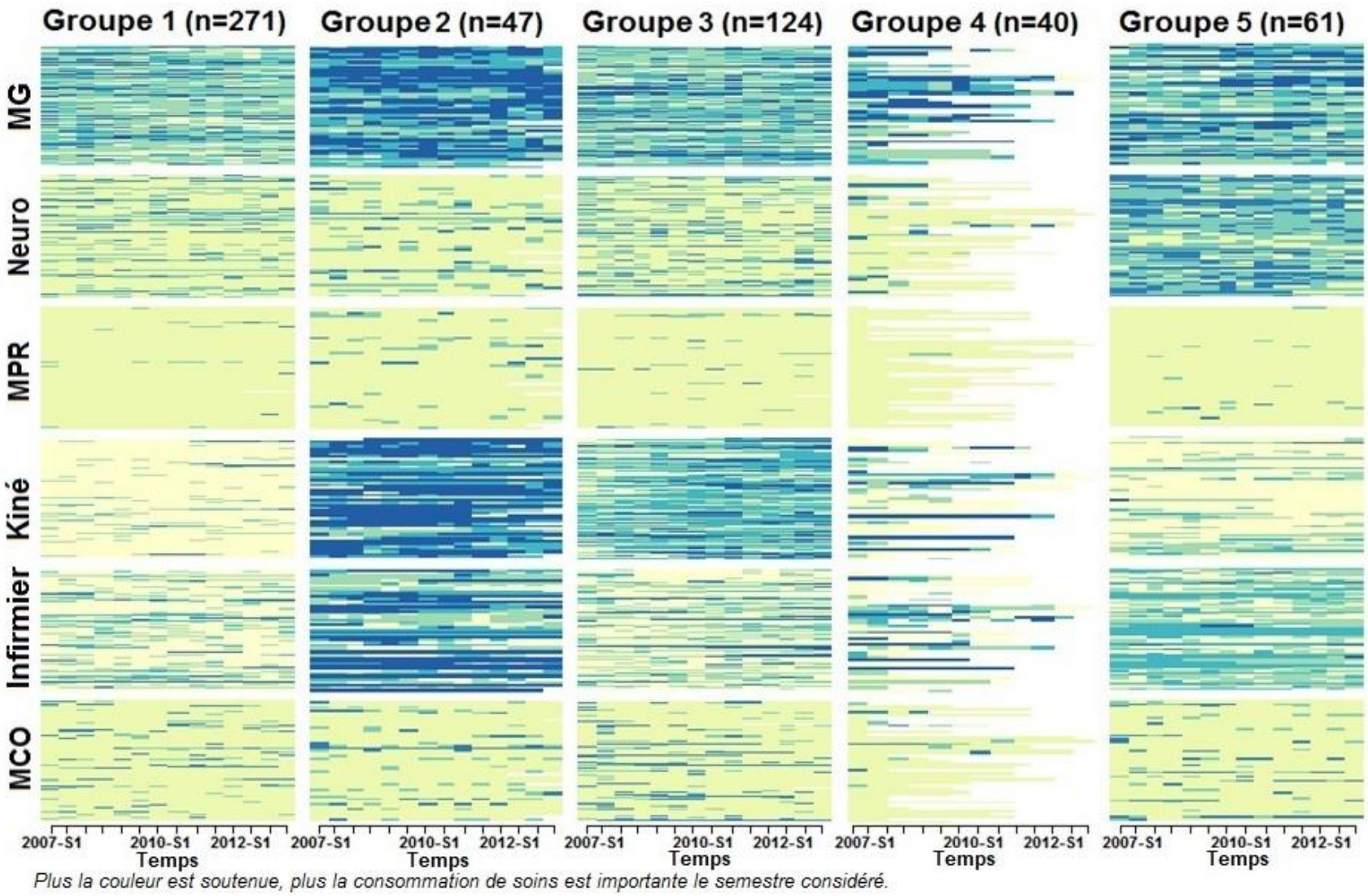


Figure 32-Index plots des groupes obtenus après application de la MCSA sur la population des individus identifiés en janvier 2007 dans l'EGB avec un coût d'indel de 1 et des coûts de substitution basés sur les transitions (n=543)

Table 26-Caractéristiques des clusters de la partition en cinq groupes obtenue suite à la MCSA appliquée sur les patients SEP identifiés en janvier 2007 dans l'EGB (n=543)

	Cluster					
	1	**2**	**3**	**4**	**5**	**p-value [a]**
	n=271 (49,9%)	n=47 (8,7%)	n=124 (22,8%)	n=40 (7,4%)	n=61 (11,2%)	
Nombre de femmes (%)	201 (74,2%)	38 (80,9%)	86 (69,4%)	26 (65,0%)	49 (80,3%)	0,254
Année de naissance [b]	1963 (1928-1988)	1953 (1921-1971)	1955,5 (1925-1981)	1945 (1916-1968)	1961 (1937-1983)	**<0,001**
Nombre de patients ALD SEP [c] (%)	252 (93,0%)	41 (87,2%)	122 (98,4%)	36 (90,0%)	58 (95,1%)	**0,028**
Durée depuis la mise en ALD [b, c, d] (années)	2001 (1979-2014)	1995 (1979-2006)	1999 (1981-2012)	1992,5 (1977-2006)	2001,5 (1990-2012)	**<0,001**
Au moins une autre ALD [c, e] (%)	71 (26,2%)	14 (29,8%)	30 (24,2%)	17 (42,5%)	20 (32,8%)	0,185
Durée de suivi [b] (années)	6,9 (2,3-7,0)	7,0 (5,0-7,0)	7,0 (4,4-7,0)	2,6 (0,0-5,9)	7,0 (6,9-7,0)	**<0,001**
Nombre de décès (%)	3 (1,1%)	5 (10,6%)	3 (2,4%)	40 (100,0%)	0 (0,0%)	**<0,001**
Au moins une délivrance d'un traitement de fond (%)	151 (55,7%)	22 (46,8%)	71 (57,3%)	8 (20,0%)	54 (88,5%)	**<0,001**
Au moins une délivrance de natalizumab (%)	34 (12,5%)	7 (14,9%)	26 (21,0%)	3 (7,5%)	9 (14,8%)	0,177
Part du suivi sous traitement [b, f] (%)	62,5 (3,3-100,0)	59,2 (5,5-100,0)	54,0 (6,6-100,0)	45,0 (13,8-98,7)	82,4 (5,5-100,0)	**0,009**

	Cluster					
	1	**2**	**3**	**4**	**5**	**p-value [a]**
	n=271 (49,9%)	n=47 (8,7%)	n=124 (22,8%)	n=40 (7,4%)	n=61 (11,2%)	
Nombre de patients avec score EDSS « estimé » supérieur à 5 (EDSS>5) [g] (%)	75 (27,7%)	42 (89,4%)	81 (65,3%)	21 (52,5%)	19 (31,1%)	**0,034**
Consommation de soins						
Recours aux généralistes [h]	6,33	14,75	7,25	10,51	9,40	
Recours aux neurologues [h, i]	0,82	0,56	0,99	0,74	2,75	
Recours aux spécialistes MPR [h, j]	0,05	0,22	0,12	0,02	0,05	
Au moins une hospitalisation liée à la SEP en MCO (%)	144 (53,1%)	25 (53,2%)	80 (64,5%)	11 (27,5%)	32 (52,5%)	**0,002**
Hospitalisations liées à la SEP en MCO [h, k]	0,41	0,60	0,93	0,39	0,63	
Recours aux kinésithérapeutes [h, l]	3,88	113,88	57,35	43,30	9,93	
Recours aux infirmiers [h, l]	18,44	128,50	21,59	65,76	21,99	

[a] P-value résultant de la comparaison des différents groupes selon le test du Khi-deux, exact de Fisher ou de Kruskal-Wallis. [b] Médiane (minimum-maximum). [c] ALD: Affection de Longue Durée. [d] Calculée aux dernières nouvelles (31/12/2013 ou date de décès). [e] ALD autre que la SEP. [f] Pour les patients ayant au moins une prescription de traitement de fond. [g] Score EDSS « estimé » basé sur l'algorithme défini dans la section V. [h] Recours médian par personne-année (minimum-maximum) [i] Neurologue libéral et neurologue hospitalier imputé. [j] MPR : Médecine Physique et de Réadaptation. [k] Calculé pour les patients ayant au moins une hospitalisation liée à la SEP au cours du suivi. [l] Recours quantifié en nombre jours avec au moins un acte paramédical effectué.

VI.2.3. MCSA sur le SNDS

Dans un dernier temps, la MCSA a été appliquée sur la population des individus pouvant être identifiés comme ayant une SEP dans le SNDS avant le 1er janvier 2010 et ayant au moins huit semestres de consommation de soins consécutifs, soit 64 373 patients (57,1%). Le sex-ratio était de 2,7 femmes pour 1 homme avec un âge médian de 49 ans en 2010 (Table 27). Parmi ces patients, 30 555 (47,5%) présentaient un score EDSS, estimé à partir de la prescription d'aides techniques à la marche, strictement supérieur à 5 au cours du suivi. Au total, 2 285 décès (3,5%) ont eu lieu durant le suivi, à un âge médian de 67,0 ans. Au cours de la période d'étude s'étalant de 2010 à 2015, la quasi-totalité des patients (98,7%) ont eu recours au moins une fois à un médecin généraliste. De manière analogue, 35 816 patients (55,6%) à un neurologue libéral, 27 269 (42,4%) patients à un neurologue hospitalier et 48 655 (75,6%) patients aux deux. Le nombre médian de consultations avec un médecin généraliste par personne-année était de 5,7 consultations (4,8 pour les hommes et 5,8 pour les femmes) alors qu'il était de 0,8 consultation avec un neurologue (libéral ou hospitalier) (0,7 pour les hommes et 0,8 pour les femmes). La population d'étude (n=64 373) étant privée des cas incidents, elle était plus âgée et présentait un handicap plus important au cours du suivi que la population totale de 112 745 patients.

Le package *TraMineR* ne permettant pas de calculer d'appliquer l'optimal matching à des populations d'une taille supérieure à 46 341 séquences, un échantillon aléatoire formé de 35 000 patients issus de la population source a été constitué (Table 27).

Table 27-Caractéristiques des populations globale (N=112 745), identifiée avant le 1er janvier 2010 dans le SNDS (n=64 373) et de l'échantillon considéré pour l'étude du recours aux soins grâce à la MCSA (n=35 000)

Caractéristique	**Population totale identifiée dans le SNDS N=112 745**	**Patients identifiés avant le 1er janvier 2010 n=64 373**	**Patients considérés dans l'échantillon d'étude n=35 000**
Nombre de femmes (%)	79 947 (70,9%)	47 108 (73,2%)	25 629 (73,2%)
Année de naissance [a]	1964 (1912-2011)	1961 (1912-2003)	1961 (1915-2003)
Age à l'identification de la SEP [a] (années)	40,0 (1,0-101,0)	39,0 (3,0-92,0)	39,0 (5,0-86,0)
Présence d'une ALD SEP au cours du suivi [b] (%)	96 451 (85,5%)	64 319 (99,9%)	34 977 (99,9%)
Durée entre la mise en ALD et l'identification de la SEP [a, b, c] (années)	6,0 (0,0-53,0)	9,0 (0,0-53,0)	9,0 (0,0-53,0)
Durée de suivi [a] (années)	6,0 (0,0-6,0)	6,0 (3,5-6,0)	6,0 (3,5-6,0)

Caractéristique	Population totale identifiée dans le SNDS N=112 745	Patients identifiés avant le 1er janvier 2010 n=64 373	Patients considérés dans l'échantillon d'étude n=35 000
Nombre de décès sur la période d'étude (%)	5 031 (4,5%)	2 285 (3,5%)	1 223 (3,5%)
Régime de rattachement aux dernières nouvelles (%)			
Régime général hors CMU [d, e]	90 541 (80,3%)	54 032 (83,9%)	29 361 (83,4%)
Sections locales mutualistes	8 126 (7,2%)	4 522 (7,0%)	2 434 (7,0%)
Régime agricole	4 152 (3,7%)	2 376 (3,7%)	1 324 (3,8%)
Régime Social des Indépendants	4 253 (3,8%)	1 793 (2,8%)	951 (2,7%)
Bénéficiaire de la CMU [d, e]	4 262 (3,8%)	1 450 (2,3%)	815 (2,3%)
Autres régimes	1 411 (1,3%)	200 (0,3%)	115 (0,3%)
Score de Charlson ajusté sur l'âge [f] (%)			
0	49 735 (44,1%)	27 675 (43,0%)	15 081 (43,1%)
1-2	37 980 (33,7%)	25 631 (39,8%)	13 920 (39,8%)
3-4	12 501(11,1%)	8 067 (12,5%)	4 393 (12,6%)
≥5	3 533 (3,1%)	1 706 (2,7%)	921 (2,6%)
Manquant	8 996 (8,0%)	1 294 (2,0%)	685 (2,0%)
Au moins une délivrance d'un traitement de fond [g] (%)	60 559 (53,7%)	36 558 (56,8%)	19 878 (56,8%)
Au moins une délivrance d'un traitement de fond spécifique (%)	53 457 (47,4%)	31 332 (48,7%)	16 957 (48,4%)
Au moins une délivrance de natalizumab (%)	8 240 (7,3%)	5 831 (9,1%)	3 167 (9,0%)
Nombre de patients avec score EDSS « estimé » supérieur à 5 (EDSS>5) [h] (%)	44 149 (39,2%)	30 555 (47,5%)	16 669 (47,6%)

Caractéristique	Population totale identifiée dans le SNDS N=112 745	Patients identifiés avant le 1er janvier 2010 n=64 373	Patients considérés dans l'échantillon d'étude n=35 000
FDep de la commune de résidence en 2009 [i] (%)			
1er quintile (communes les plus favorisées)	20 939 (18,6%)	11 374 (17,7%)	6 180 (17,7%)
2ème quintile	20 522 (18,2%)	11 674 (18,1%)	6 346 (18,1%)
3ème quintile	21 353 (18,9%)	12 474 (19,4%)	6 775 (19,4%)
4ème quintile	21 210 (18,8%)	12 383 (19,2%)	6 734 (19,2%)
5ème quintile (communes les plus défavorisées)	20 929 (18,6%)	12 055 (18,7%)	6 564 (18,8%)
Manquant	7 792 (6,9%)	4 413 (6,9%)	2 401 (6,9%)
APL de la commune de résidence en 2010 [a, j] (%)			
Médecin généraliste	70,9 (0,0-617,2)	71,3 (0,0-617,2)	71,3 (0,0-617,2)
Kinésithérapeute	67,4 (0,0-631,6)	67,0 (0,0-631,6)	67,1 (0,0-436,3)
Infirmier	67,2 (0,0-701,2)	67,4 (0,0-701,2)	67,4 (0,0-701,2)
Pharmacie	36,4 (0,0-824,5)	36,3 (0,0-824,5)	36,3 (0,0-824,5)
Manquant (%)	7 007 (6,2%)	4 077 (6,3%)	2 214 (6,3%)
Consommation de soins pour les patients suivis au moins un an	n=106 176 (94,2%)		
Recours aux généralistes [k]	6,17 (0,00-360,04)	5,67 (0,00-281,22)	5,67 (0,00-281,22)
Recours aux neurologues libéraux [k]	0,17 (0,00-100,63)	0,17 (0,00-34,17)	0,17 (0,00-34,17)
Recours aux neurologues hospitaliers [k]	0,00 (0,00-18,57)	0,00 (0,00-8,67)	0,00 (0,00-8,67)
Recours aux neurologues libéraux et hospitaliers [k]	0,83 (0,00-101,54)	0,83 (0,00-34,17)	0,83 (0,00-34,17)
Au moins une hospitalisation liée à la SEP en MCO (%)	56 200 (52,9%)	30 446 (47,3%)	16 630 (47,5%)
Nombre de séjours hospitaliers liés à la SEP en MCO [k, l]	0,67 (0,17-35,78)	0,67 (0,00-32,51)	0,67 (0,00-24,50)
Nombre de jours d'hospitalisations liées à la SEP [l, m]	1,60 (0,17-255,12)	1,50 (0,17-95,17)	1,50 (0,17-85,10)

Caractéristique	Population totale identifiée dans le SNDS N=112 745	Patients identifiés avant le 1er janvier 2010 n=64 373	Patients considérés dans l'échantillon d'étude n=35 000
Au moins un séjour en SSR lié à la SEP (%)	19 924 (18,8%)	15 435 (24,0%)	8 319 (23,8%)
Recours aux kinésithérapeutes [k, n]	1,67 (0,00-1691,36)	3,33 (0,00-837,84)	3,33 (0,00-837,84)
Recours aux infirmiers [k, n]	3,35 (0,00-7604,59)	3,50 (0,00-4700,78)	3,50 (0,00-4700,78)
Recours global avec un praticien [k, o]	15,50 (0,00-579,27)	14,34 (0,00-418,92)	14,34 (0,00-418,92)

[a] Médiane (minimum-maximum). [b] ALD : Affection de Longue Durée. [c] Calculée pour les patients ayant été identifié en premier grâce à leur statut ALD (N=92 684). [d] Couverture Maladie Universelle. [e] Bénéficiaire de la CMU sur la période d'étude 2010-2015. [f] Score de Charlson selon la définition de Bannay et al. [g] Traitements hors AMM inclus (azathioprine, méthotrexate, mycophénolate mofétil). [h] Score EDSS « estimé » basé sur l'algorithme défini dans la section V. [i] FDep : Indice de défavorisation sociale. [j] APL : Accessibilité Potentielle Localisée. [k] Recours médian par personne-année (minimum-maximum). [l] Calculé pour les patients ayant au moins une hospitalisation liée à la SEP au cours du suivi. [m] Nombre médian de jours d'hospitalisations liées à la SEP par personne-année (minimum-maximum). [n] Recours quantifié en nombre d'actes paramédicaux effectués. [o] Ensemble des consultations en libéral et externes toutes spécialités confondues.

Les bornes choisies pour la catégorisation du recours aux soins par dimension sont indiquées dans la Table 28. Elles ont été fixées en accord avec la distribution des différents recours aux soins sur la période d'étude pour les 64 373 individus considérés dans la population globale.

Table 28-Bornes de la catégorisation des dimensions considérées dans la MCSA sur la population identifiée avant le 1er janvier 2010 dans le SNDS (n=64 373)

Dimension	Bornes
Consultations avec un généraliste	0, 1, 2, [3 ; 6], ≥7
Consultations avec un neurologue libéral ou hospitalier (généraliste hospitalier exclu)	0, 1, ≥2
Nombre d'actes de kinésithérapie	0, [1 ; 10], [11 ; 40], ≥41
Nombre d'actes infirmiers	0, [1 ; 4], [5 ; 25], ≥26
Nombre jours d'hospitalisations liées à la SEP dans le PMSI MCO	0, 1, [2 ; 3], ≥4

Les coûts de substitution, une matrice par dimension de la séquence, ont été estimés en se basant sur les transitions observées dans l'ensemble des 64 373 séquences, afin de tirer parti du maximum de

données de consommation de soins à disposition (Table 29). Le coût de l'ensemble des indels a été fixé à 1.

Table 29-Matrices des coûts de substitution pour la MCSA sur l'échantillon aléatoire de 35 000 patients issus de la population identifiée avant le 1er janvier 2010 dans le SNDS basés sur les transitions observées dans l'ensemble des données (N=64 373)

Consultations avec un généraliste					
	→ 0	**→ 1**	**→ 2**	**→ [3 ; 6]**	**→ ≥7**
0 →	0	1,53	1,76	1,84	1,98
1 →	1,53	0	1,56	1,65	1,96
2 →	1,76	1,56	0	1,46	1,92
[3 ; 6] →	1,84	1,65	1,46	0	1,49
≥7 →	1,98	1,96	1,92	1,49	0
Consultations avec un neurologue (libéral ou hospitalier imputé)					
	→ 0	**→ 1**	**→ ≥2**		
0 →	0	1,55	1,77		
1 →	1,55	0	1,49		
≥2 →	1,77	1,49	0		
Nombre de jours avec au moins un acte de kinésithérapie					
	→ 0	**→ [1 ; 10]**	**→ [11 ; 40]**	**→ ≥41**	
0 →	0	1,51	1,83	1,98	
[1 ; 10] →	1,51	0	1,61	1,93	
[11 ; 40] →	1,83	1,61	0	1,63	
≥41 →	1,98	1,93	1,63	0	
Nombre de jours avec au moins un acte infirmier					
	→ 0	**→ [1 ; 4]**	**→ [5 ; 25]**	**→ ≥26**	
0 →	0	1,46	1,80	1,97	
[1 ; 4] →	1,46	0	1,72	1,95	
[5 ; 25] →	1,80	1,72	0	1,87	
≥26 →	1,97	1,95	1,87	0	
Nombre d'hospitalisations liées à la SEP dans le PMSI MCO					
	→ 0	**→ 1**	**→ [2 ; 3]**	**→ ≥4**	
0 →	0	1,29	1,42	1,75	
1 →	1,29	0	1,83	1,87	
[2 ; 3] →	1,42	1,83	0	1,75	
≥4 →	1,75	1,87	1,75	0	

L'optimal matching et la CAH ont conduit à une partition en cinq clusters (Figure 33). La qualité de la partition était très faible avec des valeurs d'ASWw et de HC étant respectivement égales à 0,047 et 0,283.

Les caractéristiques des cinq groupes de la typologie finale sont rassemblées dans la Table 30. Le premier groupe formé de 3 628 individus (10,4%) était le plus jeune avec un âge médian de 42 ans au début de la période d'étude en 2010. Ces patients étaient majoritairement traités (87,1%) et 65,6% d'entre eux avaient eu au moins une perfusion de natalizumab à l'hôpital pouvant expliquer le plus fort nombre de jours d'hospitalisation en PMSI MCO que dans les autres groupes avec en médiane 8,3 jours d'hospitalisation par personne-année. De plus, 56,1% des patients de ce groupe ont eu un score EDSS « estimé » indiquant un handicap moteur au cours du suivi traduction de l'activité de la maladie. Au vu de ces caractéristiques, le groupe 1 a été qualifié de « SEP actives sous contrôle ». Le cluster 2 (n=9 442 (26,9%)) était formé de patients ayant un âge médian de 45 ans en 2010 et étant eux aussi majoritairement traités (78,6% ayant eu au moins une délivrance d'un traitement de fond au cours du suivi) avec le plus souvent des injectables (73,8%). Concernant leur consommation de soins, ils avaient un fort recours aux neurologues avec 2,0 consultations par personne-année, d'où leur classification en « SEP sous contrôle ». Les 10 139 cas SEP (29,0%) du groupe 3 étaient âgés de 51 ans en 2010 et seulement 34,6% d'entre eux avaient eu au moins une délivrance d'un traitement de fond au cours du suivi. En accord avec ce faible nombre de traités, ils avaient un faible recours aux neurologues (0,2 consultations par personne-année), ne bénéficiaient d'aucun acte de kinésithérapie en médiane au cours du suivi et avaient un faible recours global aux soins avec 12,5 consultations avec un professionnel de santé par personne-année. Ces informations ont conduit à qualifier ce groupe de « SEP peu active ». Le quatrième groupe était composé de 9 373 individus (26,8%), dont seulement 36,2% avaient eu au moins une délivrance d'un traitement de fond spécifique à la SEP au cours du suivi. Ils avaient un fort recours aux actes de kinésithérapies avec en médiane 50,2 actes par personne-année et 6,7 consultations par personne-année avec un médecin généraliste. De plus, 40,7% des patients de ce groupe avaient eu au moins un séjour lié à la SEP en SSR, et 70,8% un score EDSS « estimé » strictement supérieur à 5, indiquant un handicap moteur. Par conséquent, le groupe 4 a été rapproché du phénotype de « SEP progressive ».Enfin, le dernier groupe de 2 418 patients (6,9%) était le plus âgé (âge médian de 57 ans en 2010) avec une mise en ALD ancienne (année médiane de mise en ALD=1996). De plus, ils étaient peu traités (30,9% avec une thérapeutique spécifique) et avaient un fort recours aux médecins généralistes avec en médiane 9,5 consultations par personne-année et près de 20 consultations par personne-année avec un professionnel de santé toutes spécialités confondues. Ils effectuaient également un grand nombre d'actes auprès de kinésithérapeutes et infirmiers avec des nombre d'actes médians par personne-année respectifs de 89,2 et 440,9. De plus, une aide technique à la marche a été repérée pour 90,5% des individus de ce groupe. Pour ces raisons, ils ont été classés

comme étant des « SEP progressives » comme le groupe 4 mais avec un recours aux soins plus important. De plus, il est apparu que ces patients ayant le recours aux soins le plus élevé des cinq groupes, avaient les plus fortes valeurs d'APL pour les spécialités considérées.

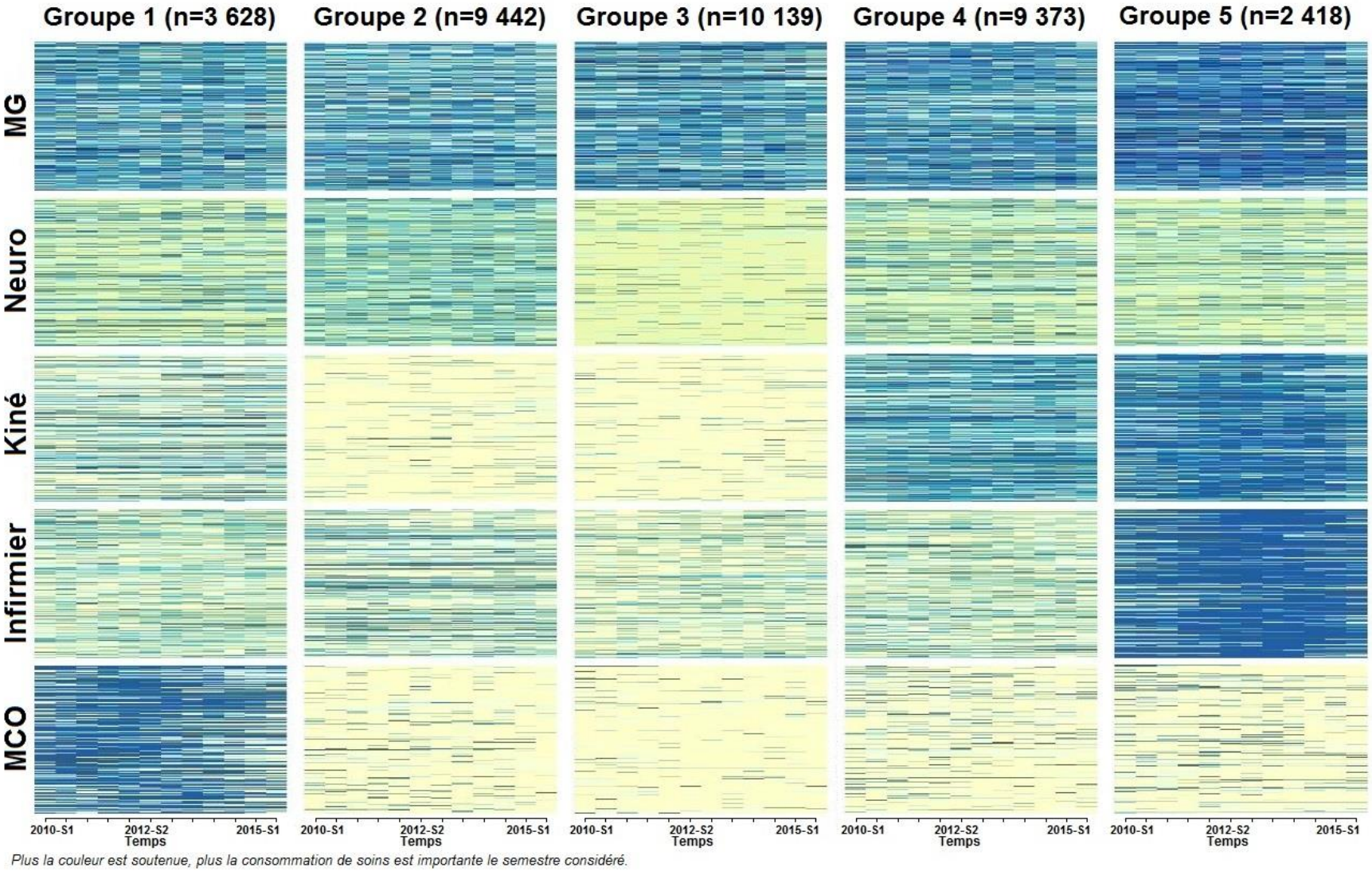


Figure 33-Index plots des clusters de la partition en cinq groupes obtenue suite à la MCSA sur l'échantillon aléatoire de patients SEP identifiés avant janvier 2010 dans le SNDS (N=35 000)

Table 30-Caractéristiques des clusters de la partition en cinq groupes obtenue suite à la MCSA appliquée sur l'échantillon de patients SEP identifiés avant janvier 2010 dans le SNDS (n=35 000)

	Cluster					
Caractéristiques	**1**	**2**	**3**	**4**	**5**	**p-value[a]**
	n=3 628	n=9 442	n=10 139	n=9 373	n=2 418	
	(10,4%)	(26,9%)	(29,0%)	(26,8%)	(6,9%)	
Nombre de femmes (%)	2 527 (69,7%)	7 066 (74,8%)	7 540 (74,4%)	6 738 (71,9%)	1 758 (72,7%)	**<0,001**
Année de naissance [b]	1968	1965	1959	1956	1953	**<0,001**
	(1932-2001)	(1915-2000)	(1919-2003)	(1919-2000)	(1917-1991)	
Age à l'identification de la SEP [b] (années)	33,0 (8,0-70,0)	37,0 (8,0-82,0)	39,0 (6,0-82,0)	42,0 (5,0-86,0)	42,0 (13,0-81,0)	**<0,001**
Présence d'une ALD SEP au cours du suivi [c] (%)	3 626 (99,9%)	9 440 (100,0%)	10 135 (100,0%)	9 361 (99,9%)	2 415 (99,9%)	**0,023**
Durée entre la mise en ALD et l'identification de la SEP [b, c, d] (années)	7,0 (0,0-36,0)	7,0 (0,0-47,0)	9,0 (0,0-47,0)	10,0 (0,0-49,0)	14,0 (0,0-53,0)	**<0,001**
Durée de suivi [b] (années)	6,0 (3,5-6,0)	6,0 (3,5-6,0)	6,0 (3,5-6,0)	6,0 (3,5-6,0)	6,0 (3,5-6,0)	**<0,001**
Nombre de décès sur la période d'étude (%)	47 (1,3%)	105 (1,1%)	172 (1,7%)	634 (6,8%)	265 (11,0%)	**<0,001**
Régime de rattachement aux dernières nouvelles (%)						**<0,001**
Régime général hors CMU [e, f]	3 037 (83,7%)	7 897 (83,6%)	8 559 (84,4%)	7 829 (83,5%)	2 039 (84,3%)	
Sections locales mutualistes	245 (6,8%)	669 (7,1%)	677 (6,7%)	692 (7,4%)	151 (6,2%)	
Régime agricole	97 (2,7%)	357 (3,8%)	372 (3,7%)	391 (4,2%)	107 (4,4%)	
Régime Social des Indépendants	88 (2,4%)	242 (2,6%)	240 (2,4%)	304 (3,2%)	77 (3,2%)	
Bénéficiaire de la CMU [e, f]	145 (4,0%)	254 (2,7%)	256 (2,5%)	128 (1,4%)	32 (1,3%)	
Autres régimes	16 (0,4%)	23 (0,2%)	35 (0,3%)	29 (0,3%)	12 (0,5%)	

	Cluster					
Caractéristiques	**1**	**2**	**3**	**4**	**5**	**p-value[a]**
	n=3 628 (10,4%)	n=9 442 (26,9%)	n=10 139 (29,0%)	n=9 373 (26,8%)	n=2 418 (6,9%)	
Score de Charlson ajusté sur l'âge [g] (%)						**<0,001**
0	2 300 (63,4%)	5 404 (57,2%)	3 986 (39,3%)	2 863 (30,5%)	528 (21,8%)	
1-2	1 056 (29,1%)	3 177 (33,6%)	4 178 (41,2%)	4 364 (46,6%)	1 145 (47,3%)	
3-4	190 (5,2%)	569 (6,0%)	1 387 (13,7%)	1 663 (17,7%)	584 (24,2%)	
≥5	32 (0,9%)	104 (1,1%)	290 (2,9%)	357 (3,8%)	138 (5,7%)	
Manquant	50 (1,4%)	188 (2,0%)	298 (2,9%)	126 (1,3%)	23 (1,0%)	
Au moins une délivrance d'un traitement de fond [h] (%)	3 161 (87,1%)	7 417 (78,6%)	3 504 (34,6%)	4 696 (50,1%)	1 100 (45,5%)	**<0,001**
Au moins une délivrance d'un traitement de fond spécifique (%)	2 934 (80,9%)	6 997 (74,1%)	2 891 (28,5%)	3 389 (36,2%)	746 (30,9%)	**<0,001**
Au moins une délivrance de natalizumab (%)	2 381 (65,6%)	267 (2,8%)	114 (1,1%)	332 (3,5%)	73 (3,0%)	**<0,001**
Nombre de patients avec score EDSS « estimé » supérieur à 5 (EDSS>5) [i] (%)	2 034 (56,1%)	2 634 (27,9%)	3 179 (31,4%)	6 634 (70,8%)	2 188 (90,5%)	**<0,001**

	Cluster					
Caractéristiques	**1**	**2**	**3**	**4**	**5**	**p-value[a]**
	n=3 628	n=9 442	n=10 139	n=9 373	n=2 418	
	(10,4%)	(26,9%)	(29,0%)	(26,8%)	(6,9%)	
FDep de la commune de résidence en 2009 [j] (%)						**<0,001**
1er quintile (communes les plus favorisées)	628 (17,3%)	1 624 (17,2%)	1 718 (16,9%)	1 816 (19,4%)	394 (16,3%)	
2ème quintile	640 (17,6%)	1 826 (19,3%)	1 766 (17,4%)	1 681 (17,9%)	433 (17,9%)	
3ème quintile	712 (19,6%)	1 820 (19,3%)	1 931 (19,0%)	1 825 (19,5%)	487 (20,1%)	
4ème quintile	667 (18,4%)	1 821 (19,3%)	2 019 (19,9%)	1 757 (18,7%)	470 (19,4%)	
5ème quintile (communes les plus défavorisées)	728 (20,1%)	1 677 (17,8%)	2 029 (20,0%)	1 684 (18,0%)	446 (18,4%)	
Manquant	253 (7,0%)	674 (7,1%)	676 (6,7%)	610 (6,5%)	188 (7,8%)	
APL de la commune de résidence en 2010 [b, k] (%)						
Médecin généraliste	71,8	70,7	70,6	71,9	76,4	**<0,001**
	(0,0-307,8)	(0,0-216,0)	(0,0-617,2)	(0,0-205,7)	(0,0-182,4)	
Kinésithérapeute	67,6	65,8	65,2	69,1	73,9	**<0,001**
	(0,0-242,9)	(0,0-301,5)	(0,0-436,3)	(0,0-436,3)	(0,0-436,3)	
Infirmier	68,8	66,9	66,2	65,8	77,7	**<0,001**
	(0,0-623,0)	(0,0-682,7)	(0,0-623,0)	(0,0-623,0)	(0,0-701,2)	
Pharmacie	36,6	35,5	36,2	36,6	37,0	**<0,001**
	(0,0-140,3)	(0,0-292,9)	(0,0-824,5)	(0,0-200,0)	(0,0-105,5)	
Manquant (%)	230 (6,3%)	631 (6,7%)	611 (6,0%)	575 (6,1%)	167 (6,9%)	

	Cluster					
Caractéristiques	**1**	**2**	**3**	**4**	**5**	**p-value[a]**
	n=3 628	n=9 442	n=10 139	n=9 373	n=2 418	
	(10,4%)	(26,9%)	(29,0%)	(26,8%)	(6,9%)	
Consommation de soins pour les patients suivis au moins un an						
Recours aux généralistes [l]	5,17	4,67	5,50	6,17	9,51	**<0,001**
	(0,00-64,88)	(0,00-64,21)	(0,00-80,22)	(0,00-281,35)	(0,00-105,36)	
Recours aux neurologues libéraux [l]	0,17	1,33	0,00	0,17	0,17	**<0,001**
	(0,00-34,19)	(0,00-32,85)	(0,00-5,50)	(0,00-14,34)	(0,00-16,18)	
Recours aux neurologues hospitaliers [l]	0,17	0,00	0,00	0,00	0,00	**<0,001**
	(0,00-5,67)	(0,00-8,67)	(0,00-3,84)	(0,00-6,00)	(0,00-4,50)	
Recours aux neurologues libéraux et hospitaliers [l]	0,83	2,00	0,17	1,00	0,67	**<0,001**
	(0,00-34,19)	(0,00-32,85)	(0,00-5,50)	(0,00-14,84)	(0,00-16,18)	
Au moins une hospitalisation liée à la SEP en MCO (%)	3 621 (99,8%)	4 170 (44,2%)	2 715 (26,8%)	4 830 (51,5%)	1 294 (53,5%)	**<0,001**
Nombre de séjours hospitaliers liés à la SEP en MCO [l, m]	7,00	0,50	0,33	0,50	0,50	**<0,001**
	(0,17-24,52)	(0,17-13,01)	(0,17-11,01)	(0,17-12,84)	(0,17-14,84)	
Nombre de jours d'hospitalisations liées à la SEP en MCO [m, n]	8,34	0,83	0,83	1,33	1,83	**<0,001**
	(0,17-81,72)	(0,17-21,85)	(0,17-21,85)	(0,17-85,16)	(0,17-36,52)	
Au moins un séjour en SSR lié à la SEP (%)	1 303 (35,9%)	1 033 (10,9%)	919 (9,1%)	3 814 (40,7%)	1 250 (51,7%)	**<0,001**

	Cluster					
Caractéristiques	**1**	**2**	**3**	**4**	**5**	**p-value[a]**
	n=3 628	n=9 442	n=10 139	n=9 373	n=2 418	
	(10,4%)	(26,9%)	(29,0%)	(26,8%)	(6,9%)	
Recours aux kinésithérapeutes [l, o]	8,34	0,00	0,00	50,20	89,23	**<0,001**
	(0,00-281,35)	(0,00-98,56)	(0,00-129,80)	(0,00-621,94)	(0,00-838,22)	
Recours aux infirmiers [l, o]	4,50	3,00	1,83	4,00	440,87	**<0,001**
	(0,00-1345,05)	(0,00-3655,90)	(0,00-2162,09)	(0,00-1386,58)	(2,00-4702,92)	
Recours global avec un praticien [l, p]	14,84	13,68	12,51	15,51	19,18	**<0,001**
	(0,17-110,24)	(0,17-152,43)	(0,00-223,48)	(0,00-419,11)	(0,83-155,53)	

[a] P-value résultant de la comparaison des différents groupes selon le test du Khi-deux, exact de Fisher ou de Kruskal-Wallis. [b] Médiane (minimum-maximum). [c] ALD : Affection de Longue Durée. [d] Calculée pour les patients ayant été identifié en premier grâce à leur statut ALD (N=92 684). [e] Couverture Maladie Universelle. [f] Bénéficiaire de la CMU sur la période d'étude 2010-2015. [g] Score de Charlson selon la définition de Bannay et al. [h] Traitements hors AMM inclus (azathioprine, méthotrexate, mycophénolate mofétil). [i] Score EDSS « estimé » basé sur l'algorithme défini dans la section V. [j] FDep : Indice de défavorisation sociale. [k] APL : Accessibilité Potentielle Localisée. [l] Recours médian par personne-année (minimum-maximum). [m] Calculé pour les patients ayant au moins une hospitalisation liée à la SEP au cours du suivi. [n] Nombre médian de jours d'hospitalisations liées à la SEP par personne-année (minimum-maximum). [o] Recours quantifié en nombre d'actes paramédicaux effectués.

[p] Ensemble des consultations en libéral et externes toutes spécialités confondues.

VI.2.4. MCSA sur des données canadiennes

L'ensemble des opinions et conclusions exprimées dans cette section sont propres aux auteurs, et ne reflètent pas les opinions ou les politiques des régisseurs des données (All inferences, opinions, and conclusions drawn in this section are those of the authors, and do not reflect the opinions or policies of the Data Steward(s)).

VI.2.4.1. Contexte

En France, il est très difficile d'apparier des données cliniques issues de cohorte aux données des bases médico-administratives. En effet, la procédure peut s'avérer complexe et peut ne pas être autorisée par la CNIL, contrairement aux provinces canadiennes où cet appariement est possible, notamment en Colombie Britannique (BC) où une base clinique recensant l'ensemble des cas SEP a été mise en place depuis 1980 et est depuis chaînée à des données médico-administratives. Il est donc apparu intéressant d'envisager un séjour de recherche sur place afin de profiter de cette opportunité de données cliniques et médico-administratives chaînées.

C'est pourquoi, une mobilité internationale au sein de l'équipe PiMS (Pharmacoepidemiology in Multiple Sclerosis) dirigée par Helen Tremlett située à l'Université de Colombie Britannique (UBC) de Vancouver, Canada, a eu lieu du 10 Juillet au 15 Septembre 2017. Cette mobilité a été possible grâce aux financements obtenus auprès du Réseau Doctoral en Santé Publique animé par l'EHESP, de l'EA 7449 REPERES et de l'Ecole des Docteurs de l'Université Bretagne-Loire (UBL) et le Conseil Régional de Bretagne. L'objectif de ce séjour était dans un premier temps d'identifier une typologie de parcours de soins des personnes ayant une SEP à partir des données médico-administratives de la province de BC. Puis, il s'agissait de caractériser cette typologie en utilisant des données cliniques chaînées aux données de remboursement (date de début de la SEP, niveau de handicap, …), qui sont connues dans la littérature pour influencer la consommation de soins. L'intérêt majeur de ce projet était la possibilité de croiser des données de consommation de soins et des données cliniques sur l'historique de la maladie chez chacun des patients, ce qui est actuellement impossible en France.

VI.2.4.2. Méthodes

Les données utilisées lors de la mobilité étaient issues d'une cohorte rétrospective chaînant des données médico-administratives et des données cliniques de la base BCMS (British Columbia Multiple Sclerosis). Cette dernière a été créée en 1980 et recueille depuis cette date de manière prospective les données cliniques des personnes ayant une SEP de la province canadienne de BC et ayant été en contact avec une des cliniques de la SEP de la province. Les données médico-administratives et cliniques ont été chaînées grâce au numéro unique de chacun des patients commun aux deux bases, et qui a été supprimé après la jointure des bases. Les données médico-administratives étaient issues du Medical Service Plan donnant des informations sur les consultations avec des physiciens et le motif de recours (219), de la Discharge Abstract Database contenant les données

relatives aux séjours hospitaliers (220), de la base PharmaNet recueillant les informations sur les prescriptions médicamenteuses de la province (221). A ces données, ont été ajoutées des informations sur le statut vital des patients issues de la base BC Vital Statistics Agency (222) et sur leur émigration et/ou immigration de la province à partir des fichiers du BC Ministry of Health's Registration and Premium Billing (223). Enfin des données sur le niveau socio-économique du lieu de résidence des patients étaient à disposition grâce à la base Census Geodata (224). Les données pour l'ensemble des sources étaient disponibles du 1er janvier 1996 au 31 décembre 2008, soit une période de suivi maximale de 13 ans. L'ensemble du chaînage entre les bases a été effectué par la plate-forme Population Data BC. L'étude a reçu l'accord du comité de recherche clinique de UBC et l'utilisation des données a été autorisée par le Ministère de la Santé de Colombie Britannique, la BC Vital Statistics Agency et la base BC PharmaNet.

Une première partie du séjour de recherche a été consacrée à la constitution des cohortes de patients nécessaires à l'analyse statistique. En effet, dans le cadre de cette étude, deux cohortes distinctes ont été créées : une première cohorte (cohorte 1) composée des patients vivants en BC avec un démarrage de la SEP entre janvier 1996 et décembre 2004 ; et une seconde (cohorte 2) formée des patients vivants au 1er janvier 1996 en BC et ayant une SEP. Le début de la période d'étude était spécifique à chacune des deux cohortes : pour la cohorte 1, le début du suivi a été fixé au premier jour du mois du déclenchement de la SEP ; alors que pour la cohorte 2 il a été fixé au 1er janvier 1996. Cependant dans les deux cas, la fin du suivi a été fixée à soit la fin décembre 2008, soit la mort du patient ou l'émigration du patient dans une autre province du Canada ou un autre pays, définie comme une sortie des fichiers du BC Ministry of Health's Registration and Premium Billing de strictement plus de 93 jours consécutifs. Les patients de la cohorte 1 devaient avoir au moins une valeur de score EDSS, mesurant le niveau de handicap principalement moteur, durant leur suivi. Quant aux patients de la seconde cohorte, ils devaient avoir au moins une valeur de score EDSS durant le suivi ou dans les 3 ans pré-inclusion. De plus, les patients devaient être présents dans les bases médico-administratives à leur inclusion et au moins un an pré-inclusion, et avoir au minimum 4 ans de suivi dans ces dernières.

Les différentes variables de consommation de soins ainsi que celles servant à la caractérisation ont été créées dans un second temps en se basant sur les travaux déjà effectués dans le cadre de la thèse sur les données françaises. Le parcours de soins a été divisé en six dimensions : consultations chez un généraliste, un neurologue, un spécialiste de médecine interne, un spécialiste d'un des symptômes de la SEP (urologue, médecin de médecine physique et de réadaptation, ophtalmologue), un psychiatre et des séjours hospitaliers tous diagnostics confondus. Pour chacune des deux cohortes, l'unité de temps choisie était une période de six mois. Une variable composite catégorielle, correspondant à la somme du nombre de consultations ou séjours durant une unité de temps, a été créée pour chaque dimension

du parcours de soins afin de représenter le recours aux soins (Table 31). Le nombre minimal de valeurs dans un parcours de soins d'un spécialiste était donc de 9 et le maximum de 26.

Table 31-Bornes de la catégorisation des dimensions considérées dans la multichannel sequence analysis sur la population identifiée dans les bases canadiennes (cohortes 1 et 2)

Dimension	Bornes
Consultations avec un généraliste	0, [1 ; 3], [4 ; 6], ≥7
Consultations avec un neurologue	0, 1, ≥2
Consultations avec un spécialiste de médecine interne	0, ≥1
Consultations avec un spécialiste des symptômes de la SEP (urologue, ophtalmologue, MPR)	0, 1, ≥2
Consultations avec un psychiatre ou un neuropsychiatre	0, ≥1
Séjours à l'hôpital (tous diagnostics confondus)	0, 1, ≥2

MPR : Médecine Physique et de Réadaptation

Les variables de caractérisation des typologies finales étaient issues de la base BCMS, notamment le sexe, l'année de naissance, l'année de début de la SEP, l'âge au début de la SEP, le taux annualisé de poussées, ou de la base Census Geodata pour le quantile du niveau socio-économique à l'inclusion. Les scores EDSS observés au cours du suivi ont été catégorisés en trois classes (*mild* = [0 ; 2,5], *moderate* = [3 ; 5,5] et *severe* = [6 ;10]). De plus, le score ARMSS (Age Related Multiple Sclerosis Severity Score) défini par Manouchehrinia et al. (225) a été estimé. Les cinq comorbidités les plus fréquentes dans la SEP (hypertension, hyperlipidémie, diabète, maladie chronique des poumons, troubles de l'humeur et de l'anxiété) ont été identifiées dans les bases médico-administratives à partir d'algorithmes utilisés au Canada (109,226). La prescription de traitements de fond de la maladie a également été incluse dans les paramètres de caractérisation des typologies.

Les coûts des indels ont tous été fixés à 1 et les coûts de substitution ont été basés sur les transitions observées au sein de la base. Les distances ont été normalisées selon la méthode décrite par Yujian et Bo. Une CAH selon le critère de Ward a enfin été effectuée avec une sélection du nombre optimal de clusters basée sur la maximisation de l'ASWw et la minimisation du HC.

VI.2.4.3. Résultats

Descriptif des deux cohortes construites

Au total, 1048 et 3180 patients ont été identifiés respectivement dans les cohortes 1 et 2 (Table 39, Table 40 & Table 41-Annexe G). Concernant la cohorte 1, 76,0% des patients étaient des femmes, correspondant à un sex-ratio F : H de 3,2 : 1, et ils avaient un âge médian de 36,6 ans au démarrage de la SEP pour une durée de suivi médiane de 9,6 ans (étendue : 4,0-13,0) sur une durée maximale de 13

ans. Deux tiers des patients de cette première cohorte ont eu au moins une délivrance d'un traitement de fond de la SEP au cours du suivi. Aucun patient ne présentait un handicap à l'inclusion (score EDSS à l'inclusion faible/*mild*.

Les patients de la cohorte 2 étaient à 73,4% des femmes avec un âge médian de 44,5 ans en 1996. La SEP avait débuté en médiane 11,0 ans avant l'inclusion, ce qui expliquait le fait qu'à l'inclusion 41,9% de ces patients présentait un score EDSS supérieur ou égal à 3 et que de la moitié des patients ayant une SEP rémittente avaient une forme secondairement progressive à la fin du suivi d'une durée médiane de 13,0 ans. Contrairement à la cohorte 1, seuls 31,8% des patients de la seconde cohorte au reçu au moins une délivrance d'un traitement de fond de la SEP au cours du suivi.

Les consommations de soins des deux cohortes étaient similaires dans le recours aux généralistes, spécialistes de médecine interne, spécialistes des symptômes principaux de la SEP. La principale différence entre les deux cohortes provenait du recours aux neurologues, pouvant être due à l'absence de traitement de fond de la cohorte 2.

MCSA appliquée à la cohorte 1

La MCSA a conduit à une typologie en douze classes pour la cohorte 1 (Table 32) dont les caractéristiques sont présentées dans la table Table 42 disponible dans l'Annexe G.

Table 32-Répartition des patients dans les différents groupes suite à l'application de la multichannel sequence analysis sur les individus de la cohorte 1 (N=1 048)

Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
n (%)	221 (21.1%)	63 (6.0%)	59 (5.6%)	200 (19.1%)	105 (10.0%)	58 (5.5%)	88 (8.4%)	44 (4.2%)	83 (7.9%)	32 (3.1%)	81 (7.7%)	14 (1.3%)

Le cluster 1 correspondait à des patients peu traités avec une maladie active caractérisée par de nombreuses poussées résultant en une évolution du handicap au cours du suivi. Les patients du second cluster avaient un fort recours aux spécialistes des symptômes de la SEP cohérent avec le handicap s'installant rapidement après le démarrage de la maladie. Le cluster 3 pouvait s'interpréter comme des patients ayant des comorbidités se traduisant par un important recours aux généralistes. Les patients du cluster 4 effectuaient peu de recours aux différents professionnels de santé et effectuaient des poussées ce qui pourrait correspondre à une SEP peu contrôlée. De même, les patients du cluster 5 avaient un faible recours aux soins, cependant ces patients faisaient peu de poussées et étaient majoritairement traités faisant état d'une maladie contrôlée contrairement au cluster précédent. Les patients du cluster 6 avaient des comorbidités pouvant expliquer leur fort recours aux généralistes. De plus, ils avaient une maladie ayant démarré tardivement ce qui est cohérent avec l'installation rapide du handicap et le

recours important aux spécialistes des symptômes de la SEP. Le cluster 7 rassemblait des patients avec un faible recours aux soins. A l'opposé, les patients du cluster 8 recouraient fortement aux neurologues alors qu'ils effectuaient peu de poussées. Le cluster 9 regroupait des patients ayant des comorbidités, une consommation élevée de soins et présentant une installation rapide du handicap. Les membres du cluster 10 avaient un fort recours aux psychiatres ainsi que des poussées fréquentes au cours de leur suivi correspondant à une maladie active. Les patients regroupés dans le cluster 11 avaient peu de comorbidités avec un peu voire une absence de handicap traduit par un faible recours à l'ensemble des professionnels de santé. Enfin les patients du cluster 12 avaient un fort recours à l'ensemble des professionnels de santé pouvant être expliqué par l'apparition rapide du handicap suite au démarrage de la maladie et donc à une maladie sévère et rapidement évolutive.

MCSA appliquée à la cohorte 2

Les individus de la cohorte 2 ont été répartis en six groupes suite à l'application de la MCSA (Table 33). Les caractéristiques de ces groupes sont présentées dans la Table 43 de l'Annexe G.

Table 33-Répartition des patients dans les différents groupes suite à l'application de la multichannel sequence analysis sur les individus de la cohort 2 (N= 3180)

Cluster	**1**	**2**	**3**	**4**	**5**	**6**
n (%)	121 (3.8%)	1622 (51.0%)	251 (7.9%)	352 (11.1%)	629 (19.8%)	205 (6.4%)

La caractéristique discriminante du cluster 1 de la cohorte 2 était un fort recours à des psychiatres au cours du suivi, peut-être dû à une maladie que l'on pourrait qualifier d'active (fort nombre de poussées et changement de traitements de fond au cours du suivi). Le deuxième cluster majoritaire correspondait plus au parcours de soins classique d'un patient SEP au début de la maladie et sans comorbidités avec une consommation régulière mais modérée auprès de l'ensemble des professionnels impliqués. Les clusters 3 et 6 de cette seconde cohorte pouvaient avoir une interprétation similaire, c'est-à-dire une SEP secondairement progressive entraînant un niveau de handicap élevé et de nombreuses hospitalisations sur la fin du suivi. La principale différence entre les deux clusters résidait dans la présence de comorbidités dans le cluster 6 contrairement au 3. Le cluster 4 correspondait à des patients avec un handicap augmentant au cours du temps se traduisant par un fort recours aux professionnels spécialisés dans les symptômes de la SEP. Les patients du cinquième cluster avaient un recours très important aux généralistes vraisemblablement à cause de comorbidités, mais avec une évolution lente de la SEP, c'est-à-dire peu d'évolution du handicap.

En comparant d'un point de vue global les niveaux de consommation de soins entre les différents clusters, on pouvait noter que les clusters 1 et 2 décrivaient des profils de consommation proches différant par le recours aux psychiatres. Les clusters 3 et 6 se ressemblaient également, la principale différence venant de la présence de comorbidités chez les patients.

Un article pour la valorisation des résultats est actuellement en cours d'écriture pour une soumission prochaine.

VI.3. Discussion

VI.3.1. Discussion des résultats

Le but de ce travail était dans un premier temps de construire les parcours de soins des patients SEP en France en utilisant une méthode adaptée, puis de construire une typologie formée de ces derniers afin de disposer d'informations sur la consommation de soins des individus ayant une SEP en France. Pour ce faire, une méthodologie originale en santé a été utilisée, tout d'abord l'analyse de séquences unidimensionnelle puis dans un deuxième temps la multichannel sequence analysis. Ces dernières ont été appliquées sur trois populations différentes en France avec à chaque fois des ajouts méthodologiques rendant les résultats plus robustes.

Comme attendu, les individus ayant une SEP sont apparus comme ayant une consommation de soins plus importante que celle de la population générale. Ainsi, en ne considérant que le recours aux médecins généralistes et aux neurologues, les nombres médians de consultations étaient respectivement de 9,6 et 0,9 par personne-année dans les données du SNDS sur 2010-2015. A l'inverse, l'OCDE (Organisation de Coopération et de Développement Economique) a reporté un nombre moyen en population générale de 6,5 consultations par personne-année sur la même période pour toutes spécialités confondues (227), soit très inférieur à la population SEP considérée. Cette surconsommation, notamment de médecins généralistes, comparée à la population générale est cohérente avec les résultats de plusieurs études canadiennes (106,107). Par ailleurs, le recours aux soins plus important des femmes ayant une SEP par rapport aux hommes ayant eux aussi la maladie est cohérent avec les études en population générale française (228) et canadienne (107). Les consommations de soins estimées pour l'année 2013 sont également en cohérence avec les résultats avancés par Lefeuvre et al. (105) (5,0 consultations en médiane avec un médecin généraliste *vs* 6,0 consultations dans l'étude française).

L'analyse de séquences unidimensionnelle sur la population de l'EGB a montré cinq groupes de consommation de soins. Cependant, cette méthode n'ayant permis l'utilisation que d'une seule dimension pour décrire le recours aux soins, les groupes obtenus dans la partition finale étaient

principalement déterminés par un état de consommation de soins majoritaire. Ainsi, environ un tiers des patients avaient une forte consommation de soins, probablement due à la SEP elle-même, c'est-à-dire un fort recours aux neurologues et une utilisation élevée des traitements de fond, mais aussi à des comorbidités. A l'inverse, les groupes 2 à 4 correspondaient à des consommations que l'on pouvait qualifiées de faible à intermédiaire. Et enfin, le cluster 1 correspondait aux patients les plus âgés en fin de vie. Cette première classification ne permettait donc pas de différencier de réels profils de consommation de soins, mais plutôt des groupes de niveaux de consommation homogènes.

L'utilisation de la MCSA a permis d'avoir une vision plus précise des différences de recours entre individus. Dans les deux applications qui en ont été faites, une sur l'EGB et une sur le SNDS, une partition en cinq groupes a été obtenue. Ces groupes au vu des consommations de soins et des caractéristiques démographiques et cliniques à disposition ont chacun été rapprochés d'un phénotype de patient SEP. En comparant les partitions de l'EGB et du SNDS, les groupes obtenus sont apparus comme similaires, excepté le groupe 4 obtenu dans l'EGB n'ayant pas de correspondance. En effet, ce cluster avait été vraisemblablement créé du fait de la faible longueur des séquences le constituant. Cinq groupes de recours aux soins ont pu être identifiés. Ainsi, un groupe majoritaire correspondait aux SEP peu actives caractérisé par une faible consommation de soins auprès des neurologues. Deux groupes de patients avec une SEP sous contrôle ont pu être observés, un premier correspondant aux SEP actives et donc traitées à l'hôpital avec des soins de kinésithérapie ponctuels, et un second correspondant aux SEP traitées en ville avec des traitements injectables majoritairement. Ce dernier groupe a par ailleurs souvent recours aux neurologues avec près de 2 consultations par personne-année. Enfin, un quatrième groupe évoquant une SEP progressives avec une forte consommation auprès du domaine paramédical et un important recours aux soins a pu être constaté. Ce dernier peut par ailleurs être subdivisé en deux sous-groupes ayant des caractéristiques similaires se différenciant principalement par le recours aux actes infirmiers. Par ailleurs, dans l'analyse menée sur le SNDS, le groupe ayant la plus forte consommation de soins (groupe 5) était celui ayant les valeurs d'APL les plus élevées. Cette constatation pose la question de la présence ou non d'un effet de surconsommation liée à l'offre de soins. En effet, il a été montré que la distance d'accès aux soins était un frein à la consommation notamment fonction du niveau socio-économique d'un individu (229). De plus, il serait intéressant de mesurer la réelle nécessité de ces soins, notamment les soins infirmiers très importants, et savoir s'ils répondent à un véritable besoin physiologique du patient nécessaire à son bien-être ou seulement à un confort (230). Il est important également de penser au concept opposé, c'est-à-dire une sous-consommation du fait d'un potentiel refus de soins pour des raisons personnelles, d'ordre socio-économique (231), mais aussi par une mauvaise connaissance de son état de santé (232,233). Cependant la majorité de ces paramètres sont inaccessibles via les données utilisées dans ce travail mais pourrait être recueilli dans des études qualitatives.

Dans toutes les analyses menées, la qualité de la partition était faible avec une valeur d'ASWw inférieure à 0,20, ce qui pourrait suggérer une absence de structure. Cependant cette qualité est à relativiser compte-tenu du grand nombre de séquences considérées, notamment dans le SNDS où 35 000 séquences étaient utilisées pour la classification, et de la forte variabilité intra-séquence. Cette variabilité peut s'expliquer par le fait du caractère quasi-aléatoire de la consommation de soins, lié à des facteurs de la maladie, comme son ancienneté, son activité et son avancée, et les effets indésirables des traitements mais aussi à des facteurs extérieurs tels que les infections virales ou bactériologiques par exemple. Cette hétérogénéité du parcours de soins rend donc plus difficile leur comparaison que lors d'études sur d'autres thématiques comme l'emploi ou la situation maritale où une certaine stabilité peut être constatée, entraînant alors des partitions de meilleure qualité.

VI.3.2. Forces et limites

VI.3.2.1. Forces et limites des méthodes d'analyse de séquences

Peu utilisées en épidémiologie et plus largement dans le domaine de la santé, les méthodes d'analyse de séquences se sont révélées être parfaitement adaptées pour l'étude des parcours de soins. Leur principal avantage est la prise en compte du parcours de soins global comme étant l'unité statistique d'intérêt (234). Cette approche holistique permet une comparaison simultanée entre individus de la quantité de consommation de soins mais de son évolution au cours du temps. De plus, grâce aux développements récents de ces méthodes, plusieurs types de recours auprès de différents professionnels de santé peuvent à présent être considérés dans le cadre de la MCSA. L'analyse de séquences s'oppose ainsi aux méthodes plus classiquement utilisées en épidémiologie, considérant chacun des évènements indépendamment les uns des autres ou ne permettant pas une analyse simultanée de la dimension temporelle et de consommations soins.

L'apport de la MCSA par rapport à l'analyse de séquences unidimensionnelle a été constaté au travers des deux exemples sur la population de l'EGB. En effet, l'analyse de séquences unidimensionnelle avait permis de donner un premier aperçu des variations de consommation de soins au sein de la population SEP. Cependant le niveau macroscopique de ces informations limitait l'identification de phénotypes SEP pouvant correspondre à ces groupes. A l'opposé, l'utilisation de la MCSA a permis l'identification de différences à un niveau plus fin, notamment concernant les recours aux soins paramédicaux importants dans le traitement de la SEP, permettant ainsi une meilleure compréhension et l'identification de phénotypes cliniques pertinents.

Par ailleurs, la procédure d'application des deux méthodes (unidimensionnelle et multidimensionnelle) telle que décrite dans cette section est simple à mettre en place et laisse toute la liberté à l'utilisateur dans le choix de l'estimation de la mesure de dissimilarité, par la fixation des coûts notamment (234).

Cette flexibilité permet ainsi une adaptation à la majorité des questions de recherche en épidémiologie, qu'elles portent sur un aspect longitudinal ou sur la réalisation d'un évènement de façon concomitante. Dans notre cas, l'étude portait sur les parcours de soins et donc la succession des consommations, mais il est possible d'envisager une étude ayant pour point de départ le diagnostic de la SEP et ayant pour but de vérifier le respect d'une prise en charge théorique nécessitant alors la recherche d'évènements ayant lieu à un temps défini depuis le diagnostic. Ainsi, avec une simple variation des coûts des indels, une réponse peut être apportée à cette dernière. De plus, l'application sur plusieurs échantillons a pu permettre d'identifier des choix méthodologiques adaptés à la problématique des parcours de soins. Ainsi, dans la phase de construction des séquences, il a été fait le choix d'un alphabet synthétique basé sur des catégories de consommation de soins afin de diminuer la variabilité intra-séquence déjà très forte et ainsi permettre d'augmenter la qualité de la partition finale. Le fait de tronquer à droite les séquences en cas de décès a également permis d'éviter la création d'un cluster dédié à ces patients, même s'il est apparu qu'un tel cluster existait dans la typologie obtenue avec la MCSA appliquée sur l'EGB. C'est pourquoi, lors de l'application sur le SNDS, il a été fait le choix dans un premier temps de ne garder que des parcours ayant au moins huit semestres de consommation de soins consécutifs et dans un deuxième temps de normaliser la mesure de dissimilarité afin de limiter les différences résiduelles dues à la longueur des parcours. Cette deuxième étape a été réalisée en utilisant la méthode Yujian et Bo (211) permettant de garantir l'inégalité triangulaire, contrairement aux autres méthodes décrites dans la littérature (169). Concernant la mesure de dissimilarité, le maximum de données à disposition a été pris en compte dès que possible lors du calcul des coûts de substitution afin de bénéficier de la meilleure estimation de ces derniers et d'envisager l'ensemble des transitions possibles.

Néanmoins, plusieurs limites inhérentes à la méthodologie utilisée peuvent être notées. Tout d'abord, des limites logicielles peuvent être mises en évidence. Ainsi, la méthode implémentée sous R ne permet pas l'étude de plus de 45 000 séquences, notamment pour le calcul de la matrice de dissimilarité, d'où notre limitation à un échantillon de 35 000 séquences. Toutefois, il est très rare d'avoir à disposition une masse de données aussi importante en analyse de séquences (134,203). Ce problème est donc plus lié à une conséquence directe de l'utilisation des bases de données médico-administratives que de la méthode en elle-même puisqu'il suffirait d'optimiser le code des fonctions pour pouvoir permettre la prise en compte d'un nombre de séquences plus important.

Par ailleurs, des choix méthodologiques peuvent être discutables. En effet, les séquences considérées dans les différentes analyses sont assez courtes (7 états en analyse de séquences unidimensionnelle sur l'EGB, 14 avec la MCSA sur l'EGB et 12 sur le SNDS) et sont en-deçà des recommandations effectuées dans la littérature (134). Cependant, ce choix est lié aux recours aux soins parfois faibles

aux différents professionnels de santé, notamment au neurologue. Il est également en rapport avec une volonté d'éviter la succession d'états d'absence de consommation de soins et aussi de n'avoir que deux états pour certaines dimensions importantes, c'est-à-dire une absence de recours aux soins *vs* un recours. En outre, d'autres mesures de dissimilarités auraient pu être choisies à la place de l'optimal matching. Néanmoins, cette dernière est apparue comme étant la plus adaptée à notre problématique, suite au développement de l'expertise sur la méthode, permettant de ne pas seulement considérer ni la plus longue sous-séquence commune ni le nombre d'états identiques, mais au contraire de permettre un entre-deux favorisant néanmoins légèrement la plus longue sous-séquence commune.

D'autres méthodologies statistiques auraient pu être utilisées dans le cadre de cette étude car permettant la prise en compte de la dimension temporelle du parcours de soins et conduisant à une classification elles-aussi. Ces dernières sont les méthodes d'analyse à classes latentes (LCA - *latent class analysis*) (235) et les méthodes de k-means appliquées aux trajectoires (kml) (236). La première repose sur l'hypothèse qu'il existe des paramètres non mesurés communs à un groupe d'individus (la classe latente) permettant d'expliquer la succession de valeurs observées. La seconde est une extension des k-means à des données longitudinales permettant d'obtenir une classification des trajectoires, pouvant être uni- ou multidimensionnelle. A notre connaissance, trois comparaisons entre l'analyse de séquences et les méthodes de LCA existent dans la littérature (185,237,238), ces dernières ayant montré des résultats de clustering similaires selon les deux méthodes. De plus, ces études ont montré que l'analyse de séquences était plus facile à utiliser et à exécuter que la LCA et que contrairement à cette dernière, elle ne nécessite pas la spécification d'un modèle *a priori*. En effet, cette dernière a besoin d'un modèle statistique afin de pouvoir trouver les classes latentes nécessitant donc le choix d'un type de régression. En addition, une hypothèse importante de la LCA est l'indépendance locale conditionnellement à la classe (185), c'est-à-dire qu'un état à un instant t doit être indépendant des états le précédant. Or, cette hypothèse est difficile à assumer dans le cas d'un parcours de soins puisque même si une partie importante du recours aux soins est liée à l'activité de la maladie, une part non négligeable est liée au patient lui-même entraînant une corrélation dans les valeurs. Afin de s'affranchir de cette limite, Mikolai et Lyons-Amos ont proposé l'utilisation des *latent class growth models* (238) prenant en compte les mesures répétées et donc l'ordre dans les évènements (239).

Les modèles kml, quant à eux, permettent dans une extension de la méthode une étude multidimensionnelle (236). Ces modèles ont pour but de synthétiser plusieurs trajectoires construites à partir de variables continues corrélées en une seule variable nominale formant des groupes de trajectoires. Ces méthodes apparaissent donc comme étant parfaitement adaptées à la question de recherche considérée dans ce travail. Elles permettraient notamment de s'affranchir de la classification du recours aux soins conduisant à une perte d'information. Néanmoins, le fait de devoir considérer le

recours aux soins comme variable continue pourrait conduire à une difficulté importante d'identification de groupes homogènes, du fait de l'importante variabilité inter- et intra-parcours. Par ailleurs, la méthode des k-means sur laquelle repose kml s'est révélée être celle conduisant au plus grand nombre de mauvaises classifications dans les simulations conduites par Dlouhy et al. dans leur revue de la littérature sur l'utilisation de l'optimal matching (134). De plus, elle se révèle comme étant imparfaite car nécessitant plusieurs applications sur un même jeu de données afin de prendre en compte la possibilité d'un maximum local et d'obtenir ainsi une classification stable (236).

Néanmoins, un travail sur la comparaison de l'analyse de séquences à ces deux méthodes est nécessaire afin de comparer les résultats et hypothèses sous-jacentes, notamment dans le domaine de l'étude des parcours de soins.

VI.3.2.2. Forces et limites des données de consommation de soins

Plusieurs forces sont également liées aux hypothèses faites lors de la construction des indicateurs de recours aux soins. Un premier travail a été réalisé afin de limiter au maximum les doublons présents dans la base et ainsi une surestimation des indicateurs calculés. Ainsi, seuls les remboursements effectués par l'Assurance Maladie (régularisations comprises) ont été considérés. Les majorations des différentes consultations ont, quant à elles, été supprimées, toutes étant complémentaires d'une prestation codant pour la consultation. De plus, les indicateurs ont été construits au niveau mensuel afin de pouvoir repérer plus facilement des erreurs de remboursement éventuelles conduisant à des indicateurs négatifs, qui ont été forcés à 0. En effet, une estimation au niveau semestriel d'emblée aurait pu conduire à des compensations entre valeurs de différents mois et ainsi introduire une sous-estimation de la consommation de soins. Afin d'estimer au mieux cette consommation, les séjours hospitaliers se déroulant sur deux mois consécutifs ont été séparés et le nombre total de jours d'hospitalisation a été majoré par le nombre de jours mensuel correspondant. De manière analogue, les actes infirmiers et de kinésithérapie ont été lissés, lorsqu'ils étaient renseignés pour une période de temps définie. En outre, des incohérences entre le codage de la NGAP et des spécialités médicales ont pu être constatées pour certaines consultations (consultations cardiologiques codées pour un neurologue par exemple). Dans ce cas, la priorité a été donnée à la spécialité médicale, en supposant que la cotation NGAP différait du fait d'une surfacturation et afin d'assurer une continuité entre les consultations externes et celles effectuées en libéral. Toutefois, ces incohérences sont restées rares et ont donc peu d'impact sur nos résultats. Enfin, l'avantage majeur déjà évoqué dans les sections précédentes est l'exhaustivité des consommations de soins, entraînant de ce fait une quasi-absence de données manquantes.

A l'inverse, de nombreuses autres données de recours aux soins auraient pu être prises en compte. Les consultations avec un spécialiste MPR ont été prises en compte dans les deux applications sur l'EGB, mais pas dans l'application sur le SNDS. En effet, cette spécialité étant principalement exercée dans le milieu hospitalier SSR, seulement un faible nombre de consultations avaient pu être repérées en libéral. Ayant peu d'expertise au moment des analyses sur le PMSI SSR, il serait à présent intéressant de repérer cette prise en charge importante dans la SEP notamment du point de vue du handicap dans les consultations externes du SSR.

De façon similaire, il réside une certaine incertitude sur les consultations externes effectuées par des généralistes du PMSI MCO. Ainsi, même si cette spécialité n'existe que très peu dans les établissements de santé, une quantité importante d'actes et de consultations effectuées par ces professionnels a pu être constatée dans les ACE du MCO, car il s'agit d'une spécialité par défaut dans le PMSI. Ainsi, il est très fréquent qu'une consultation externe soit codée par défaut comme ayant eu lieu avec un médecin généraliste alors qu'elle a en réalité eu lieu avec un neurologue ou un cardiologue sans possibilité de distinction. Afin de prendre en compte cette particularité, seules les consultations externes identifiées comme ayant eu lieu avec un neurologue hospitalier ont été considérées dans la construction des parcours de soins, mais il faudrait réaliser une analyse de sensibilité en prenant en compte ces consultations. Par ailleurs, le PMSI permet de repérer quand un patient est passé par les urgences sans connaître les soins qui y ont été effectués. Néanmoins, le nombre de passage aux urgences d'un patient aurait pu permettre d'approximer un recours aux soins non programmé, comme lors d'une poussée par exemple.

Malgré les limites inhérentes à la méthodologie utilisée et aux données à disposition, ce travail a permis de quantifier pour la première fois en France le recours aux soins des personnes ayant une SEP. Plusieurs groupes homogènes de consommation de soins ont pu être identifiés et rapprochés de phénotypes cliniques, même s'ils restent à être rapprocher de la pratique clinique pour une meilleure caractérisation.

VII. Discussion générale

VII.1. Résultats principaux

Au travers de trois axes de recherche, cette thèse avait pour but de fournir des données objectives sur la population des personnes ayant une SEP en France, en particulier concernant leur prise en charge, notamment thérapeutique, et leurs consommations de soins. Afin d'accéder à ces informations, les bases médico-administratives de l'Assurance Maladie ont été utilisées. Ces dernières ont permis une (quasi-)exhaustivité tant au niveau de la population d'étude qu'au niveau des données de recours aux soins.

Ainsi, un descriptif complet des principales caractéristiques de cette population a pu être réalisé permettant une mise à jour des données françaises par rapport aux études les plus récentes effectuées par l'Assurance Maladie (40,105). Ces données viennent également confirmer certains chiffres avancés par les neurologues résultant d'observations faites durant leur pratique quotidienne, notamment concernant la part de patients traités en France mais aussi sur la prévalence du statut ALD par exemple. Ainsi, il était communément admis que la moitié des patients SEP bénéficiaient d'un traitement de fond et que près de 90% d'entre eux étaient en ALD pour la maladie, sans corroboration possible au niveau national jusqu'aux données résultant de cette thèse. Par ailleurs, cette dernière a permis de réaliser un état des lieux de l'utilisation des traitements de fond de la SEP en France et d'introduire de l'ensemble des professionnels impliqués de la prise en charge de nouvelles formes de représentations graphiques de cette évolution. L'impact de l'arrivée des traitements oraux dans la prise en charge a également pu être montré, ces résultats étant en cohérence avec ceux obtenus dans des cohortes cliniques comme celle de l'OFSEP. De plus, grâce à l'exhaustivité du SNDS, le biais de recrutement a pu être confirmé sur les modalités de recrutement des cohortes cliniques. En effet, les individus suivis dans les centres experts sont souvent plus actifs avec un handicap évoluant plus rapidement et nécessitant donc une prise en charge particulière ne reflétant pas forcément l'ensemble des SEP prises en charge, comme cela a été mis en évidence dans une étude de Debouverie et al. comparant les méthodes *population-based* et *hospital-based* (240).

Un autre intérêt des bases médico-administratives a été démontré au travers du travail exploratoire sur l'estimation du handicap moteur à partir des remboursements d'aides techniques à la marche. Ce dernier a permis un enrichissement des données avec un indicateur de quantification du niveau de handicap moteur. Cet indicateur ne permet pas d'atteindre un pouvoir fin de discrimination pour les

niveaux faibles de handicap, mais il permet une bonne identification des handicaps moteurs moyens et sévères. Con utilisation peut donc être intéressante dans d'autres études épidémiologiques sur d'autres pathologies. Ce domaine de la mesure du handicap dans le SNDS est par ailleurs aujourd'hui en plein essor avec un travail de l'Irdes sur la faisabilité de l'identification des situations de handicap à partir des données médico-administratives (projet FISH coordonné par Catherine Sermet). Nous avons été sollicités au lancement de ce projet, qui s'est orienté en premier temps sur l'estimation du handicap visuel avec un indicateur en cours d'élaboration. Le travail réalisé durant cette thèse pourra servir de base au développement d'un indicateur de handicap moteur dans un futur proche.

Il s'agit enfin du premier travail en France sur la consommation de soins dans la SEP à l'échelle nationale et plus particulièrement sur les parcours de soins, et le premier à utiliser la MCSA dans le domaine de la santé pour ce type d'étude. En l'absence de recommandations concernant le parcours de soins des individus ayant la SEP, l'établissement d'une typologie des parcours observés en France est une première étape de recherche dans ce domaine. A l'inverse, si ce travail avait porté sur une maladie avec un parcours « recommandé », telle que l'insuffisance rénale chronique par exemple, il aurait été intéressant de comparer les parcours observés avec le parcours théorique afin de mettre en évidence de potentiels écarts par rapport à ces recommandations et d'en identifier les déterminants. Ici, il s'agissait plutôt de faire un état des lieux et de donner des éléments chiffrés décrivant précisément les consommations de soins aux niveaux annuel et semestriel, en espérant qu'ils puissent contribuer à l'élaboration de recommandations. Au final, cinq groupes homogènes de consommation de soins ont pu être mis en évidence, chacun pouvant être mis plus ou moins en correspondance avec un phénotype clinique de la SEP grâce aux caractéristiques des individus le composant. De plus, au travers des différentes typologies créées et des questions méthodologiques soulevées, ce travail a permis d'identifier les éléments clés permettant l'application aux données de santé des méthodes d'analyse de séquences en santé, et plus seulement aux sciences sociales. Il reste toutefois un travail de validation de cette typologie à effectuer en confrontant ces résultats au point de vue de différents professionnels, qu'ils soient décideurs (HAS) ou acteurs (neurologues, réseaux, infirmiers, médecins généralistes, patients, aidants, …).

Ce besoin de validation pose la question de l'appropriation et de l'utilisation des résultats de ce type d'étude quantitative, en dehors de la sphère académique. Un travail sur ce thème avec une application spécifique à la SEP a été conduit dans le cadre du Module Interprofessionnel de Santé Publique (MIP) organisé par l'EHESP par un groupe de sept élèves fonctionnaires de l'école (241). Pour ce faire, les résultats de la typologie de parcours obtenus dans l'EGB suite à l'application de la MCSA ont été présentés au cours d'entretiens semi-directifs à différents acteurs et décideurs de la prise en charge de la SEP en France. Ces entretiens ont montré qu'il existe un réel intérêt pour ce type de résultats

permettant d'avoir une vision globale de la situation en France et de contribuer au débat, mais aussi un intérêt pour décliner les données au niveau régional, qui permettrait d'évaluer l'organisation des soins, peut-être leur efficience et l'impact des réseaux SEP. Les neurologues ont plutôt validé les interprétations de nos groupes, mais ces derniers ont également eu plusieurs interrogations sur la méthodologie utilisée et ont remarqué le manque de données cliniques et médico-sociales permettant de valider pleinement les conclusions avancées (241). Il faut donc rester prudent dans l'interprétation. Ces difficultés d'appropriation des résultats ont également pu être constatées auprès des membres de la HAS rencontrés, qui ont encouragé le couplage de ces résultats avec des apports de la littérature et des experts, afin qu'ils ne reposent pas uniquement sur la statistique. Ces personnes étaient d'ailleurs très intéressées par nos travaux, le parcours de soins dans la SEP étant à leur agenda, et il donc probable qu'elles reviennent vers notre équipe pour en rediscuter. Cela révèle également l'importance des enquêtes qualitatives permettant l'étude des pratiques des professionnels de santé et de mesurer les retentissements psychologiques et sociétaux liés à la maladie par exemple. Par ailleurs, l'importance de confronter les résultats à la clinique a été soulignée, notamment lorsqu'un neurologue conseille de retourner au dossier clinique d'un échantillon de patients. Cependant, un des points revenant le plus dans les entretiens est l'impossibilité d'identifier un unique « parcours de soins type » dans la SEP du fait de l'évolution imprévisible de la maladie et du grand nombre de professionnels de santé impliqués dans la prise en charge. Ce point de variation inter-individuelle a été souligné tout au long de ce travail notamment par la faible qualité des partitions et la variabilité des séquences thérapeutiques, et permet donc d'expliquer la difficulté dans la définition de recommandations de prise en charge pour la SEP.

VII.2. Intérêt des données médico-administratives

Tout d'abord, les bases de données médico-administratives permettent un accès à une population quasi-exhaustive. On peut ainsi aisément supposer que le nombre d'individus ayant une SEP en France et n'étant pas affiliés à un des régimes couverts par le SNDS est minime, d'où un manque d'exhaustivité très limité (un peu moins de 3%) et ne laissant pas présumer d'un biais de sélection dans ce travail. Il n'y a pas de raison de penser que individus couverts par un des régimes absents du SNDS aient un recours aux soins et des caractéristiques différentes. Par ailleurs, la longue période d'inclusion sur six années consécutives permet de repérer à la fois les cas prévalents, y compris ceux ayant un faible recours aux soins, mais aussi les cas incidents. Un algorithme permettant d'identifier ces cas incidents pourrait par ailleurs être développé, comme c'est déjà le cas au Canada (242), afin de pouvoir mener une étude d'incidence et la date de « diagnostic » (alors approximée par la première mention de SEP dans la prise en charge) permettrait d'avoir une date commune pour une analyse de séquences potentielle sur la prise en charge au début de la maladie. Par ailleurs, l'identification des patients a été réalisée sur la base de quatre critères complémentaires dans le repérage des différentes

formes cliniques de SEP. La population finale identifiée est alors plus complète que celle se trouvant dans les cohortes cliniques françaises se basant sur un recrutement auprès des neurologues (principalement en CHU ou en centre expert). En outre, les résultats concernant la population d'étude sont cohérents avec les études de prévalence dans la littérature en prenant en compte l'augmentation du nombre de cas liée à l'incidence de la maladie. Ainsi, Foulon et al. avançaient une prévalence au niveau national de 151,2 cas pour 100 000 au 31 décembre 2012 (40), qui est en accord avec notre mesure de 162,2 cas pour 100 000 au 31 décembre 2015 (l'incidence de la SEP étant supérieure à la mortalité conduisant donc à une prévalence en augmentation). De plus, un travail spécifique a été réalisé sur l'identification des individus ayant une SEP pédiatrique avec un diagnostic précoce (âge à l'identification inférieur à 8 ans) afin de consolider la population d'étude. Le lien de filiation n'étant pas à disposition dans le SNDS, il n'était pas possible d'être certain que c'était bien l'enfant qui avait une SEP et non l'un de ses parents, notamment lors de l'identification via le critère portant sur les traitements et l'identification dans le PMSI MCO (la question aurait été similaire avec le repérage grâce au PMSI SSR, mais aucun enfant n'a été identifié qu'avec ce critère). En effet, un enfant n'ayant pas de carte vitale et étant donc sur la carte vitale de ses parents, nous avons pensé qu'une erreur à la pharmacie pouvait attribuer à un enfant le remboursement d'un traitement spécifique de la SEP alors que le bénéficiaire est en réalité un des deux parents. De manière analogue, une suspicion de diagnostic ou une erreur de codage à l'hôpital pourrait conduire à de faux positifs. C'est pourquoi les critères d'identification pour les enfants avaient été modifiés à au moins deux délivrances d'un traitement spécifique à la SEP et à une identification croisée avec au moins un autre critère pour ceux étant identifiés via le PMSI MCO.

Cependant, même si l'algorithme d'identification des individus ayant une SEP mêle quatre critères complémentaires, le risque d'inclure des faux positifs reste néanmoins plausible. En effet, un quart de la population n'a été identifiée qu'avec un seul critère, ce qui peut entraîner des doutes sur leur réel statut de malade. Parmi ces patients, ceux identifiés par le biais de l'ALD SEP et d'un traitement de fond spécifique peuvent être considérés comme ayant réellement une SEP, puisque dans le premier cas le diagnostic est vérifié par un médecin de l'Assurance Maladie lors de l'attribution du statut et dans le second cas les traitements considérés dans l'algorithme sont spécifiques à la maladie. Le diagnostic est également plus sûr pour les 4 066 patients identifiés avec un diagnostic SEP dans au moins deux séjours hospitaliers différents du fait de la redondance du diagnostic. Le doute concerne donc principalement les individus identifiés uniquement grâce aux deux PMSI (soit 11 652 patients (10,3%)) et plus particulièrement les 7 586 (6,7% de la population totale) n'ayant qu'un seul séjour permettant leur identification, pouvant correspondre à un faux diagnostic ou à une suspicion de SEP. Il serait intéressant de réaliser une analyse de sensibilité afin d'étudier l'impact de ces patients dont le diagnostic de SEP peut être mis en doute sur les résultats présentés précédemment, notamment

concernant les parcours de soins, ce qui reviendrait à appliquer les mêmes règles que lors de la sélection des individus présentant une identification pédiatrique précoce de la SEP.

Un autre atout de l'utilisation des bases de données médico-administratives est l'exhaustivité concernant les données de consommation de soins, limitant ainsi le biais de mémoire, souvent présent dans les études observationnelles. En effet, le SNDS étant une base comptable liée aux remboursements, les données sont recueillies de manière systématique et prospective grâce à la carte vitale et des feuilles de soins, entraînant ainsi un délai réduit entre la réalisation de l'acte et son renseignement dans la base. Une alternative consistant à réaliser un recueil rétrospectif direct des données auprès des patients serait compliquée à mettre en œuvre, exposerait au biais de mémoire. Une autre méthode envisageable aurait pu être un recueil prospectif de la part du patient grâce à un cahier dans lequel il aurait rentré l'ensemble de ses consommations, présentant alors un risque d'oubli et/ou de lassitude de la part de ce dernier. Cependant, ces méthodes de recueil se font nécessairement sur des effectifs réduits de quelques dizaines de patients et à des coûts très élevés. Il est toutefois indéniable que le recueil de données auprès du patient permet l'obtention d'un plus grand nombre d'informations, notamment cliniques (forme de la maladie, score EDSS par exemple) lesquelles sont absentes du SNDS, et qui peuvent se révéler utiles voire très utiles selon la question de recherche. Ce travail ayant visé à étudier la consommation de soins et l'ensemble des évènements d'intérêt étant remboursés, tout comme l'intégralité des traitements de la SEP (sauf la mitoxantrone, l'alemtuzumab et le cyclophosphamide remboursés mais non identifiables dans le SNDS), l'utilisation des données médico-administratives nous semble pertinente avec plus d'avantages que d'inconvénients recensés.

Néanmoins, l'utilisation des données médico-administratives nécessite la prise en compte d'une importante volumétrie et de données complexes. En effet, comme indiqué ci-dessus, ces dernières ne sont pas dédiées à une utilisation en épidémiologie mais ont un but original purement comptable, ce qui induit un travail de transformation des variables afin d'obtenir les paramètres d'intérêt. De plus, l'importante volumétrie n'a pas permis d'estimer les mêmes paramètres sur l'ensemble des populations d'étude, ne permettant pas une comparaison directe des résultats entre eux notamment concernant l'étude des parcours de soins. En effet, cette masse de données (196 221 798 prestations, et respectivement 1 008 940 et 3 601 892 séjours hospitaliers et ACE en MCO en 2010-2015 dans notre population totale dans le SNDS (N=112 745)) nécessite une approche adaptée et un codage de haut niveau afin de pouvoir obtenir des résultats en un temps de calcul raisonnable de l'ordre de quelques minutes voire secondes. Afin d'atteindre cette vitesse de traitement, la librairie *data.table* adaptée au *big data* (243) a été utilisée accompagnée d'une optimisation du code. Afin de donner un ordre de grandeur, le calcul du nombre mensuel de consultations avec un médecin généraliste et des données relatives à ces consultations (isolement des dates d'exécution et de prescription par exemple) pour

l'ensemble des 112 745 individus pour la période 2010-2015 prenait 10 minutes après optimisation (19 694 369 prestations de consultations à traiter) alors qu'il fallait 25 minutes sur l'échantillon de 1 000 patients de l'EGB au début de la thèse pour le même travail à partir du format de base *data frame*. D'où là encore, le développement d'une expertise et de compétences dans le management des données au fil de ce travail de thèse.

Par ailleurs, il est préconisé par l'Assurance Maladie de travailler avec une période de latence de 27 mois entre la fin de la période d'étude et l'extraction des données, afin de respecter le délai maximum pour le remboursement d'un acte de soins (un individu a 27 mois pour envoyer une feuille de soins et se faire rembourser). Or, ce délai n'a pas été respecté dans notre analyse sur les données du SNDS. Cependant, nous avons considéré que le délai de plus de 18 mois entre la fin de la période d'étude (2015 pour rappel) et la date d'extraction au mois d'octobre 2017 était suffisant du fait du statut d'ALD de la maladie. En effet, ce statut entraîne un règlement de la majorité des soins par tiers payant grâce à la carte vitale, le patient n'ayant aucun frais à avancer, d'où leur traitement rapide par l'Assurance Maladie et donc leur entrée plus rapide dans le SNDS. De plus, la période d'étude avait été définie afin de pouvoir étudier l'arrivée des nouveaux traitements oraux de la SEP, notamment le teriflunomide et le diméthyl fumarate, même si pour ce dernier il aurait fallu bénéficier d'une extraction de l'année 2016. Une étude du taux de données manquantes de l'extraction du fait du non-respect de la période de latence serait intéressante afin de quantifier la part d'information perdue, même si cette dernière doit être fortement négligeable. En effet, une rapide analyse des données de prestations a révélé que 88,0% des prestations effectuées au mois de décembre 2015 (3 515 819 prestations) avaient été remboursées en février 2016 et se chiffre s'élevait à 98,7% au mois de juin 2016. Sur un recul plus long en considérant les prestations exécutées en décembre 2011 (2 532 677 prestations), 83,0% étaient remboursées au bout de deux mois, 96,9% au bout de six mois et 98,7% au bout d'un an. En complément nous avons identifié une étude de Santé publique France de 2017 sur la qualité des données du SNIIRAM, qui avait montré une complétude des données de 98% à 2 mois qui avoisinait les 100% à 6 mois, sur le remboursement d'un antidiarrhéique dans le cas d'un épisode aigu de gastro-entérite (244). Ce médicament n'ayant pas de lien avec une ALD, on peut supposer que la complétude des actes liés à des ALD est obtenue dans les mêmes ordres de grandeur si ce n'est plus rapide.

Enfin, l'utilisation du SNDS a conduit à un questionnement méthodologique lors de cette thèse portant sur l'intérêt des méthodes statistiques inférentielles classiques. En effet, les méthodes actuellement à disposition en statistique sont utilisées afin de réaliser de l'inférence sur une population plus large. Qu'en est-il lorsque la population est exhaustive entraînant alors une puissance statistique suffisante pour détecter des violations d'hypothèses conditionnant l'usage d'un modèle (245) ou identifier la

présence de différences significative sans l'utilisation de tests statistiques (246). Ces nouveaux challenges mènent actuellement l'épidémiologie dans une crise inférentielle (246) avec la remise en cause de l'intérêt des tests statistiques usuels et celle de la suprématie de la p-valeur. On assiste alors à l'essor de nouveaux modes de pensée et d'action, créant un nouveau courant de la statistique nommé *data science*. Ainsi, une étude sur un échantillon aléatoire de la population exhaustive peut permettre l'utilisation des techniques usuelles et de réaliser ainsi de l'inférence. A l'inverse, une étude purement descriptive de la population exhaustive peut être suffisante, la fluctuation d'échantillonnage étant quasi-absente. De plus l'utilisation d'indices de parcimonie doit être prônée dans le cadre du *big* data, tout comme des techniques adaptées (246), telles que le LASSO (Least Absolute Shrinkage and Selection Operator) (247) par exemple. Dans le cadre de ce questionnement, nous avons opté pour la réalisation d'un échantillonnage aléatoire sur le SNDS afin de permettre une forme d'inférence à la population totale des individus ayant une SEP en France. Ce choix a également permis de s'affranchir du nombre maximal de séquences pouvant être traitées avec *TraMineR*. Néanmoins, on peut se poser la question de la validité d'un échantillon aléatoire représentant un peu plus de la moitié de la population d'étude (35 000 parmi les 64 373 individus soit 54,4%). Il serait ainsi intéressant de réaliser des analyses de sensibilité faisant varier la taille de l'échantillon et d'évaluer l'impact sur la partition finale. Par ailleurs, une autre conséquence de notre réflexion est la non-présentation systématique de p-valeur dans les résultats portant sur le SNDS. En effet, lorsqu'elles étaient présentées, comme lors de la phase de caractérisation des groupes obtenus avec la multichannel sequence analysis sur le SNDS, une différence infime entre deux groupes entraînait alors un test significatif.

VII.3. Réflexion sur les hypothèses formulées

Concernant à présent les indicateurs permettant de caractériser la typologie finale de parcours de soins, plusieurs choix méthodologiques ont été réalisés. Il a été fait le choix de ne pas considérer la CMU-C, alors qu'il s'agit d'une meilleure approximation du niveau socio-économique d'un individu que la CMU ou l'AME. En effet, notre population ayant un âge médian de de 46 ans en 2010 avec un premier quartile situé à 57 ans, la CMU-C ne pouvait plus être attribuée à environ un quart de notre population au cours du suivi, du fait qu'un individu âgé de plus de 60 ans ne peut plus bénéficier de ce statut, créant ainsi de nombreuses valeurs manquantes. De plus, le statut ALD prévalant sur la CMU-C au moment d'un remboursement et près de 85,5% de notre population bénéficiant d'une ALD SEP sur 2010-2015, le nombre d'individus identifiés comme bénéficiaire aurait été fortement sous-estimé. C'est pourquoi dans notre travail, la CMU moins sujette à de telles conditions a été préférée à la CMU-C. Toutefois, dans d'autres situations avec une population n'étant pas fortement en ALD et ayant moins de 60 ans, la CMU-C serait sûrement plus informative que la CMU et donc probablement

à favoriser. Par ailleurs, afin de quantifier l'offre de soins à disposition du lieu de résidence des individus, il a été fait le choix d'utiliser comme indicateur contextuel l'APL développé par l'Irdes et la DREES (159–161). Ce paramètre a été préféré à la densité médicale plus classique du fait qu'il prend en compte à la fois l'offre (densité médicale), la demande (activité réelle des différents praticiens), mais aussi le délai d'accès aux différents spécialistes. De plus, développé à une échelle plus fine que la densité médicale, car disponible au niveau communal, il permet de prendre en compte les disparités territoriales au sein d'un même département. L'APL est un indicateur communément utilisé dans les agences régionales de santé françaises pour identifier les zones à faible niveau d'accès aux soins et orienter leur action (248,249), d'où notre choix de le considérer comme variable descriptive, car pouvant s'intégrer de manière plus adéquate à la définition ou l'évaluation de parcours.

Cependant, certaines variables utilisées dans la caractérisation des typologies peuvent être discutables. Tout d'abord, un doute peut subsister concernant le lieu de résidence des individus pour lesquels il a été imputé grâce aux données du PMSI. Ce doute provient du fait du passage du code géographique PMSI au code Insee via le code postal, conditionnant les valeurs d'APL et de FDep. La transformation en code postal a supposé pour 37 patients un lieu de résidence par défaut pour les petites zones géographiques. Toutefois, il est acceptable de penser que ces communes faisant partie du même code géographique PMSI partagent de fortes similitudes du fait de leur proximité géographique et donc qu'elles aient des valeurs d'APL et de FDep proches voire similaires. De plus, la faible proportion d'individus concernés (pour rappel, 37 personnes soit 0,03% des patients) ne permettrait pas de remettre en cause les résultats. Par ailleurs, concernant ces variables socio-économiques contextuelles, que sont l'APL et le FDep, elles sont estimées au niveau de la commune de résidence d'un patient et non au niveau de l'individu. Par conséquent, elles peuvent ne pas refléter ni la réelle accessibilité aux soins d'un patient ni sa réelle situation sociale. Cependant, ces variables étant pertinentes et utilisées dans la littérature (250–253), elles ont été prises en compte comme paramètre explicatif. De plus, même si elles sont contextuelles, il a été démontré que les valeurs socio-économiques du lieu de résidence quantifiaient de la même façon les inégalités sociales que le niveau socio-économique d'un individu (254). A l'opposé, la CMU était un réel reflet du niveau socio-économique d'un patient. L'indicatrice portant sur ce statut était positive quand un patient avait bénéficié de ce dernier au moins une fois au cours de la période d'étude. Or, la CMU est un dispositif devant être renouvelé annuellement. Donc elle aurait mérité une analyse plus fine, comme par exemple une indicatrice annuelle caractérisant une adhésion au dispositif. Néanmoins, l'hypothèse choisie dans ce travail nous a paru raisonnable, partant du fait qu'un individu bénéficiant de la CMU durant un an ne verrait pas son niveau socio-économique fluctuer de manière importante au cours de la période de suivi de six années. Un raisonnement analogue a par ailleurs eu lieu concernant l'indicatrice portant sur l'AME. De plus, même si une mesure annuelle aurait pu approcher au mieux l'évolution du niveau

économique d'un individu, le nombre de variables introduites aurait alors été important et aurait potentiellement pu inutilement compliquer les analyses.

De plus, il nous semble que le choix du score de Charlson défini par Bannay et al. (168) comme mesure de comorbidités peut également être discutable. En effet, même s'il est conseillé pour les études portant sur les données du SNDS, les poids utilisés pour chacune des maladies le composant ont été définis à partir de la mortalité de la population générale française issue des données françaises du SNDS. Cela a donc entraîné des poids de 0 notamment pour le diabète, alors que celui-ci peut avoir un impact important sur la prise en charge, notamment le recours au médecin généraliste. Par conséquent, l'utilisation d'une autre version du score, comme la version proposée par Quan et al. (167) ou la version originale proposée par Charlson et al. (164) pourrait être utilisée afin de comparer l'évolution de la valeur du score selon les définitions. Des indicatrices spécifiques à certaines maladies auraient également pu être construites, codant pour l'hypertension ou le diabète par exemple, afin de mieux identifier les modifications de prise liées à d'autres maladies mais ne l'ont pas été faute de temps. Néanmoins, il est important de noter que le sujet de l'identification de comorbidités est un sujet à part entière nécessitant un travail dédié. Dans notre cas, certaines approximations ont pu être faites par l'utilisation du score de Charlson tel que défini par Bannay et al., mais il s'agissait de fournir un paramètre d'ajustement, connu pour influer fortement sur la consommation de soins, et qu'il nous paraissait donc important à prendre en compte.

Par ailleurs, l'expertise développée sur les bases a permis de consolider les paramètres reflétant la consommation de soins, l'utilisation des traitements et la prescription des aides techniques. Ainsi, l'ensemble du travail de pré-traitement, portant notamment sur la suppression des séjours ayant des GHM en erreur et des doublons dans le PMSI MCO dus aux grands regroupements d'hôpitaux, a permis dans un premier temps de réduire la volumétrie importante des bases et de ne pas surestimer la consommation hospitalière. De plus, le temps consacré à la gestion des valeurs de délivrances négatives pour les traitements et les équipements techniques a conduit au renforcement des algorithmes développés, notamment celui sur le score EDSS « estimé ». Même si des hypothèses importantes ont été faites, comme le fait qu'une délivrance négative corresponde à une délivrance réelle en l'absence d'autres prestations reliées, elles ont été réfléchies et implémentées d'après notre connaissance des bases. D'un point de vue plus général, la manipulation des données du SNDS nécessite un important de travail dans le data management et la définition d'hypothèses dues à des spécificités de la base ou à de potentielles erreurs de codage. Toutefois, cette préparation, le temps passé à la formulation d'hypothèses et de recodage sont autant d'étapes nécessaires à la conduite d'une étude rigoureuse sur des données médico-administratives.

Enfin, une des limites majeures de ce travail est l'absence de variables cliniques. En effet, ces données si elles étaient présentes seraient de précieux indicateurs pour apprécier l'état de santé global d'un individu et pourraient être utilisées comme paramètre d'ajustement. Dans notre cas, il manque notamment des informations sur la date de début de la maladie ou la survenue de poussées par exemple. Nous avons fait à certains moments dans ce travail l'approximation que la date de mise en ALD coïncidait avec le début de la maladie, cependant cette hypothèse est peu probable. En effet, avec l'arrivée des traitements de fond en 1996, une forte augmentation du nombre d'ALD pour la SEP a été constatée et reconnue par les neurologues afin de faciliter l'accès à ces thérapeutiques. Suite à des discussions avec des cliniciens, il est apparu que cette hypothèse semblait valide pour les patients mis en ALD quasiment dès leur diagnostic au cours des deux dernières décennies, pour leur permettre une meilleure prise en charge thérapeutique précoce. A l'inverse, cela était beaucoup moins vrai avant les années 2000, du fait du peu de traitement disponible et d'une moins bonne connaissance de la maladie (à savoir qu'un traitement précoce est plus efficace qu'un traitement retardé (255,256)). Par ailleurs, afin de fournir un proxy de la survenue de poussées, un travail exploratoire, non présenté dans la thèse car nécessitant des améliorations, a été mené sur leur identification dans les données en se basant sur des algorithmes définis pour des bases médico-administratives étrangères (257–260). Ce nouvel indicateur permettrait alors de mieux comprendre l'impact de ces évènements sur la prise en charge et le parcours de soins. Cependant, il faut garder à l'esprit que ces algorithmes (estimation du score EDSS, survenue de poussées, incidence, …) ne sont que des moyens détournés permettant uniquement d'obtenir une estimation parfois imparfaite d'une mesure clinique indisponible dans les bases.

VII.4. Perspectives

Afin de s'affranchir de la majorité des limites discutées précédemment, le chaînage avec une cohorte peut être envisagé. En effet, les données cliniques à disposition permettraient d'un côté de combler les lacunes du SNDS concernant ces indicateurs, mais aussi d'un autre côté de valider plusieurs algorithmes d'estimation de paramètres cliniques. Cette validation sur une maladie pourrait alors conduire à l'utilisation de l'algorithme dans d'autres maladies où un chaînage avec une cohorte est impossible du fait de la non-existence de celle-ci par exemple. Ce type de chaînage existe déjà pour la cohorte Constances où à terme 200 000 personnes auront des données issues d'examens de santé et de questionnaires chaînées à leur données de consommation de soins issus du SNDS (261). Comme évoqué précédemment, dans le cas de la SEP, un projet de chaînage entre les données du SNDS et la cohorte EDMUS de l'OFSEP est en discussion actuellement. Ce chaînage aurait plusieurs avantages, le premier étant la validation des algorithmes développés dans ce travail, notamment celui sur le score EDSS et celui sur les poussées. Il permettrait également de mieux documenter le biais de recrutement de la cohorte OFSEP en ajoutant les patients suivis hors des centres experts. Par le croisement des

patients identifiés, il donnerait l'occasion de valider l'algorithme d'identification des SEP défini par l'Assurance Maladie et utilisé dans ce travail et permettrait également de détecter s'il conduit à des faux positifs. Enfin, ce chaînage serait utile pour prouver la complémentarité des données cliniques et médico-administratives dans la réalisation d'études épidémiologiques et pharmaco-épidémiologiques. Il faut également souligner que ce dernier représente un challenge technique et constitue un projet à lui seul, notamment à cause de la phase réglementaire à réaliser pour obtenir l'autorisation de chaînage.

VIII. Conclusion

L'analyse des parcours de soins est encore un domaine de recherche très récent, à l'interface entre épidémiologie, organisation des soins et management des organisations de santé. La jeunesse de cette question de recherche explique principalement l'absence de méthodes adaptées pour décrire, représenter et analyser ces parcours afin de permettre à terme la définition de recommandations de prise en charge. Cette lacune oblige notamment à faire preuve d'interdisciplinarité et de transversalité afin d'aller chercher des méthodes adaptées issues d'autres disciplines telles que les sciences sociales ou l'informatique pour pouvoir les adapter à l'épidémiologie et à l'étude des parcours. Cet élargissement du mode de pensée est d'autant plus important avec l'essor qui se développe autour des parcours de soins et la transformation du système de santé dans les quatre prochaines années telle que présentée par le Ministère des Solidarités et de la Santé avec l'enjeu d'une tarification au parcours et non plus à l'acte (262). L'interdisciplinarité ne doit pas seulement intervenir dans l'analyse statistique mais également dans la construction des parcours au travers de la complémentarité des informations recueillies via des études quantitatives et qualitatives, mais aussi issues de la connaissance clinique des professionnels de santé.

Les bases de données médico-administratives sont un bon moyen d'accès à l'information d'un point de vue quantitatif. Leur complexité, leur volumétrie et l'importante variabilité des données à disposition nécessitent une expertise et une phase d'appropriation et de traitement de l'information afin d'en tirer des indicateurs spécifiques aux parcours et à une pathologie donnée. Il est important de garder à l'esprit que ces données ont été détournées de leur but originel et qu'elles ne sont donc pas recueillies dans un but de recherche en épidémiologie, nécessitant le développement d'hypothèses de travail basée sur la discussion avec des professionnels de santé. Enfin, elles possèdent l'atout de permettre l'accès à l'exhaustivité de l'information souvent souhaitée en recherche, posant toutefois des questions méthodologiques importantes sur l'obsolescence potentielle des méthodes statistiques classiques nécessitant peut-être un changement de paradigme.

Références

1. Système nerveux [Internet]. Encyclopédie Larousse; 2018 [cited 2018 Jun 5]. Available from: http://www.larousse.fr/encyclopedie/divers/système_nerveux/73116
2. Mieux comprendre le système nerveux central [Internet]. SEP ensemble; 2018 [cited 2018 Jun 5]. Available from: https://www.sep-ensemble.fr/la-sep-c-est-quoi/comprendre-systeme-nerveux-central
3. Charcot JM. Histologie de la sclérose en plaques. Gaz Hôpitaux. 1868;41:554, 557–558, 566.
4. Vaincre ensemble la SEP [Internet]. Fondation pour l'Aide à la Recherche sur la Sclérose en Plaques (ARSEP); 2018 [cited 2018 Jun 6]. Available from: https://www.arsep.org/
5. Sclérose en plaques [Internet]. Institut National de la Santé Et de la Recherche Médicale (Inserm); 2018 [cited 2018 Jun 12]. Available from: http://www.inserm.fr/thematiques/neurosciences-sciences-cognitives-neurologie-psychiatrie/dossiers-d-information/sclerose-en-plaques-sep
6. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: Guidelines for research protocols. Ann Neurol. 1983;13(3):227–31.
7. McDonald WI, Compston A, Edan G, Goodkin D, Hartung H-. P, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. Ann Neurol. 2001;50(1):121–7.
8. Polman CH, Reingold SC, Edan G, Filippi M, Hartung H-PP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Ann Neurol. 2005;58(6):840–6.
9. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011;69(2):292–302.
10. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018;17(2):162–73.
11. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Neurology. 1996;46(4):907–11.
12. Types of MS. National Multiple Sclerosis Society; 2018.
13. Goldenberg MM. Multiple Sclerosis Review. Pharm Ther. 2012;37(3):175–84.

14. Weinshenker BG, Bass B, Rice GPA, Noseworthy J, Carriere W, Baskerville J, et al. The natural history of multiple sclerosis: a geographically based study. Brain. 1989;112(1):133–46.
15. de Sa JCC, Airas L, Bartholome E, Grigoriadis N, Mattle H, Oreja-Guevara C, et al. Symptomatic therapy in multiple sclerosis: a review for a multimodal approach in clinical practice. Ther Adv Neurol Disord. 2011;4(3):139–68.
16. van Asch P. Impact of Mobility Impairment in Multiple Sclerosis 2 - Patients' Perspectives. Eur Neurol Rev. 2011;6(2):115.
17. Gallien P, Nicolas B, Guichet A. Le point sur la sclérose en plaques. Kinésithérapie, la Rev. 2012;12(125):17–22.
18. Marrie RA, Cutter G, Tyry T, Vollmer T, Campagnolo D. Disparities in the management of multiple sclerosis-related bladder symptoms. Neurology. 2007;68(23):1971–8.
19. Grzegorski T, Losy J. Cognitive impairment in multiple sclerosis - a review of current knowledge and recent research. Rev Neurosci. 2017 Nov;28(8).
20. Minden SL, Frankel D, Hadden L, Perloff J, Srinath KP, Hoaglin DC. The Sonya Slifka Longitudinal Multiple Sclerosis Study: methods and sample characteristics. Mult Scler J. 2006;12(1):24–38.
21. Fiest KM, Fisk JD, Patten SB, Tremlett H, Wolfson C, Warren S, et al. Fatigue and Comorbidities in Multiple Sclerosis. Int J MS Care. 2016;18(2):96–104.
22. Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray TJ. The impact of fatigue on patients with multiple sclerosis. Can J Neurol Sci. 1994;21(1):9–14.
23. Pepper G, Butzkueven H, Dhib-Jalbut S, Giovannoni G, Havrdová E, Hobart J, et al. Brain health: a guide for people with multiple sclerosis. Oxford PharmaGenesis Ltd.; 2016.
24. Confavreux C, Vukusic S, Moreau T, Adeleine P. Relapses and Progression of Disability in Multiple Sclerosis. N Engl J Med. 2000;343(20):1430–8.
25. Confavreux C, Vukusic S, Adeleine P. Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. Brain. 2003;126(4):770–82.
26. Kurtzke JF. Rating neurological impairment in multiple sclerosis: an expanded disability status scale. Neurology. 1983;33:1444–52.
27. Buzzard K, Broadley S, Butzkueven H. What Do Effective Treatments for Multiple Sclerosis Tell Us about the Molecular Mechanisms Involved in Pathogenesis? Int J Mol Sci. 2012 Oct;13(12):12665–709.
28. Freedman MS, Selchen D, Prat A, Giacomini PS. Managing Multiple Sclerosis: Treatment Initiation, Modification, and Sequencing. Can J Neurol Sci / J Can des Sci Neurol. 2018;1–15.
29. Pardo G, Jones DE. The sequence of disease-modifying therapies in relapsing multiple sclerosis: safety and immunologic considerations. J Neurol. 2017;
30. Le Page E, Edan G. Induction or escalation therapy for patients with multiple sclerosis? Rev

Neurol (Paris). 2018;174(6):449–57.

31. Edan G, Le Page E. Induction Therapy for Patients with Multiple Sclerosis: Why? When? How? CNS Drugs. 2013;27(6):403–9.
32. Berger JR, Cree BA, Greenberg B, Hemmer B, Ward BJ, Dong VM, et al. Progressive multifocal leukoencephalopathy after fingolimod treatment. Neurology. 2018;90(20):e1815--e1821.
33. Wiertlewski S. La poussée - L'identifier, la gérer. 2015.
34. Le Page E, Veillard D, Laplaud DA, Hamonic S, Wardi R, Lebrun C, et al. Oral versus intravenous high-dose methylprednisolone for treatment of relapses in patients with multiple sclerosis (COPOUSEP): a randomised, controlled, double-blind, non-inferiority trial. Lancet. 2015 Sep;386(9997):974–81.
35. Ben-Zacharia AB. Therapeutics for multiple sclerosis symptoms. Mt Sinai J Med. 2011;78(2):176–91.
36. Rico-Lamy A. Traitements symptomatiques dans la sclérose en plaques. 2018.
37. Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor B V, et al. Atlas of Multiple Sclerosis 2013: A growing global problem with widespread inequity. Neurology. 2014 Sep;83(11):1022–4.
38. Trojano M, Lucchese G, Graziano G, Taylor B V, Simpson Jr S, Lepore V, et al. Geographical Variations in Sex Ratio Trends over Time in Multiple Sclerosis. PLoS One. 2012;7(10):e48078.
39. MS International Federation. Atlas of MS. 2015.
40. Foulon S, Maura G, Dalichampt M, Alla F, Debouverie M, Moreau T, et al. Prevalence and mortality of patients with multiple sclerosis in France in 2012: a study based on French health insurance data. J Neurol. 2017;264(6):1185–92.
41. Leray E, Vukusic S, Debouverie M, Clanet M, Brochet B, de Sèze J, et al. Excess Mortality in Patients with Multiple Sclerosis Starts at 20 Years from Clinical Onset: Data from a Large-Scale French Observational Study. PLoS One. 2015;10(7):e0132033.
42. Kurtzke JF. MS epidemiology world wide. One view of current status. Acta Neurol Scand. 1995;91(S161):23–33.
43. Fromont A, Binquet C, Sauleau EA, Fournel I, Bellisario A, Adnet J, et al. Geographic variations of multiple sclerosis in France. Brain. 2010;133(7):1889–99.
44. Kurtzke JF, Delasnerie-Lauprêre N. Reflection on the geographic distribution of multiple sclerosis in France. Acta Neurol Scand. 1996;93(2–3):110–7.
45. Fromont A, Binquet C, Sauleau E, Fournel I, Despalins R, Rollot F, et al. National estimate of multiple sclerosis incidence in France (2001-2007). Mult Scler J. 2012;18(8):1108–15.
46. Yaouanq J, Tron I, Kerbrat A, Leray E, Hamonic S, Merienne M, et al. Register-based

incidence of multiple sclerosis in Brittany (north-western France), 2000–2001. Acta Neurol Scand. 2015;131(5):321–8.

47. El Adssi H, Debouverie M, Guillemin F, LORSEP Group. Estimating the prevalence and incidence of multiple sclerosis in the Lorraine region, France, by the capture-recapture method. Mult Scler. 2012;18(9):1244–50.

48. McKay KA, Jahanfar S, Duggan T, Tkachuk S, Tremlett H. Factors associated with onset, relapses or progression in multiple sclerosis: A systematic review. Neurotoxicology. 2016;

49. Hansen T, Skytthe A, Stenager E, Petersen HC, Brønnum-Hansen H, Kyvik KO. Concordance for multiple sclerosis in Danish twins: an update of a nationwide study. Mult Scler. 2005;11(5):504–10.

50. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC, Study CC, et al. Twin concordance and sibling recurrence rates in multiple sclerosis. Proc Natl Acad Sci. 2003;100(22):12877–82.

51. Muñoz-Culla M, Irizar H, Otaegui D. The genetics of multiple sclerosis: review of current and emerging candidates. Appl Clin Genet. 2013;6:63–73.

52. Patsopoulos NA, the Bayer Pharma MS Genetics Working Group, the Steering Committees of Studies Evaluating IFN$β$-1b and a CCR1-Antagonist, ANZgene Consortium, GeneMSA, International Multiple Sclerosis Genetics Consortium, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann Neurol. 2011;70(6):897–912.

53. Goldberg P. Multiple sclerosis: vitamin D and calcium as environmental determinants of prevalence. Int J Environ Stud. 1974;6(1):19–27.

54. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. Ann Neurol. 2007;61(6):504–13.

55. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. Ann Neurol. 2007;61(4):288–99.

56. Hernán MA, Oleky MJ, Ascherio A. Cigarette Smoking and Incidence of Multiple Sclerosis. Am J Epidemiol. 2001;154(1):69–74.

57. Lauer K. Environmental risk factors in multiple sclerosis. Expert Rev Neurother. 2010;10(3):421–40.

58. Belbasis L, Bellou V, Evangelou E, Ioannidis JP, Tzoulaki I. Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses. Lancet Neurol. 2015;14(3):263–73.

59. Andersen O, Lygner PE, Bergstrom T, Andersson M, Vahlne A. Viral infections trigger multiple sclerosis relapses: a prospective seroepidemiological study. J Neurol. 1993;240(7):417–22.

60. Sotelo J, Ordonez G, Pineda B. Varicella-zoster virus at relapses of multiple sclerosis. J

Neurol. 2007;254(4):493–500.

61. Salvi F, Bartolomei I, Smolensky MH, Lorusso A, Barbarossa E, Malagoni AM, et al. A seasonal periodicity in relapses of multiple sclerosis? A single-center, population-based, preliminary study conducted in Bologna, Italy. BMC Neurol. 2010;10:105.
62. Cook RD. Multiple sclerosis: is the domestic cat involved? Med Hypotheses. 1981;7(2):147–54.
63. Artemiadis AK, Anagnostouli MC, Alexopoulos EC. Stress as a risk factor for multiple sclerosis onset or relapse: a systematic review. Neuroepidemiology. 2011;36(2):109–20.
64. Mehrpour M, Shams-Hosseini NS, Rezaali S, Sahraiian MA, Taki S. Effect of Air Pollutant Markers on Multiple Sclerosis Relapses. Iran J Public Heal. 2013;42(10):1167–73.
65. Watad A, Azrielant S, Soriano A, Bracco D, Abu Much A, Amital H. Association between seasonal factors and multiple sclerosis. Eur J Epidemiol. 2016;1–9.
66. Heydarpour P, Amini H, Khoshkish S, Seidkhani H, Sahraian MA, Yunesian M. Potential impact of air pollution on multiple sclerosis in Tehran, Iran. Neuroepidemiology. 2014;43(3–4):233–8.
67. Oikonen M, Laaksonen M, Laippala P, Oksaranta O, Lilius EM, Lindgren S, et al. Ambient air quality and occurrence of multiple sclerosis relapse. Neuroepidemiology. 2003;22(1):95–9.
68. Angelici L, Piola M, Cavalleri T, Randi G, Cortini F, Bergamaschi R, et al. Effects of particulate matter exposure on multiple sclerosis hospital admission in Lombardy region, Italy. Environ Res. 2016;145:68–73.
69. Marrie RA, Hanwell H. General Health Issues in Multiple Sclerosis: Comorbidities, Secondary Conditions, and Health Behaviors. Contin Minneap. 2013;19(4, Multiple Sclerosis):1046–57.
70. Marrie RA, Patten SB, Berrigan LI, Tremlett H, Wolfson C, Warren S, et al. Diagnoses of Depression and Anxiety Versus Current Symptoms and Quality of Life in Multiple Sclerosis. Int J MS Care. 2018;20(2):76–84.
71. Marrie RA, Cohen J, Stuve O, Trojano M, Sørensen PS, Reingold S, et al. A systematic review of the incidence and prevalence of comorbidity in multiple sclerosis: Overview. Mult Scler J. 2015;21(3):263–81.
72. Zhang T, Tremlett H, Zhu F, Kingwell E, Fisk JD, Bhan V, et al. Effects of physical comorbidities on disability progression in multiple sclerosis. Neurology. 2018 Jan;10.1212/WNL.0000000000004885.
73. McKay KA, Tremlett H, Fisk JD, Zhang T, Patten SB, Kastrukoff L, et al. Psychiatric comorbidity is associated with disability progression in multiple sclerosis. Neurology. 2018 Mar;10.1212/WNL.0000000000005302.
74. Conway DS, Thompson NR, Cohen JA. Influence of hypertension, diabetes, hyperlipidemia, and obstructive lung disease on multiple sclerosis disease course. Mult Scler J. 2016;

75. Marrie RA. Comorbidity in multiple sclerosis: implications for patient care. Nat Rev Neurol. 2017;
76. Kowalec K, McKay KA, Patten SB, Fisk JD, Evans C, Tremlett H, et al. Comorbidity increases the risk of relapse in multiple sclerosis: A prospective study. Neurology. 2017;
77. McKay KA, Marrie RA, Fisk JD, Patten SB, Tremlett H. Comorbidities Are Associated with Altered Health Services Use in Multiple Sclerosis: A Prospective Cohort Study. Neuroepidemiology. 2018;1–10.
78. World Health Organization. Innovative care for chronic conditions : building blocks for actions : global report. 2002.
79. Ministère de la Santé, de la Jeunesse, des Sports et de la Vie associative. Plan 2007-2011 pour l'amélioration de la qualité de vie des personnes atteintes de maladies chroniques. 2007.
80. Vanhaecht K. The impact of clinical pathways on the organisation of care processes. Katholieke Universiteit Leuven; 2007.
81. Zander KS, Bower KA, Etheredge MLS. Nursing Case Management: Blueprints for Transformation. New England Medical Center Hospitals; 1987.
82. Schrijvers G, van Hoorn A, Huiskes N. The care pathway: concepts and theories: an introduction. Int J Integr Care. 2012;12(Spec Ed Integrated Care Pathways):e192.
83. Haute Autorité de Santé (HAS). Parcours de soins. Questions/Réponses. 2012.
84. Féry-Lemonnier E. Les parcours, une nécessité. Actual Doss en santé publique. 2014;88:12–5.
85. Secrétariat général des ministères chargés des affaires sociales (SGMCAS). Parcours de soins, Parcours de santé, Parcours de vie - Pour une prise en charge adaptée des patients et usagers - Lexique des parcours de A à Z. 2016.
86. Haute Autorité de Santé (HAS). Guide du parcours de soins - Maladie de Parkinson. 2016.
87. Haute Autorité de Santé (HAS). Guide du parcours de soins - Maladie rénale chronique de l'adulte. 2012.
88. Qu'est-ce qu'une affection de longue durée ? Assurance Maladie; 2015.
89. Haute Autorité de Santé (HAS). Guide - Affection de longue durée - Sclérose en plaques. 2006.
90. Haute Autorité de Santé (HAS). Actes et prestations affection de longue durée - Sclérose en plaques. 2015.
91. Haute Autorité de Santé (HAS). Programme de travail - Version révisée au 31 janvier 2018. 2018.
92. Ministère de l'Education Nationale, de l'Enseignement supérieur et de la Recherche et Ministère des Affaires Sociales, de la Santé et des Droits des Femmes. Plan Maladies Neurodégénératives 2014-2019. 2014.
93. Ministre de la Santé, de la Jeunesse, des Sports et de la Vie associative. Plan "Alzheimer et maladies" 2008-2012. 2008.

94. Société Francophone de la Sclérose en Plaques (SFSEP), editor. Regards croisés et perspectives - Parcours de soins SEP : Analyser aujourd'hui pour construire demain. In 2017.
95. Nemirovsky B, Papeix C, Moreau T, Pelletier J, Stankoff B, Zephir H. Parcours de soins : Sclérose en plaques : adapter le parcours de soins à l'hétérogénéité des soins cliniques. Le Concours Médical. 2015;137(3):191–231.
96. Les Réseaux SEP. Ligue française contre la sclérose en plaques; 2018.
97. Observatoire Français de la Sclérose en Plaques [Internet]. Observatoire Français de la Sclérose en Plaques; 2018 [cited 2018 Jun 12]. Available from: http://www.ofsep.org/fr/
98. Halper J. Comprehensive Care in Multiple Sclerosis - A Patient-centred Approach. Eur Neurol Rev. 2008;3(2):72.
99. The Consortium of Multiple Sclerosis Centers. Comprehensive Care in Multiple Sclerosis. 2010.
100. Ross AP, Halper J, Kennedy P, Walker T. Treating the Whole Patient: Applying Comprehensive Care in MS. Monson N, editor. Delaware Media Group; 2012. (Counseling Points - Enhancing Patient Communication for the MS Nurse; vol. 8).
101. Oreja-Guevara C, Miralles A, Garcia-Caballero J, Noval S, Gabaldón L, Esteban-Vasallo MD, et al. Clinical pathways for the care of multiple sclerosis patients. Neurol (English Ed. 2010;25(3):156–62.
102. Al-Khedr A, Blanchard-Dauphin A, Donze C, Hautecoeur P, Kriseler M, Galles D Le, et al. La sclérose en plaques - Guide pratique des dispositifs en région Hauts-de-France. 2017.
103. Agence Nationale d'Accréditation et d'Evaluation en Santé (ANAES) et Fédération Française de Neurologie (FFN). Conférence de consensus - La sclérose en plaques. 2001.
104. Steenbrugghe FH-V. Prise en charge globale de la sclérose en plaques. Université de Bordeaux; 2016.
105. Lefeuvre D, Rudant J, Foulon S, Alla F, Weill A. Healthcare expenditure of multiple sclerosis patients in 2013: A nationwide study based on French health administrative databases. Mult Scler J - Exp Transl Clin. 2017;3(3):2055217317730421.
106. Pohar SL, Jones CA, Warren S, Turpin K V, Warren K. Health status and health care utilization of multiple sclerosis in Canada. Can J Neurol Sci. 2007;34(2):167–74.
107. Marrie RA, Yu N, Wei Y, Elliott L, Blanchard J. High rates of physician services utilization at least five years before multiple sclerosis diagnosis. Mult Scler J. 2013;19(8):1113–9.
108. Wijnands JMA, Kingwell E, Zhu F, Zhao Y, Högg T, Stadnyk K, et al. Health-care use before a first demyelinating event suggestive of a multiple sclerosis prodrome: a matched cohort study. Lancet Neurol. 2017;16(6):445–51.
109. McKay KA, Tremlett H, Zhu F, Kastrukoff L, Marrie RA, Kingwell E. A population-based study comparing multiple sclerosis clinic users and non-users in British Columbia, Canada. Eur

J Neurol. 2016;23(6):1093–100.

110. Jones E, Pike J, Marshall T, Ye X. Quantifying the relationship between increased disability and health care resource utilization, quality of life, work productivity, health care costs in patients with multiple sclerosis in the US. BMC Health Serv Res. 2016;16(1):1–9.

111. Marrie RA, Elliott L, Marriott J, Cossoy M, Tennakoon A, Yu N. Comorbidity increases the risk of hospitalizations in multiple sclerosis. Neurology. 2014;84(4):350–8.

112. Ytterberg C, Lundqvist S, Johansson S. Use of health services in people with multiple sclerosis with and without depressive symptoms: a two-year prospective study. BMC Health Serv Res. 2013 Sep;13(1).

113. Yermakov S, Davis M, Calnan M, Fay M, Cox-Buckley B, Sarda S, et al. Impact of increasing adherence to disease-modifying therapies on healthcare resource utilization and direct medical and indirect work loss costs for patients with multiple sclerosis. J Med Econ. 2015;18(9):711–20.

114. Direction de la stratégie, des études et des statistiques (DSES). Personnes prises en charge pour sclérose en plaques (SEP) en 2013. 2015.

115. Direction de la stratégie, des études et des statistiques (DSES). Personnes prises en charge pour sclérose en plaques (SEP) en 2015. 2017.

116. Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS). Améliorer la qualité du système de santé et maîtriser les dépenses : Propositions de l'Assurance Maladie pour 2015. 2014.

117. Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS). Améliorer la qualité du système de santé et maîtriser les dépenses : Propositions de l'Assurance Maladie pour 2016. 2015.

118. Lebrun-Frenay C, Kobelt G, Berg J, Capsa D, Gannedahl M. New insights into the burden and costs of multiple sclerosis in Europe: Results for France. Mult Scler J. 2017;23(2_suppl):65–77.

119. Com-Ruelle L, Nestrigue C, Le Guen N, Bricard D, Le Neindre C, Lucas-Gabrielli V. Parcours de soins des personnes hospitalisées pour un accident vasculaire cérébral - Premiers Résultats. In: Questions d'économie de la Santé. IRDES; 2018.

120. Hammel P, Coriat R, Lledo G, de Bausset M, Selosse M, Obled S, et al. Care pathway of patients with metastatic pancreatic cancer in daily practice in France: Results from the REPERE national survey. Bull Cancer. 2017;104(4):321–31.

121. Yatim F, Ferrua M, Fourcade A, Di Palma M, Minvielle E. Barriers to cancer care pathways in France: A qualitative case study. J Clin Oncol. 2016;34(15_suppl):e18032–e18032.

122. Josseran L, Baptiste C, Bensmail D, Karam P. Patient Care Pathway for Post-Stroke Spasticity and Bont Management in French Hospitals Through the Prism of PMSI Data. Value Heal. 2017

Oct;20(9):A629--A630.

123. Tuppin P, Samson S, Fagot-Campagna A, Woimant F. Care pathways and healthcare use of stroke survivors six months after admission to an acute-care hospital in France in 2012. Rev Neurol (Paris). 2016;172(4–5):295–306.

124. Le Meur N, Gao F, Bayat S. Mining care trajectories using health administrative information systems: the use of state sequence analysis to assess disparities in prenatal care consumption. BMC Heal Serv Res. 2015;15(1):200.

125. Nuemi G, Afonso F, Roussot A, Billard L, Cottenet J, Combier E, et al. Classification of hospital pathways in the management of cancer: Application to lung cancer in the region of burgundy. Cancer Epidemiol. 2013;37(5):688–96.

126. Egho E, Jay N, Raïssi C, Nuemi G, Quantin C, Napoli A. An Approach for Mining Care Trajectories for Chronic Diseases. In: Peek N, Marín Morales R, Peleg M, editors. Artificial Intelligence in Medicine. Springer Berlin Heidelberg; 2013. p. 258–67. (Lecture Notes in Computer Science; vol. 7885).

127. Defossez G, Rollet A, Dameron O, Ingrand P. Temporal representation of care trajectories of cancer patients using data from a regional information system: an application in breast cancer. MC Med Informatics Decis MakingBMC Med Informatics Decis Mak. 2014;14:24.

128. Vogt V, Scholz SM, Sundmacher L. Applying sequence clustering techniques to explore practice-based ambulatory care pathways in insurance claims data. Eur J Public Health. 2017 Oct;

129. Cordesse V, Jametal T, Guy C, Lefebvre S, Roussel M, Ruggeri J, et al. Analysis of clinical pathway in changing and disabling neurological diseases. Rev Neurol (Paris). 2013;4085(6):e1-529.

130. Davies N, Manthorpe J, Sampson EL, Iliffe S. After the Liverpool Care Pathway—development of heuristics to guide end of life care for people with dementia: protocol of the ALCP study. BMJ Open. 2015;5(9).

131. Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. Rev Epidemiol Sante Publique. 2017 Oct;65:S149--S167.

132. Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc J-L, Sailler L. French health insurance databases: What interest for medical research? La Rev Médecine Interne. 2015;36(6):411–7.

133. Haute Autorité de Santé (HAS). Rapport d'analyse prospective 2018 - De nouveaux choix pour soigner mieux. 2018.

134. Dlouhy K, Biemann T. Optimal matching analysis in career research: A review and some best-practice recommendations. J Vocat Behav. 2015;90:163–73.
135. Studer M, Ritschard G. What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. J R Stat Soc Ser A (Statistics Soc. 2015;179(2):481–511.
136. Direction de la Stratégie, des Études et des Statistiques. Présentation générale du SNIIRAM. 2016.
137. Agence Technique de l'Information sur l'Hospitalisation (ATIH). Hospitalisation - Chiffres clés 2016. 2017.
138. Karnofsky DA, Abelmann WH, Craver LF, Burchenal JH. The use of the nitrogen mustards in the palliative treatment of carcinoma.With particular reference to bronchogenic carcinoma. Cancer. 1948 Nov;1(4):634–56.
139. Tuppin P, de Roquefeuil L, Weill A, Ricordeau P, Merlière Y. French national health insurance information system and the permanent beneficiaries sample. Rev Epidemiol Sante Publique. 2010;58(4):286–90.
140. de Roquefeuil L, Studer A, Neumann A, Merlière Y. L'Échantillon généraliste de bénéficiaires : représentativité, portée et limites. In: Points de repère. Caisse Nationale de l'Assurance Maladie (CNAM); 2009.
141. Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. Pharmacoepidemiol Drug Saf. 2017;26(8):954–62.
142. Définition : Qu'est-ce que le Big Data ? Le big data - Le magazine cloud & big data; 2018.
143. Présentation générale du Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM) - Formation "Architecture et données du SNIIRAM." 2016.
144. Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés. Cahier des charges de la norme NOEMIE (Norme Ouverte d'Échanges entre la Maladie et les Intervenants Extérieurs) pour les organismes complémentaires. 2016.
145. Agence Technique de l'Information sur l'Hospitalisation (ATIH). Classification statistique internationale des maladies et des problèmes de santé connexes 10ème version (CIM-10) à l'usage du PMSI en France. 2015.
146. Union nationale des caisses d'assurance maladie (Uncam). Nomenclature générale des actes professionnels (NGAP) - Version du 11 mars 2016. 2016.
147. Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS). Classification Commune des Actes Médicaux (CCAM) - Version 53. 2018.
148. Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS) & Direction Déléguée à la Gestion de l'Offre de Soins (DDGOS). Table Nationale de codage de Biologie -

Version 48. 2018.

149. Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS). Liste des Produits et Prestationsremboursables (LPP) - Version 501. 2018.

150. FINESS - Fichier National des Etablissements Sanitaires et Sociaux. Ministère de l'action et des comptes publics, Ministère des solidarités et de la santé, Ministère du travail, Direction de la Recherche, des Etudes, de l'Evaluation et de Statistiques (DREES); 2018.

151. Lefeuvre D, Rudant J, Foulon S, Weill A, Alla F. Prise en charge des patients atteints de sclérose en plaques en France en 2013–Consommations de soins et montants remboursés par l'Assurance maladie. Rev Epidemiol Sante Publique. 2016;64, Supple:S26–7.

152. Lefeuvre D, Rudant J, Foulon S, Weill A, Alla F. Coût de la sclérose en plaques en France : description des différents postes de dépenses à partir des bases de données de l'assurance maladie. Rev Neurol (Paris). 2016;172:A84.

153. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2015.

154. Ministère des Affaires Sociales, de la Santé et des Droits des femmes. Guide méthodologique de production des informations relatives à l'activité médicale et à sa facturation en Médecine, Chirurgie, Obstétrique et Odontologie. 2015.

155. Service des Statistiques, des Etudes et des Systèmes d'Information. Constitution de codes géographiques de zones de plus de 1000 habitants à partir des codes postaux. 1997.

156. Code Postal - Code INSEE 2015 [Internet]. OpenDataSoft; 2018 [cited 2018 Jul 13]. Available from: https://data.opendatasoft.com/explore/dataset/code-postal-code-insee-2015%40public/export/

157. Estimation de la population au 1er janvier 2018 - Séries par région, département, sexe et âge de 1975 à 2018. Institut national de la statistique et des études économiques; 2018.

158. Waniez P. Philcarto. Waniez, P.; 2015.

159. Barlet M, Coldefy M, Collin C, Lucas-Gabrielli V. L'accessibilité potentielle localisée (APL) : une nouvelle mesure de l'accessibilité aux médecins généralistes libéraux. In: DREES, editor. Etudes et résultats. Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) et Institut de recherche et documentation en économie de la santé (IRDES); 2012.

160. Barlet M, Coldefy M, Collin C, Lucas-Gabrielli V. L'Accessibilité potentielle localisée (APL) : une nouvelle mesure de l'accessibilité aux soins appliquée aux médecins généralistes libéraux en France. In: DREES, editor. Etudes et recherche. Direction de la recherche, des études, de l'évaluation et des statistiques (DREES); 2012.

161. Direction de la recherche, des études, de l'évaluation et des statistiques (DREES). Mesurer l'accès aux soins de premier recours à une échelle géographique locale : l'indicateur d'Accessibilité Potentielle Localisée (APL) aux principales professionnels de soins premier

recours. 2015.

162. Rey G, Rican S, Jougla E. Mesure des inégalités de mortalité par cause de décès. Approche écologique à l'aide d'un indice de désavantage social. Bull Epidémiologique Hebd. 2011;8–9:87–90.

163. Fonds de financement de la protection complémentaire de la couverture universelle du risque maladie. Des dispositifs d'aide pour favoriser l'accès aux soins. 2018.

164. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis. 1987 Jan;40(5):373–83.

165. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol. 1992;45(6):613–9.

166. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005;43(11):1130–9.

167. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. Am J Epidemiol. 2011;173(6):676–82.

168. Bannay A, Chaignot C, Blotière PO, Basson M, Weill A, Ricordeau P, et al. The Best Use of the Charlson Comorbidity Index With Electronic Health Care Database to Predict Mortality. Med Care. 2016;54(2):188–94.

169. Gabadinho A, Ritschard G, Müller NS, Studer M. Analyzing and Visualizing State Sequences in R with TraMineR. J Stat Softw. 2011;40(4):1–37.

170. Deburghgraeve V, Brunet I. Le Gilenya - Traitement oral dans la sclérose en plaques. 2015.

171. Casey R, Rollot F, Clanet M, Brochet B, Pelletier J, Laplaud DA, et al. Observatoire Français de la Sclérose en Plaques (OFSEP): a powerful epidemiological tool. In: Multiple Sclerosis Journal. SAGE Publications; 2017. p. 85–426.

172. Defer G, de Seze J, Bouee S, Courouve L, Longin J, Payet M, et al. Outcomes and treatment management of a French cohort suffering from multiple sclerosis: A retrospective epidemiological study. Mult Scler Relat Disord. 2018 Oct;25:276–81.

173. Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS). IR_NAT_V - Table des valeurs des natures de prestation. 2018.

174. Abbott A, Forrest J. Optimal Matching Methods for Historical Sequences. J Interdiscip Hist. 1986;16(3):471.

175. Studer M, Ritschard G, Gabadinho A, Muller NS. Discrepancy Analysis of State Sequences. Sociol Methods Res. 2011;40(3):471–510.

176. Gabadinho A, Ritschard G. Analyzing state sequences with probabilistic suffix trees: the PST

R package. J Stat Softw. 2016;72(3):1–39.

177. Blanchard P, Bühlmann F, (eds.) J-AG. Advances in Sequence Analysis: Theory, Method, Applications. 1st ed. Blanchard P, Bühlmann F, Gauthier J-A, editors. Springer International Publishing; 2014. (Life Course Research and Social Policies 2).

178. Gauthier J-A, Valarino I. Approche séquentielle multidimensionnelle de l'activation du statut-maitre sexué lors de la transition à la parentalité. In: Devenir parent, devenir inégaux, transition à la parentalité et inégalités de genre. First. Zurich, Switzerland: Seismo; 2016. p. 48–73.

179. Pollock G. Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis. J R Stat Soc Ser A (Statistics Soc. 2007;170(1):167–83.

180. Struffolino E, Studer M, Fasang AE. Gender, education, and family life courses in East and West Germany: Insights from new sequence analysis techniques. Adv Life Course Res. 2016;29:66–79.

181. Studer M, Struffolino E, Fasang AE. Estimating the Relationship between Time-varying Covariates and Trajectories: The Sequence Analysis Multistate Model Procedure. Sociol Methodol. 2018 Jan;008117501774712.

182. McVicar D, Anyadike-Danes M. Predicting successful and unsuccessful transitions from school to work by using sequence methods. J R Stat Soc Ser A (Statistics Soc. 2002;165(2):317–34.

183. Anyadike-Danes M, McVicar D. You'll never walk alone: Childhood influences and male career path clusters. Labour Econ. 2005;12(4):511–30.

184. Brzinsky-Fay C. Lost in Transition? Labour Market Entry Sequences of School Leavers in Europe. Eur Sociol Rev. 2007;23(4):409–22.

185. Han Y, Liefbroer A, Elzinga C. Comparing methods of classifying life courses: sequence analysis and latent class analysis. Longit Life Course Stud. 2017 Oct;8(4):319–41.

186. Gosselin A, du Loû AD, Lelièvre E. How to use sequence analysis for life course epidemiology? An example on HIV-positive Sub-Saharan migrants in France. J Epidemiol Community Health. 2018;jech--2017--209739.

187. Gauthier J-A, Widmer ED, Bucher P, Notredame C. Multichannel sequence analysis applied to social science data. Sociol Methodol. 2010;40(1):1–38.

188. Studer M, Liefbroer AC, Mooyaart JE. Understanding Trends in Family Formation Trajectories: An Application of Competing Trajectories Analysis. Adv Life Course Res. 2018 Mar;

189. Sironi M, Barban N, Impicciatore R. Parental social class and the transition to adulthood in Italy and the United States. Adv Life Course Res. 2015;26:89–104.

190. Studer M. WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. 2013.

191. Le Meur N, Vigneau C, Lefort M, Lebbah S, Jais J-P, Daugas E, et al. Categorical state sequence analysis and regression tree to identify determinants of care trajectory in chronic disease: Example of end-stage renal disease. Stat Methods Med Res. 2018;096228021877481.
192. Gabadinho A, Ritschard G, Studer M, Müller N. Extracting and Rendering Representative Sequences. In: Fred A, Dietz JLG, Liu K, Filipe J, editors. Knowledge Discovery, Knowlege Engineering and Knowledge Management. Springer Berlin Heidelberg; 2011. p. 94–106. (Communications in Computer and Information Science; vol. 128).
193. Lesnard L, Kan MY. Investigating scheduling of work: a two-stage optimal matching analysis of workdays and workweeks. J R Stat Soc Ser A (Statistics Soc. 2011;174(2):349–68.
194. Andrade SB. Transition and Adaptation: An Analysis of Adaption Strategies Amongst Danish Farm Families from 1980–2008. Sociol Ruralis. 2015;
195. Halpin B. Imputing sequence data: Extensions to initial and terminal gaps, Stata's mi. 2013.
196. Halpin B. Multiple imputation for life-course sequence data. Limerick; 2012.
197. Biemann T, Datta DK. Analyzing Sequence Data: Optimal Matching in Management Research. Organ Res Methods. 2014;17(1):51–76.
198. Elzinga CH. Distance, Similarity and Sequence Comparison. In: Blanchard P, Bühlmann F, Gauthier J-A, editors. Advances in Sequence Analysis: Theory, Method, Applications. Springer International Publishing; 2014. p. 51–73. (Life Course Research and Social Policies).
199. Levenshtein VI. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. Sov Phys Dokl. 1966;10:707.
200. Deville J-C, Saporta G. Correspondence analysis, with an extension towards nominal time series. J Econom. 1983;22(1):169–89.
201. Elzinga CH. Sequence Similarity. Sociol Methods Res. 2003;32(1):3–29.
202. Elzinga CH. Combinatorial Representations of Token Sequences. J Classif. 2005;22(1):87–118.
203. Abbott A, Tsay A. Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect. Sociol Methods Res. 2000;29(1):3–33.
204. Sankoff D, Kruskal J. Time warps, string edits, and macromolecules: the theory and practice of sequence comparison. Reading A-W, editor. Sankoff, David and Kruskal, Joseph; 1983.
205. Lesnard L, de Saint Pol T. Introduction aux méthodes d'appariement optimal (Optimal Matching Analysis). Bull méthodologie Sociol. 2006;90:5–25.
206. Abbott A, Hrycak A. Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians{\textquotesingle} Careers. Am J Sociol. 1990;96(1):144–85.
207. Lesnard L. Setting Cost in Optimal Matching to Uncover Contemporaneous Socio-Temporal Patterns. Sociol Methods Res. 2010;38(3):389–419.
208. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the

amino acid sequence of two proteins. J Mol Biol. 1970;48(3):443–53.

209. Robette N, Bry X. Harpoon or Bait? A Comparison of Various Metrics in Fishin} for Sequence Patterns. BMS Bull Sociol Methodol / Bull Méthodologie Sociol. 2012;(116):5–24.

210. Hamming RW. Error Detecting and Error Correcting Codes. Bell Syst Tech J. 1950;29(2):147–60.

211. Yujian L, Bo L. A normalized Levenshtein distance metric. IEEE Trans Pattern Anal Mach Intell. 2007;29(6):1091–5.

212. Elzinga CH, Studer M. Normalization of distance and similarity in sequence analysis. Sociol Methods Res. In press.

213. Batagelj V. Generalized Ward and Related Clustering Problems. In: Bock HH, editor. Classification and Related Methods of Data Analysis. Amsterdam, the Netherlands: North-Holland; 1988. p. 67–74.

214. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. New York: Wiley; 1990. (Wiley Series in Probability and Statistics).

215. Hubert L, Levin J. A general statistical framework for assesing categorical clustering in free recall. Psychol Bull. 1976;83:1072–1080.

216. Milligan G, Cooper M. An examination of procedures for determining the number of clusters in a data set. Psychometrika. 1985;50(2):159–179.

217. Hennig C, Liao TF. Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification. 2010.

218. Hubert L, Arabie P. Comparing Partitions. J Classif. 1985;2:193–218.

219. British Columbia Ministry of Health [creator] (2012) Medical Services Plan (MSP) Payment Information File Population Data BC [publisher]. Data Extract. MOH (2012). http://www.popdata.bc.ca/data

220. British Columbia Ministry of Health [creator] (2012) Discharge Abstract Database (Hospital Separations). Population Data BC [publisher]. Data Extract. MOH (2012). http://www.popdata.bc.ca/data

221. British Columbia Ministry of Health [creator] (2012). PharmaNet: BC Ministry of Health publisher]; 2012. Data Extract MOH (2012). http://www.popdata.bc.ca/data

222. BC Vital Statistics Agency [creator] (2012). Vital Statistics Deaths: Population Data BC [publisher]. Data Extract. MOH (2012). http://www.popdata.bc.ca/data

223. British Columbia Ministry of Health [creator] (2012) Consolidation File (MSP Registration & Premium Billing). Population Data BC [publisher]. Data Extract. MOH (2012). http://www.popdata.bc.ca/data

224. Statistics Canada (2014) Annual Income Estimates for Census Families and Individuals (T1 Family File).

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4105&lang=en&db=IMDB&dbg=f&adm=8&dis=2. Accessed 11 August2014

225. Manouchehrinia A, Westerlind H, Kingwell E, Zhu F, Carruthers R, Ramanujam R, et al. Age Related Multiple Sclerosis Severity Score: Disability ranked by age. Mult Scler. 2017;1352458517690618.

226. Marrie RA, Miller A, Sormani MP, Thompson A, Waubant E, Trojano M, et al. Recommendations for observational studies of comorbidity in multiple sclerosis. Neurology. 2016;86(15):1446–53.

227. OECD Health Statistics. Health care utilisation. 2016.

228. Montaut A. Santé et recours aux soins des femmes et des hommes. 2010.

229. Tonnelier F, Vigneron E. Géographie de la santé en France. In: Que sais-je ? PUF; 1999. p. 127.

230. Maslow A. A Theory of Human Motivation. Psychol Rev. 1943;50:370–96.

231. Com-Ruelle L, Crestin B, Dumesnil S. L'asthme en France selon les stades de sévérité. In: Questions d'économie de la Santé. Centre de Recherche, d'Etude et de Documentation en Economie de la Santé (CREDES); 2000.

232. Le Pape A, Lecomte T. Prévalence et prise en charge médicale de la dépression en 1996-1997. In: Questions d'économie de la Santé. Centre de Recherche, d'Etude et de Documentation en Economie de la Santé (CREDES); 1999.

233. Rieckmann P, Centonze D, Elovaara I, Giovannoni G, Havrdová E, Kesselring J, et al. Unmet needs, burden of treatment, and patient engagement in multiple sclerosis: A combined perspective from the MS in the 21st Century Steering Group. Mult Scler Relat Disord. 2018 Jan;19:153–60.

234. Billari FC. Sequence analysis in demographic research. Can Stud Popul. 2001;28(2):439–58.

235. Hagenaars JA, McCutcheon AL, editors. Applied Latent Class Analysis. Cambridge University Press; 2002.

236. Genolini C, Alacoque X, Sentenac M, Arnaud C, others. kml and kml3d: R Packages to Cluster Longitudinal Data. J Stat Softw. 2015;65(1):1–34.

237. Barban N, Billari FC. Classifying life course trajectories: a comparison of latent class and sequence analysis. J R Stat Soc Ser C (Applied Stat. 2012;61(5):765–84.

238. Mikolai J, Lyons-Amos M. Longitudinal methods for life course research: A comparison of sequence analysis, latent class growth models, and multi-state event history models for studying partnership transitions. Longit Life Course Stud. 2017;8(2):191–208.

239. Magidson J, Vermunt J. Latent class models. In: Kaplan D, editor. The Sage handbook of quantitative methodology for the social sciences. Thousands Oakes: Sage; 2004. p. 175–98.

240. Debouverie M, Laforest L, Van Ganse E, Guillemin F. Earlier disability of the patients

followed in Multiple Sclerosis centers compared to outpatients. Mult Scler J. 2008 Oct;15(2):251–7.

241. Bardy J, Billion M, Bourrelier T, Chahbi F, Gaillard H, Majka C, et al. La recherche quantitative au service de l'harmonisation du parcours de soins de la sclérose en plaques. Module Interprofessionnel de santé publique - Ecole des Hautes Etudes en Santé Publique; 2018.
242. Marrie RA, Yu N, Blanchard J, Leung S, Elliott L. The rising prevalence and changing age distribution of multiple sclerosis in Manitoba. Neurology. 2010;74(6):465–71.
243. Dowle M, Srinivasan A. data.table: Extension of `data.frame`. 2018.
244. Beaudeau P, Bentayeb M, Corso M, Rambaud L, C. G. Les données de l'entrepôt de cas de gastro-entérite médicalisés issues du SNIIRAM : description, qualité et utilisation. 2017.
245. Austin PC. Statistical power to detect violation of the proportional hazards assumption when using the Cox regression model. J Stat Comput Simul. 2017 Nov;88(3):533–52.
246. Falissard B. Quelle inférence pour l'épidémiologie à l'heure des big data ? In: Big data et protection sociale. Revue française des affaires sociales - La documentation Française; 2017. p. 127–32.
247. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser B. 1994;58:267–88.
248. Nouveau zonage pour renforcer l'aide à l'installation de médecins en Pays de la Loire. Agence Régionale de Santé Pays de la Loire; 2018.
249. Vergier N, Chaput H. Déserts médicaux : comment les définir ? Comment les mesurer ? 2017.
250. Ducros D, Nicoules V, Chehoud H, Bayle A, Souche A, Tanguy M, et al. Les bases médico-administratives pour mesurer les inégalités sociales de santé. Sante Publique (Paris). 2015;27(3):383–94.
251. Windenberger F, Rican S, Jougla E, Rey G. Spatiotemporal association between deprivation and mortality: trends in France during the nineties. Eur J Public Health. 2011 Mar;22(3):347–53.
252. Temam S, Varraso R, Pornet C, Sanchez M, Affret A, Jacquemin B, et al. Ability of ecological deprivation indices to measure social inequalities in a French cohort. BMC Public Health. 2017;17(1).
253. Castell L, Dennevault C. Qualité et accès aux soins : que pensent les Français de leurs médecins ? In: Etudes et Résultats. Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques (DREES); 2017.
254. Dominguez-Berjon F, Borrell C, Rodriguez-Sanz M, Pastor V. The usefulness of area-based socioeconomic measures to monitor social inequalities in health in Southern Europe. Eur J Public Health. 2005;16(1):54–61.

255. Leray E, Yaouanq J, Le Page E, Coustans M, Laplaud D, Oger J, et al. Evidence for a two-stage disability progression in multiple sclerosis. Brain. 2010;133(7):1900–13.
256. Tintore M. New options for early treatment of multiple sclerosis. J Neurol Sci. 2009;277:S9--S11.
257. Marriott JJ, Chen H, Fransoo R, Marrie RA. Validation of an algorithm to detect severe MS relapses in administrative health databases. Mult Scler Relat Disord. 2018 Jan;19:134–9.
258. Colais P, Agabiti N, Davoli M, Buttari F, Centonze D, De Fino C, et al. Identifying Relapses in Multiple Sclerosis Patients through Administrative Data: A Validation Study in the Lazio Region, Italy. Neuroepidemiology. 2017;48(3–4):171–8.
259. Bergamaschi R, Agnello M, Colombo E, Della Giovanna M, Montomoli C, Nava A, et al. Detection of clinical relapses in multiple sclerosis cohorts: construction and validation of a model based on administrative data. Neurol Sci. 2014;35(2):265–9.
260. Chastek BJ, Oleen-Burkey M, Lopez-Bresnahan M V. Medical chart validation of an algorithm for identifying multiple sclerosis relapse in healthcare claims. J Med Econ. 2010;13(4):618–25.
261. Zins M, Goldberg M. The French CONSTANCES population-based cohort: design, inclusion and follow-up. Eur J Epidemiol. 2015 Oct;30(12):1317–28.
262. Ministère des Solidarités et de la Santé. Ma santé 2022 - Un engagement collectif. 2018.
263. La base de données en ligne des prescripteurs libéraux [Internet]. VIDAL; 2015 [cited 2015 Sep 20]. Available from: https://www.vidal.fr/

Annexes

Annexe A-Systèmes fonctionnels et Echelle du score EDSS

Table 34-Description des scores fonctionnels et des critères d'attribution des scores

Score	Critères
Fonction pyramidale	
0	Normal
1	Signes anormaux sans handicap
2	Handicap minime
3	Paraparésie ou hémiparésie légère ou modérée
4	Paraparésie ou hémiparésie marquée ; tétraparésie modérée ; ou monoplégie
5	Paraplégie, hémiplégie, ou tétraplégie marquée
6	Tétraplégie
Fonction cérébelleuse	
0	Normal
1	Anormal sans handicap
2	Légère ataxie
3	Ataxie modérée des membres ou du tronc
4	Ataxie sévère des quatre membres
5	Impossibilité de réaliser des mouvements coordonnés en raison de l'ataxie
X	A ajouter quand le déficit moteur (grade 3 ou plus du score pyramidal) interfère avec l'examen

Score	Critères
Fonction sensitive	
0	Normal
1	Diminution du sens vibratoire ou seulement altération de la reconnaissance de figure dessinée sur la peau, affectant un ou deux membres
2	Légère diminution au touché-piqué ou au sens de position, et/ou diminution modérée du sens vibratoire, dans un ou deux membres ; ou diminution du sens vibratoire (ou de la reconnaissance de figure dessinée sur la peau) dans trois ou quatre membres
3	Diminution modérée au touché-piqué ou au sens de position, et/ou perte du sens vibratoire, dans un ou deux membres ; ou diminution légère du sens tactile nociceptif et/ou diminution modérée de tous les tests proprioceptifs de trois ou quatre membres.
4	Diminution marquée au touché-piqué ou perte de la proprioception, isolées ou associées dans un ou deux membres ; ou diminution modérée au touché-piqué et/ou diminution sévère de la proprioception dans plus de deux membres
5	Perte (essentiellement) de la sensibilité dans un ou deux membres ; ou diminution modérée au touché-piqué et/ou perte de la proprioception de l'ensemble du corps en-dessous de la tête
6	Perte de la sensibilité en-dessous de la tête
Fonction du tronc cérébral	
0	Normal
1	Signes uniquement
2	Nystagmus modéré ou autre handicap léger
3	Nystagmus sévère, faiblesse extra-oculaire marquée, ou handicap modéré d'autres nerfs crâniens
4	Dysarthrie marquée ou autre handicap marqué
5	Incapacité de déglutir ou parler

Score	Critères
Fonction sphinctérienne	
0	Normal
1	Légère dysurie, impériosité ou rétention
2	Dysurie, impériosité modérées, rétention urinaire ou constipation modérée, ou incontinence urinaire rare
3	Incontinence urinaire fréquente
4	Nécessite des sondages quasi-constants
5	Perte de fonction urinaire
Fonction visuelle	
0	Normal
1	Scotome avec acuité visuelle corrigée > 0,67
2	Scotome sur l'œil le plus atteint avec une acuité visuelle entre 0,34 et 0,67
3	Large scotome sur l'œil le plus atteint, ou réduction du champ visuel, avec acuité visuelle comprise entre 0,2 et 0,33
4	Œil le plus atteint avec diminution marquée du champ visuel, acuité visuelle entre 0,1 et 0,2 ; ou grade 3 avec acuité visuelle du meilleur œil < 0,3
5	Acuité visuelle de l'œil le plus atteint < 0,1 ; ou grade 4 avec acuité visuelle du meilleur œil < 0,3
6	Grade 5 avec acuité visuelle du meilleur œil < 0,3
X	À ajouter quand il y a pâleur temporale
Fonction cérébrale	
0	Normal
1	Altération isolée de l'humeur (n'affecte pas le score EDSS)
2	Diminution légère de l’idéation
3	Diminution modérée de l’idéation
4	Diminution marquée de l’idéation
5	Démence
Autres fonctions	
0	Normal
1	Toute autre perturbation neurologique attribuable à la SEP

Table 35-Echelle du score EDSS

Score	Critères
0	Examen neurologique normal (tous systèmes fonctionnels (SF) à 0 ; SF 1 mental acceptable)
1,0	Absence de handicap fonctionnel, signes minimes d'atteinte d'une des fonctions (SF 1, à l'exclusion du SF mental)
1,5	Absence de handicap fonctionnel, signes minimes dans plus d'un SF (plus d'un SF 1, à l'exclusion du SF mental)
2,0	Handicap minime d'un des SF (1 SF 2, les autres 0 ou 1)
2,5	Handicap minime dans 2 SF (2 SF 2, les autres 0 ou 1)
3,0	Handicap modéré dans un SF (1 SF score 3, les autres 0 ou 1) ; ou atteinte minime de 3 ou 4 fonctions (3 ou 4 SF 2, les autres 0 ou 1), mais malade totalement ambulatoire
3,5	Totalement ambulatoire, mais atteinte modérée dans un SF (SF 3) et 1 ou 2 SF 2 ; ou 2 SF 3 ; ou 5 SF 2 (les autres 0 ou 1)
4,0	Malade totalement autonome pour la marche, vaquant à ses occupations 12h par jour malgré une gêne fonctionnelle relativement importante : 1 SF 4 (les autres 0 ou 1), ou association de niveaux inférieurs dépassant les limites des degrés précédents. Capable de marcher 500m environ sans aide ni repos
4,5	Malade autonome pour la marche, vaquant à ses occupations la majeure partie de la journée, capable de travailler une journée entière, mais pouvant parfois être limité dans ses activités ou avoir besoin d'une aide minime, handicap relativement sévère : un SF 4 (les autres 0 ou 1), ou association de niveaux inférieurs dépassant les limites des degrés précédents. Capable de marcher 300m environ sans aide ni repos
5,0	Capable de marcher environ 200m sans aide ni repos, handicap suffisamment sévère pour entraver l'activité d'une journée normale. (En général un SF 5, les autres 0 ou 1, ou association de niveaux plus faibles dépassant ceux du grade 4,0)
5,5	Capable de marcher environ 100m sans aide ni repos ; handicap suffisamment sévère pour empêcher l'activité d'une journée normale.(En général un SF 5, les autres 0 ou 1, ou association de niveaux plus faibles dépassant ceux du grade 4,0)
6,0	Aide unilatérale (canne, canne anglaise, béquille) constante ou intermittente nécessaire pour parcourir environ 100m avec ou sans repos intermédiaire. (En général association de SF comprenant plus de 2 SF 3+)
6,5	Aide permanente bilatérale (cannes, cannes anglaises, béquilles) nécessaire pour marcher 20m sans s'arrêter. (En général association de SF comprenant plus de 2 SF 3+)

Score	Critères
7,0	Incapable de marcher plus de 5m même avec aide ; essentiellement confiné au fauteuil roulant ; fait avancer lui-même son fauteuil et effectue le transfert ; est au fauteuil roulant au moins 12h par jour. (En général association de SF comprenant plus d'un SF 4+ ; très rarement, SF 5 pyramidal seulement)
7,5	Incapable de faire plus de quelques pas ; strictement confiné au fauteuil roulant ; a parfois besoin d'une aide pour le transfert ; peut faire avancer lui-même son fauteuil mais ne peut y rester toute la journée ; peut avoir besoin d'un fauteuil électrique. (En général association de SF comprenant plus de d'un SF 4+)
8,0	Essentiellement confiné au lit ou au fauteuil, ou promené en fauteuil par une autre personne ; peut rester hors du lit la majeure partie de la journée ; conserve la plupart des fonctions élémentaires ; conserve en général l'usage effectif des bras. (En général SF 4+ dans plusieurs systèmes)
8,5	Confiné au lit la majeure partie de la journée ; garde un usage partiel des bras ; conserve quelques fonctions élémentaires. (En général SF 4+ dans plusieurs systèmes)
9,0	Patient grabataire ; peut communiquer et manger. (En général SF 4+ dans plusieurs systèmes)
9,5	Patient totalement impotent, ne peut plus manger ou avaler, ni communiquer. (En général SF 4+ dans presque tous les systèmes)
10	Décès lié à la SEP

Annexe B-Traitements de fond de la SEP

Table 36-Informations sur les traitements de la SEP classés selon leur ordre d'admission et leur date d'autorisation de mise sur le marché

Spécialité	Substance active	Posologie	Voie d'administration	AMM[a]	Nombre de jour de traitement/boite	Prix en euros[b]
			Traitements de 1ère ligne			
Betaferon®	Interféron bêta-1b	1 fois/2 jours	Sous-cutanée	30/11/1995	30	874,13
Cellcept®†	Mycophénolate mofétil	4 cps/jour	Per os	14/02/1996	13	58,47 à 194,42
Avonex®	Interféron bêta-1a	1 fois/semaine	Intra-musculaire	13/03/1997	28	820,78
Methotrexate®†	Methotrexate	3 à 10 cps/semaine	Per os	31/12/1997	30	4,55
Rebif®	Interféron bêta-1a	3 fois/semaine	Sous-cutanée	04/05/1998	28	727,38 à 835,98
Copaxone®	Acétate de glatiramère	1 fois/jour	Sous-cutanée	26/03/2004	28	829,20
Imurel®†	Azathioprine	3 cps/jour	Per os	22/04/2004	30	9,50 à 26,59
Extavia®	Interféron bêta-1b	1 fois/2 jours	Sous-cutanée	20/05/2008	30	759,76
Aubagio®	Teriflunomide	1 fois/jour	Per os	26/08/2013	28	827,00
Tecfidera®	Diméthyl fumarate	2 fois/jour	Per os	30/01/2014	7 ou 28	267,87 à 1017,68
Plegridy®	Peginterféron bêta-1a	1 fois/2 semaines	Sous-cutanée	18/07/2014	28	779,22
Ocrevus®	Ocrelizumab	1 fois/6 mois	Perfusion	08/01/2018	-	Prix libre

Spécialité	Substance active	Posologie	Voie d'administration	AMM[a]	Nombre de jour de traitement/boite	Prix en euros[b]
Traitements de 2ème ligne						
Mabthera®†	Rituximab	1 fois/4 semaines	Perfusion	02/06/1998	-	263,67
Tysabri®	Natalizumab	1 fois/4 semaines	Perfusion	27/06/2006	-	1724,94
Gilenya®	Fingolimod	1 fois/jour	Per os	17/03/2011	28	1905,69
Campath®	Alemtuzumab	1 fois/4 semaines	Perfusion	ATU [c]	-	-
Lemtrada®	Alemtuzumab	5 jours puis 3 jours 12 mois plus tard	Perfusion	12/09/2013 [d]	-	Prix libre
Traitements de 3ème ligne						
Novantrone®	Mitoxantrone	1 fois/4 semaines	Perfusion	31/01/1985	-	Prix libre
Elsep®	Mitoxantrone	1 fois/5 semaines	Perfusion	29/10/2003	-	Prix libre

[a] Autorisation de Mise sur le Marché. [b] Prix indiqués en 2015 sur la plate-forme en ligne VIDAL (263). [c] Autorisation Temporaire d'Utilisation. [d] AMM européenne.

†Spécialité non spécifique à la SEP.

Annexe C-Codes CIP des traitements de fond délivrés en ville

Table 37-Codes CIP utilisées pour le repérage des traitements de fond délivrés en ville

Spécialité	CIP 7	CIP 13	Nombre de jours par boîte
Traitements spécifiques à la SEP			
Avonex®	3432326	3400934323263	28
Avonex®	2160908	3400921609080	28
Betaferon®	3404092	3400934040924	30
Betaferon®	3772022	3400937720229	30
Rebif®	2219079	3400922190792	28
Rebif®	2219091	3400922190914	28
Rebif®	2219116	3400922191164	28
Rebif®	3474164	3400934741647	28
Rebif®	3474170	3400934741708	28
Rebif®	3508087	3400935080875	28
Rebif®	3508093	3400935080936	28
Rebif®	3759027	3400937590273	28
Rebif®	3931466	3400939314662	28
Rebif®	3931472	3400939314723	28
Rebif®	3474158	3400934741586	28
Rebif®	3508070	3400935080707	28
Extavia®	2699435	3400926994358	30
Extavia®	3865545	3400938655452	30
Copaxone®	3638401	3400936384019	28
Copaxone®	3585521	3400935855213	28
Gilenya®	4177876	3400941778766	28
Gilenya®	4177853	3400941778537	28
Aubagio®	2749989	3400927499890	28
Tecfidera®	2749788	3400927497889	7
Tecfidera®	2749794	3400927497940	28
Plegridy®	2794932	3400927949326	28
Plegridy®	2794955	3400927949555	28
Plegridy®	2794903	3400927949036	28
Plegridy®	2794926	3400927949265	28

Spécialité	CIP 7	CIP 13	Nombre de jours par boîte
Traitements non spécifiques à la SEP			
Imurel®	3641432	3400936414327	30
Imurel®	3641490	3400936414907	30
Imurel®	3052500	3400930525005	30
Methotrexate®	3067068	3400930670682	30
Cellcept®	3595293	3400935952936	13
Cellcept®	5616303	3400956163038	13
Cellcept®	5596167	3400955961673	13
Cellcept®	4950448	3400949504480	13
Cellcept®	3595258	3400935952585	13
Cellcept®	5596173	3400955961734	13
Cellcept®	4950454	3400949504541	13
Cellcept®	4951100	3400949511006	13
Cellcept®	5619394	3400956193943	13
Cellcept®	3595270	3400935952707	13
Cellcept®	3941625	3400939416250	13
Cellcept®	3961125	3400939611259	13
Cellcept®	3953315	3400939533155	13
Cellcept®	3957402	3400939574028	13
Cellcept®	4991803	3400949918034	13
Cellcept®	4945619	3400949456192	13
Cellcept®	4917959	3400949179596	13
Cellcept®	4929715	3400949297153	13
Cellcept®	4946599	3400949465996	13
Cellcept®	4946636	3400949466368	13
Cellcept®	3998950	3400939989501	13
Cellcept®	4991789	3400949917891	13
Cellcept®	4979392	3400949793921	13
Cellcept®	4945648	3400949456482	13
Cellcept®	2179167	3400921791679	13
Cellcept®	4921286	3400949212866	13
Cellcept®	4921211	3400949212118	13
Cellcept®	4924451	3400949244515	13
Cellcept®	3850041	3400938500417	13

Spécialité	CIP 7	CIP 13	Nombre de jours par boîte
Cellcept®	3850064	3400938500646	13
Cellcept®	3740039	3400937400398	13
Cellcept®	3957425	3400939574257	13
Cellcept®	3921261	3400939212616	13
Cellcept®	4914145	3400949141456	13
Cellcept®	4951092	3400949510924	13

Annexe D-Traitements de fond dans l'OFSEP

L'ensemble des résultats présentés dans cette annexe constituent des données non publiées étant le fruit d'un travail en parallèle de la thèse en collaboration avec l'OFSEP sur les séquences thérapeutiques dans la SEP.

Abstract soumis au 26th Annual Meeting of the European Charcot Foundation qui s'est tenu à Baveno, Italie, du 15 au 17 novembre 2018.

Complementary use of administrative and clinical data in studying disease-modifying therapies in MS in France in 2010-2015

Roux J[1,2], Casey R[3], Vukusic S[3,4,5], Leray E[1,2], on behalf of OFSEP investigators

[1] Univ Rennes, EHESP, REPERES (Pharmacoepidemiology and health services research) - EA 7449, F-35000 Rennes, France.

[2] Inserm CIC-P 1414, University Hospital of Rennes, F-35000 Rennes, France

[3] Univ Lyon, Université Claude Bernard Lyon 1, Hospices Civils de Lyon, Fondation EDMUS, OFSEP, Centre de Recherche en Neurosciences de Lyon, F-69000 Lyon, France

[4] Service de Neurologie, sclérose en plaques, pathologies de la myéline et neuro-inflammation, and Fondation Eugène Devic EDMUS contre la Sclérose en Plaques, Hôpital Neurologique Pierre Wertheimer, Hospices Civils de Lyon, Lyon, F-6977, France

[5] Centre des Neurosciences de Lyon, Inserm 1028 et CNRS UMR5292, Lyon, F-69003, France

Introduction

In the recent years, therapeutic management has evolved with the increasing variety of available disease-modifying therapies (DMTs). Most of data comes from clinical series, with the advantage of a strong data quality but the drawback of a potential recruitment bias. At the opposite, administrative data are exhaustive and population-based but suffer from a lack of clinical data. Considering both types of data may be relevant to better describe use of DMTs and highlight potential variations in clinical practice.

Objectives

To describe the use of DMTs in real-life settings of persons with MS (PwMS) in France over 2010-2015 period by using two different data sources: the OFSEP database ("Observatoire Français de la Sclérose en Plaques") and the national health insurance database (which covers 97% of the general population).

Methods

The two databases include respectively all PwMS from the OFSEP cohort seen in 2010-2015 (N=36,329) and all PwMS from the French national health insurance system ("Système National des Données de Santé") covered in 2010-2015 (N=112,745). DMTs accessible in both datasets were beta-interferons, glatiramer acetate, fingolimod, teriflunomide, dimethylfumarate, natalizumab, azathioprine, mycophenolate mofetil, methotrexate while rituximab, alemtuzumab, cyclophosphamide and mitoxantrone were available only through OFSEP data. The proportion of PwMS receiving each DMT was calculated and compared between the two databases.

Results

In total, 60,559 (53.7%) PwMS in the SNDS and 23,187 (63.8%) in the OFSEP database received at least one DMT over 2010-2015. In addition, 1,002 more PwMS were identified as treated with rituximab, alemtuzumab, cyclophosphamide or mitoxantrone in OFSEP database only. Injectable DMTs were the most prescribed DMTs whatever the dataset (respectively 38.2% and 43.1% in SNDS and OFSEP). In addition, 23.1% of OFSEP patients were prescribed an oral medication versus 20.4% of SNDS patients.

Discussion

The use of multiple sources permitted to access complementary information on DMT use. Population-based datasets such as SNDS may be helpful to describe practices outside expert centers that mainly contribute to OFSEP.

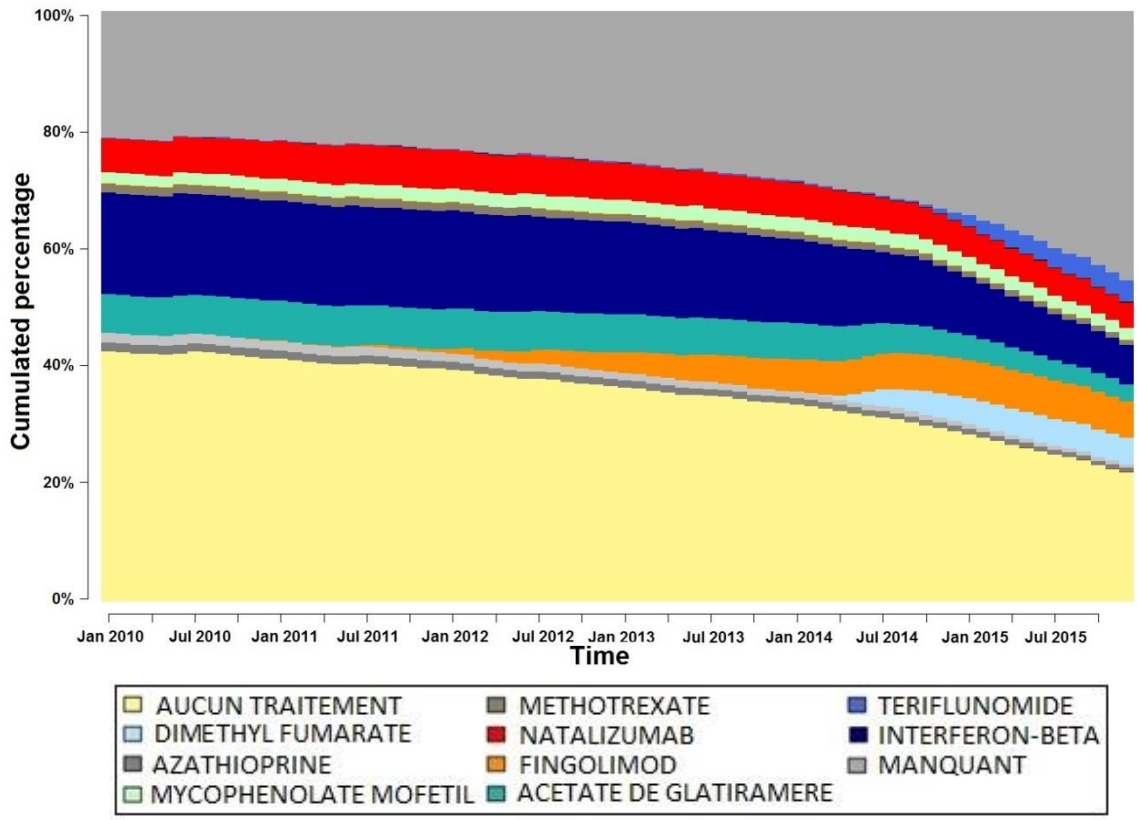


Figure 34-Chronogramme de la répartition des traitements sur la période d'étude 2010-2015 selon la substance active dans la cohorte de l'OFSEP (N=36 329)

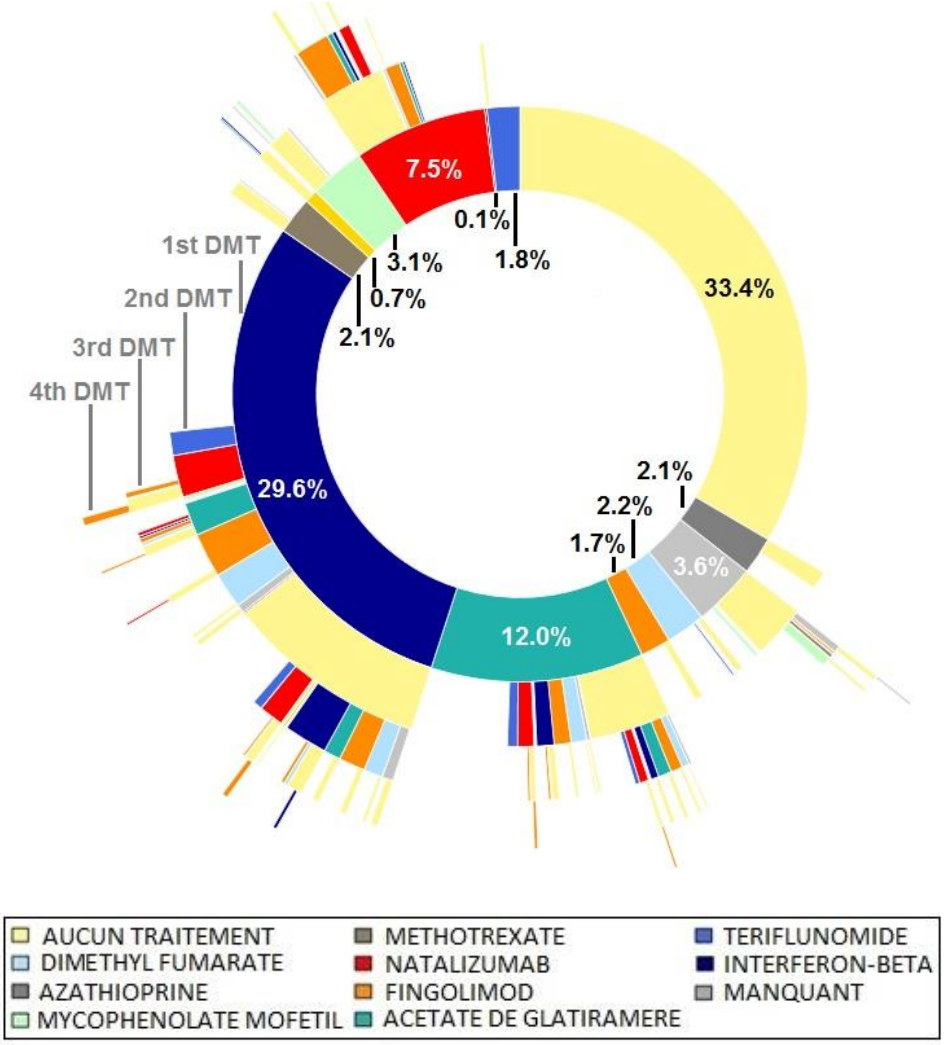


Figure 35-Sunburst des séquences thérapeutiques observées sur la période d'étude 2010-2015 dans la cohorte de l'OFSEP (N=36 329) ***(DMT=traitement de fond)***

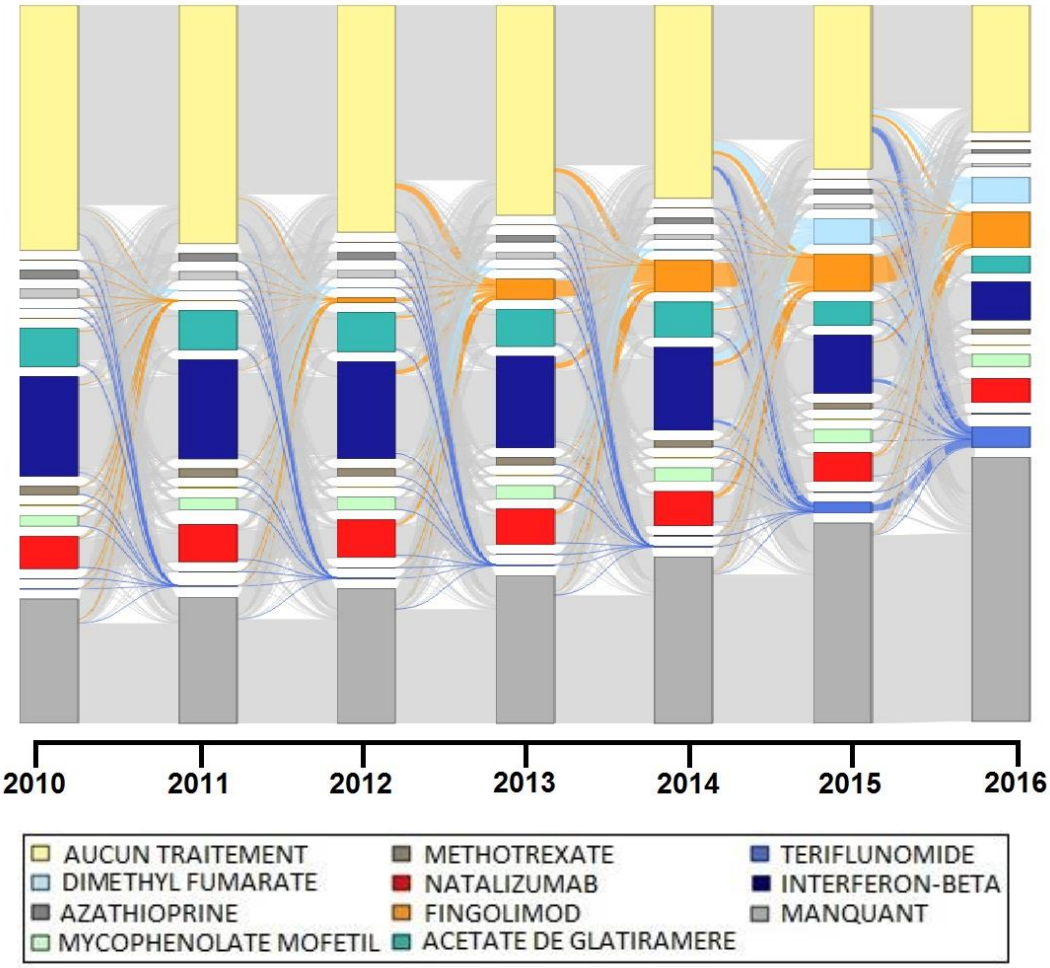


Figure 36-Diagramme de Sankey de la répartition des traitements sur la période d'étude 2010-2015 selon la substance active dans la cohorte de l'OFSEP (N=36 329)

Annexe E-Liste des codes LPP considérés pour l'estimation du score EDSS

Table 38-Codes LPP des différents matériels considérés pour le calcul du score EDSS « estimé »

Code LPP	Matériel	Libellé
1200764	Canne	Canne tripode ou quadripode réglable en hauteur, à l'achat
1270463	Canne	Canne en bois verni, a l'achat
1296787	Canne	Canne métallique réglable, avec appui antebrachial ou poignet en T, à l'achat
1225646	Déambulateur	Déambulateur, location hebdomadaire, < ou = 26 semaines
1260418	Déambulateur	Déambulateur, location hebdomadaire, > 26 semaines
1285619	Déambulateur	Déambulateur, achat
1204800	VHP manuel	VHP avec 2 accessoires, location hebdomadaire, < ou = 52 semaines
1210917	VHP manuel	VHP avec + de 2 accessoires, location hebdomadaire, < ou = 52 semaines
1232988	VHP manuel	VHP avec 2 accessoires, location hebdomadaire, > 52 semaines
1240976	VHP manuel	VHP avec + de 2 accessoires, location hebdomadaire, > 52 semaines
1255682	VHP manuel	VHP avec 1 accessoire, location hebdomadaire, > 52 semaines
1298680	VHP manuel	VHP avec 1 accessoire, location hebdomadaire, < ou = 52 semaines
4101353	VHP manuel	VHP, propulsion manuelle, non pliant, non réversible, à dossier non inclinable
4101956	VHP manuel	VHP propulsion manuelle, forfait annuel, réparation de roues
4107723	VHP manuel	VHP, propulsion manuelle, pliant, a dossier non inclinable
4113391	VHP manuel	VHP propulsion manuelle, dispositif de propulsion par moteur électrique
4118193	VHP manuel	VHP, propulsion manuelle, pliant, a dossier inclinable
4122473	VHP manuel	VHP, propulsion manuelle, pliant, a dossier non inclinable, articulation médiane
4134364	VHP manuel	VHP, propulsion manuelle, évolutif, pour enfant à partir de 18 mois
4142530	VHP manuel	VHP, propulsion manuelle, verticalisation manuelle assistée, usage d'une main
4156928	VHP manuel	VHP propulsion manuelle, forfait annuel, autres réparations dont sellerie
4164566	VHP manuel	VHP, propulsion manuelle, non pliant, réversible, a dossier non

		inclinable
4184899	VHP manuel	VHP, propulsion manuelle, non pliant, non réversible, a dossier inclinable
4195615	VHP manuel	VHP, propulsion manuelle, verticalisation manuelle assistée, usage des 2 mains
4222803	VHP manuel	VHP, tricycle à propulsion manuelle ou podale, > ou = 38 cm et < 90 cm
4278212	VHP manuel	VHP, tricycle à propulsion manuelle ou podale, < 38 cm, < 16 ans
4302152	VHP manuel	VHP, propulsion manuelle, forfait annuel, autres réparations dont sellerie
4307824	VHP manuel	VHP, propulsion manuelle, forfait annuel, réparation de roues
4101821	VHP électrique	VHP assistance électrique a la propulsion, mobitec, e motion
4108190	VHP électrique	VHP propulsion électrique, forfait annuel, hors roues et composants électriques
4111854	VHP électrique	VHP, propulsion électrique, a dossier non inclinable
4116165	VHP électrique	VHP propulsion électrique, forfait annuel, réparation de roues
4122250	VHP électrique	VHP propulsion électrique, forfait annuel, réparation composants électriques
4122757	VHP électrique	VHP, propulsion électrique, assise adaptée, dossier réglable, vérin pneumatique
4130136	VHP électrique	VHP, propulsion électrique, assise adaptée, dossier réglable, platines crantées
4147668	VHP électrique	VHP, propulsion électrique, a dossier inclinable
4168966	VHP électrique	VHP, propulsion électrique, verticalisateur électrique
4300348	VHP électrique	VHP, propulsion manuelle, dispositif de propulsion par moteur électrique
4307994	VHP électrique	VHP, propulsion électrique, forfait annuel, réparation composants électriques
4321630	VHP électrique	VHP, assistance électrique a la propulsion, invacare, alber e-motion
4348622	VHP électrique	VHP, propulsion électrique, forfait annuel, hors roues et composants électriques
4389845	VHP électrique	VHP, propulsion électrique, forfait annuel, réparation de roues
1231782	Soulève-malade	Soulève-malade, location hebdomadaire, < ou = 32 semaines
1278654	Soulève-	Soulève-malade, location hebdomadaire, > 32 semaines

	malade	
1202266	Lit médicalisé	Lit médical, achat lit et accessoires hors barrières, home médical, fortissimo
1203923	Lit médicalisé	Lit médical, accessoires, achat de 2 barrières, invacare, octave
1207973	Lit médicalisé	Lit médical, accessoires, achat de 2 barrières, invacare, sb 750
1213896	Lit médicalisé	Lit médical, forfait location lit et ses accessoires, invacare, sb 750
1220293	Lit médicalisé	Lit médical, forfait location lit et ses accessoires, medicatlantic, lit XXL
1224285	Lit médicalisé	Lit médical, accessoires, achat de 2 barrières, medicatlantic, lit XXL
1229970	Lit médicalisé	Lit médical, forfait location lit et ses accessoires, sotec, atlas
1230966	Lit médicalisé	Lit médical, forfait location lit et ses accessoires, invacare, octave
1233858	Lit médicalisé	Lit médical, achat lit et accessoires hors barrières, invacare, octave
1233976	Lit médicalisé	Lit médical, forfait location lit et ses accessoires, home medical, fortissimo
1235662	Lit médicalisé	Lit médical, lit spécifique, achat, lit et accessoires
1236704	Lit médicalisé	Lit médical, accessoires, achat de 2 barrières, sotec, atlas
1239393	Lit médicalisé	Lit médical, réparation du moteur électrique
1241266	Lit médicalisé	Lit médical, achat, maintenance préventive
1241763	Lit médicalisé	Lit médical, lit standard, location hebdomadaire, lit et accessoires
1248512	Lit médicalisé	Lit médical, achat lit et accessoires hors barrières, medicatlantic, lit XXL
1249523	Lit médicalisé	Lit médical, lit + 135 kg, location hebdomadaire, lit et accessoires
1252206	Lit médicalisé	Lit médical, achat lit et accessoires hors barrières, invacare, sb 750
1261808	Lit médicalisé	Lit médical, achat lit et accessoires hors barrières, sotec, atlas
1265427	Lit médicalisé	Lit médical, achat, lit spécifique, maintenance curative
1270316	Lit médicalisé	Lit médical, achat du lit et de ses accessoires à l'exclusion des barrières
1278281	Lit médicalisé	Lit médical, accessoires, achat de deux barrières sur justification médicale
1283879	Lit médicalisé	Lit médical, lit enfant, location hebdomadaire, lit et accessoires
1291583	Lit médicalisé	Lit médical, accessoires, achat de 2 barrières, home medical, fortissimo
1200072	Escarres II	Escarres, coussin pneumatique, classe ii, askle, optimair 10
1200250	Escarres II	Escarres surmatelas en mousse viscoélastique, classe ii. artois, svpe
1201700	Escarres II	Escarres, coussin pneumatique, s/classe ii,physipro,supracor stimulite slimline

1202272	Escarres II	Escarres matelas mixte, classe ii, recticel, visco-maxi
1202705	Escarres II	Escarres, coussin visco, classe ii,carpenter, carpenter viscolux, standard
1203633	Escarres II	Escarres, remplacement d'une des deux housses livrées avec le coussin
1203640	Escarres II	Escarres matelas en mousse viscoélastique, classe ii,invacare,softform viscomixt
1204035	Escarres II	Escarres, coussin pneumatique, s/classe ii, roho, high profile
1204220	Escarres II	Escarres surmatelas en mousse viscoélastique, s/classe ii, carpenter,cargum mixt
1204242	Escarres II	Escarres, coussin visco, classe ii, darulift, visco'mouss, standard.
1204288	Escarres II	Escarres, coussin visco, classe ii, syst'am, p 361 cv, technique.
1206548	Escarres II	Escarres surmatelas mousse viscoélastique, s/classe ii, askle, alova dermalon fr
1206904	Escarres II	Escarres, coussin visco, classe ii, escarius, visco, standard < 38 cm
1213092	Escarres II	Escarres matelas en mousse viscoélastique, classe ii, darulift, visco mouss
1214890	Escarres II	Escarres matelas en mousse viscoélastique, classe ii, askle, alova dm
1215808	Escarres II	Escarres, coussin visco, classe ii, syst'am, p 361 cp, technique.
1216682	Escarres II	Escarres matelas en mousse viscoélastique, s/classe ii, pharma ouest, 17/17300
1216937	Escarres II	Escarres matelas en mousse viscoélastique, s/classe ii, escarius, visco
1217664	Escarres II	Escarres, coussin visco, classe ii,carpenter, carpenter viscolux, technique
1218020	Escarres II	Escarres, coussin pneumatique, s/classe ii, roho, low profile
1218280	Escarres II	Escarres, surmatelas air motorise, askle, axtair one, forfait loc hebdomadaire
1221536	Escarres II	Escarres matelas mousse viscoélastique, s/classe ii, askle, alova vcl 74hep
1222429	Escarres II	Escarres, coussin pneumatique, s/classe ii, roho, quadtro
1222808	Escarres II	Escarres matelas en mousse viscoélastique, classe ii, syst'am, p161m
1223297	Escarres II	Escarres, coussin visco, classe ii, askle, alova galbe, technique.
1223423	Escarres II	Escarres matelas en mousse viscoélastique, classe ii, askle, alova promust
1223452	Escarres II	Escarres, coussin viscoélastique, classe ii, invacare, flo-tech lite visco
1223920	Escarres II	Escarres, coussin visco, classe ii, abc, memoba m-pur, technique.

1224078	Escarres II	Escarres surmatelas en mousse viscoélastique, s/classe ii, syst'am, p161m
1224084	Escarres II	Escarres, coussin visco, classe ii, syst'am, p 361 c, standard.
1225729	Escarres II	Escarres matelas en mousse structurée, s/classe ii, hill-rom, housse
1226410	Escarres II	Escarres, coussin pneumatique, s/classe ii, roho, enhancer
1227770	Escarres II	Escarres (sur)matelas pneumatique, s/classe ii, roho, mattress
1228857	Escarres II	Escarres, coussin pneumatique, classe ii, recfrance, profil'air.
1228917	Escarres II	Escarres matelas en mousse viscoélastique, s/classe ii, tempur, discovery
1229160	Escarres II	Escarres, coussin visco, classe ii, escarius, viscotech, technique.
1229785	Escarres II	Escarres, coussin visco, classe ii, invacare, flo-tech lite visco, standard
1230972	Escarres II	Escarres, coussin pneumatique, classe ii, askle, optimair 7
1231233	Escarres II	Escarres, coussin visco, classe ii, pharmaouest, kalli visco, standard
1231285	Escarres II	Escarres, matelas air motorise,medidev,sentry 1200,forfait livraison avec lit
1231865	Escarres II	Escarres, coussin visco, classe ii, pharmaouest, pharma form prima, < 38 cm
1233143	Escarres II	Escarres, coussin visco, classe ii, darulift, visco'mouss, standard > 45 cm
1233367	Escarres II	Escarres,coussin mousse viscoélastique,classe ii,carpenter,carpenter viscolux
1234728	Escarres II	Escarres matelas mixte, multiportance, classe ii, askle, alova
1237017	Escarres II	Escarres surmatelas mousse viscoélastique, classe ii,invacare,softform visco
1237218	Escarres II	Escarres surmatelas mousse viscoélastique, classe ii, pharma ouest, pharma form
1239022	Escarres II	Escarres, coussin visco, classe ii, pharmaouest, pharma form iskio, technique
1243510	Escarres II	Escarres, coussin viscoélastique, classe ii, invacare, flo-tech contour visco
1243680	Escarres II	Escarres, coussin visco, classe ii, darulift, visco'mouss, standard < 38 cm.
1244678	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, abc, memoba anatomique
1244900	Escarres II	Escarres surmatelas en mousse viscoélastique, s/classe ii, escarius,

		visco
1245034	Escarres II	Escarres, coussin visco, classe ii, syst'am, p 361 cl, standard < 38 cm.
1245459	Escarres II	Escarres, coussin pneumatique, classe ii, askle, kineris 7 france
1245844	Escarres II	Escarres, coussin visco, classe ii, syst'am, p 361 cl, standard.
1247754	Escarres II	Escarres, coussin visco, classe ii, abc, memoba m-pur, standard < 38cm.
1248713	Escarres II	Escarres,coussin visco,classe ii,invacare,flo-tech lite visco,standard >45cm
1249138	Escarres II	Escarres surmatelas en mousse viscoélastique, s/classe ii, carpenter, cargum
1250259	Escarres II	Escarres surmatelas mousse viscoélastique, s/classe ii, askle, alova promust
1251342	Escarres II	Escarres surmatelas en mousse viscoélastique, s/classe ii, tempur, discovery
1251460	Escarres II	Escarres, coussin visco, classe ii, escarius, visco, standard > 45 cm.
1251773	Escarres II	Escarres, coussin visco, classe ii, askle, alova ergonomique, standard.
1251840	Escarres II	Escarres matelas en mousse viscoélastique, s/classe ii, carpenter, cargum
1253766	Escarres II	Escarres, coussin visco, classe ii, syst'am, p 361 cl, standard > 45 cm.
1255096	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, syst'am, p361c
1255274	Escarres II	Escarres, coussin visco, classe ii,invacare, lo-tech contour visco, technique
1256546	Escarres II	Escarres, coussin visco, classe ii, askle, alova galbe, standard < 38cm.
1257155	Escarres II	Escarres, coussin visco, classe ii, escarius, visco, standard.
1258730	Escarres II	Escarres, coussin visco, classe ii, syst'am, p 361 ca, technique.
1259177	Escarres II	Escarres, coussin visco, classe ii, abc, memoba m-pur, standard > 45 cm.
1259214	Escarres II	Escarres, coussin visco, classe ii, darulift, visco'mouss, technique.
1261725	Escarres II	Escarres, coussin visco, classe ii, pharmaouest, prima plot, technique
1262699	Escarres II	Escarres, coussin visco, classe ii, syst'am, p 361 c, standard < 38 cm.
1262713	Escarres II	Escarres, coussin visco, classe ii, pharmaouest, kalli visco, standard > 45 cm
1265090	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, darulift, visco'mouss
1266421	Escarres II	Escarres matelas en mousse viscoélastique, s/classe ii, abc, mem 10

1267343	Escarres II	Escarres, matelas air motorise, medidev, sentry 1200 v2, forfait location jour
1267449	Escarres II	Escarres surmatelas en mousse viscoélastique, classe ii, recticel, visco-topi
1269342	Escarres II	Escarres matelas en mousse viscoélastique, classe ii, pharma ouest, vitaform
1270687	Escarres II	Escarres, coussin pneumatique, classe ii, syst'am, polyair p301c
1271379	Escarres II	Escarres, coussin visco, classe ii, syst'am, p 361 c, standard > 45 cm
1272344	Escarres II	Escarres, coussin visco, classe ii, pharmaouest, kalli visco, standard < 38 cm
1272545	Escarres II	Escarres, coussin pneumatique, classe ii, syst'am, polyair p302c
1273438	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, pharma ouest, pharma form
1273875	Escarres II	Escarres, coussin pneumatique, classe ii, askle, kineris 10 france
1276951	Escarres II	Escarres, coussin en visco, classe ii, askle, alova galbe, standard > 45 cm.
1277749	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, pharma ouest, prima form
1280036	Escarres II	Escarres, coussin visco, classe ii, askle, alova galbe, standard.
1280562	Escarres II	Escarres, coussin pneumatique, s/classe ii,physipro,supracor stimulite contoured
1283260	Escarres II	Escarres matelas en mousse multistrate, classe ii, a. bauer, airsoft duolight
1284927	Escarres II	Escarres surmatelas en mousse viscoélastique, s/classe ii, abc, mem 10
1285690	Escarres II	Escarres, coussin visco, classe ii, pharmaouest, pharma form prima, > 45cm
1286381	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, askle, alova
1289652	Escarres II	Escarres, matelas air motorise, medidev, sentry 1200, forfait livraison sans lit
1290129	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, escarius, visco
1291904	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, tempur, challenger
1292789	Escarres II	Escarres, coussin visco, classe ii,carpenter, carpenter viscolux, standard >45cm
1294506	Escarres II	Escarres matelas en mousse viscoélastique, classe ii, carpenter,

		cargumixt
1294570	Escarres II	Escarres matelas en mousse multistrate, classe iii, a. bauer, airsoft duo
1294653	Escarres II	Escarres, coussin visco, classe ii, pharmaouest, pharma form prima, standard
1296103	Escarres II	Escarres, matelas air motorise, medidev, sentry 1200, forfait location jour
1296793	Escarres II	Escarres, coussin visco, classe ii,carpenter, carpenter viscolux, standard <38cm
1296876	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, abc, memoba anatomique
1298119	Escarres II	Escarres, remplacement de la housse d'origine du (sur)matelas
1298734	Escarres II	Escarres, coussin en mousse viscoélastique, classe ii, askle, alova vcae/m
1217374	Lit et escarres II	Lit, compresseur et surmatelas pneumatique, location hebdomadaire

Annexe F-Article : Use of state sequence analysis for care pathway analysis: the example of multiple sclerosis

Check for updates

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–13
© The Author(s) 2018
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0962280218772068
journals.sagepub.com/home/smm
SAGE

Use of state sequence analysis for care pathway analysis: The example of multiple sclerosis

Jonathan Roux,[1,2,3] Olivier Grimaud[1,2] and Emmanuelle Leray[1,2,3]

Abstract

The concept of care pathways is increasingly being used to enhance the quality of care and to optimize the use of resources for health care. We here propose an innovative method in epidemiology that is derived from social sciences: state sequence analysis (SSA). This method takes into account the chronology of care consumption and allows for identification of specific patterns. A process for using SSA in the health area is proposed and discussed. The main steps are: data coding, measurement of dissimilarities between sequences (focusing on optimal matching methods and the choice of related costs), and application of a clustering method to obtain a typology of sequence patterns. As an example of its use in the health area, SSA was employed to analyse care pathways of a random sample of patients with multiple sclerosis. This sample has been selected from the main French healthcare database covering the period 2007 to 2013 ($n = 1\,000$). A five-cluster typology was obtained which allowed distinction of care consumption groups. Overall, about half of the patients had low care consumption, about one quarter had medium to high consumption, and another quarter had high consumption. We conclude that state sequence analysis is an innovative and flexible methodology that is worth considering in health care research.

Keywords

State sequence analysis, epidemiology, care pathways, multiple sclerosis, administrative data

1 Introduction

Longer life expectancies and medical advances have led to a steady increase in the number of individuals living with a chronic disease.[1] To take into account the necessary coordination of the multidisciplinary care required for such patients, the concept of a "care pathway" has been defined.[2] This concept is also known as an "integrated care pathway" or a "clinical pathway" and was first used in 1985.[3] Its aim is to "enhance the quality of care across the continuum by improving risk-adjusted patient outcomes, promoting patient safety, increasing patient satisfaction, and optimizing the use of resources," according to the definition of the European Pathway Association.[2,4] In France, the commonly accepted definition is the organization of overall and continuous care for patients closest of their place of residence, with particular attention being paid to the patient themselves and their own choices.[5,6] Thus, this concept reflects the need for the coordination of health care providers (HCP) (both medical and paramedical), especially between hospital and ambulatory care.

In France, being designated with long disease duration (LDD) status allows individuals with chronic diseases to receive full healthcare coverage. In total, 30 diseases are currently included in the list of LDD, multiple sclerosis (MS) being one of them. The estimated prevalence of MS in France was 151.2/100 000 individuals on 31 December 2012, i.e. a total of about 100,000 individuals nationally,[7] with a female to male ratio ranging from 2:1 to 2.5:1.[8]

[1]METIS Department, EHESP French School of Public Health, Rennes, France
[2]UPRES EA 7449 REPERES Pharmacoepidemiology and Health Services Research, Faculty of Medicine, University of Rennes 1 & EHESP French School of Public Health, Rennes, France
[3]INSERM CIC-P 1414, CHU of Rennes, 2 Rue Henri le Guilloux, Rennes, France

Corresponding author:
Jonathan Roux, Département MéTiS – Méthodes quanTitatives en Santé publique EHESP – École des hautes études en santé publique 15 Avenue du Professeur-Léon-Bernard CS 74312 35043 Rennes Cedex, France.
Email: jonathan.roux@ehesp.fr

MS usually starts between 20 and 40 years of age, and most patients typically live with the disease for several decades. The life expectancy of individuals with MS is known to be reduced by 6–7 years compared to the general French population matched for age and gender[9] while the level of disability increases with time. MS symptoms are variable and unpredictable, and each individual's symptoms can change or fluctuate over time. The most common symptoms of this neurological disease are fatigue, difficulty with walking, as well as vision and bladder problems. MS-specific therapeutic options have undergone extensive development since the 2000s, as has the understanding of the importance of comprehensive MS care. The neurologist/general practitioner (GP) dyad is core component of MS care. Indeed, only neurologists are allowed to prescribe MS-specific disease-modifying therapies (DMTs) whereas GPs are most often involved in the day-to-day care. The aim for each patient in this context is "to get the right care provided by the right medical practitioners, at the right time, and at the right place." From a public health point of view, the item "at the right price" can be added. Due to the high cost of DMTs (ranging from €700 to €2 000 per month) and the high need for care, the financial cost of MS has increased over time on a per patient basis. This increase has resulted in a cost to society of 1.3 billion Euros in France in 2013.[10,11] In light of the number of individuals in France with MS, the costs of the disease, and the patients' need for complex and multidisciplinary care, the French National Authority for Health is currently working on a recommended care pathway in MS to guarantee appropriate and safe care for all patients. Although an exact overview of care consumption would greatly assist formulating such recommendations, this is currently lacking. Up to now, French studies of care consumption in MS have mainly focused on costs[11,12] and not on the type, amount, and chronology of the care, i.e. the key elements that constitute the care pathway.

The care pathway is unique for each patient due to the choices made by both the HCP and the patient themselves. To date, studies focusing on individual pathways have mainly remained descriptive, without taking into account the possible evolution of care consumption over time.[13,14] At the population level, descriptive data can be summarized and individuals can be categorized according to their level or the amount of care consumption. Such results will be helpful to formulate care recommendations, as well as to anticipate and foresee resources that will be needed in the future. In order to generate such categories, an appropriate methodology taking into account the temporal dimension of care pathways should be used. To our knowledge, the methods used up to now have been derived from informatics, and they have been used to study in-hospital pathways by mining multi-dimensional itemset sequential patterns and symbolic data analysis.[15,16] However, these methods are not commonly used in statistics and epidemiology. Moreover, a requirement for specific knowledge or skills in programming may limit their use in the study of care pathways, including ambulatory care.

In this paper, we propose an innovative and easy-to-apply method in epidemiology that is derived from social sciences: the state sequence analysis (SSA). This method is commonly used to identify patterns in work trajectories and to study transitions from education to work[17,18] or career patterns.[19,20] It will allow the chronology of care consumption to be assessed as well as the identification of specific patterns. To our knowledge, this methodology has been used only once in the field of health care, specifically in an attempt to identify disparities in prenatal care pathways.[21] However, this study did not provide a clear description of the analytical process. Therefore, the objective of the present manuscript was to highlight the potential and the merits of SSA in the analysis of care pathways and to propose a reference process, from data coding to clustering, through a specific application in regard to multiple sclerosis.

2 Methods

2.1 The study population

The data were obtained from the French permanent sample of health insurance care holders ("Echantillon Généraliste des Bénéficiaires (EGB)" – Agreement no.143 from the French Institute on Health Data). It is a representative random sample of 1/97th of the French national health insurance system ("Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM)"). EGB gathers data on out-of-hospital reimbursed care consumption, i.e. consultations and home visits with private HCPs, drugs dispensed, LDD status, etc. For each patient, the exact day of execution and type of care are recorded since the date of their inclusion in the database. This sample is updated periodically, i.e. new individuals are included every three months. As the care consumption by these new members is only recorded after their inclusion in the EGB, there is no retrospective data for them. Moreover, a specific characteristic of this database is the fact that patients are not removed from the sample even if they do not receive any care. There are hence no missing values regarding care consumption by patients during follow-up, and only the death of a patient results in its removal from the database. This database was linked to the French hospital discharge database ("Programme de Médicalisation des Systèmes

d'Information (PMSI)''), which compiles data on private and public in-hospital care consumption. The database contains the start and end dates of hospital admissions (including day hospital admissions) and their related International Classification of Diseases, 10th version (ICD-10) diagnoses. Outpatient consultations are also included in PMSI. However, the medical speciality is not systematically specified, leading to the exclusion of these consultations from the analysis (as we were only interested in specific specialties, as explained in the next section). Only a small amount of information regarding the patient themselves is available in these two databases, i.e. the year in which they were born, the month and the year of their death, as well as their LDD status and the corresponding year that it started, if applicable.

A four-criteria algorithm has been developed by the French health insurance system for salaried workers ''*Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés*'' to identify MS cases.[7] Patients are considered to have MS if they meet at least one of the following criteria: having MS LDD status, being prescribed at least one DMT specific for MS (e.g. beta-interferon, glatiramer acetate, fingolimod, or natalizumab), having at least one hospital admission with an ICD-10 diagnosis code G35,[22] or having received at least one disability pension payment for MS. We used these criteria to identify patients with MS in the EGB database from 1 January 2007 to 31 December 2012. Data regarding these patients were then extracted from 2007 to the end of 2013. In total, one thousand MS patients were identified, of whom 648 were prevalent cases in 2007 (and a further 42 to 94 cases identified annually from 2008 to 2012).

2.2 Data collection

The following elements were included in the care pathways: consultations with GPs (including home visits), consultations with private neurologists and physical medicine and rehabilitation (PMR) physicians (these are considered to be the two main medical specialities involved in MS care in France, in addition to GPs). MS-related hospital admissions (main or related diagnosis of G35, except in-hospital DMT injection) were also considered. Based on the drugs dispensed by pharmacists, the duration of each DMT was determined as the product of the number of packages and the corresponding posology. DMT treatments were included in the analysis if administered for at least three consecutive months, allowing for one month with no medical treatment. As DMTs can only be prescribed by a neurologist, patients receiving DMT without a consultation with a private neurologist were assigned one annual outpatient consultation as long as the treatment was ongoing.

2.3 The SSA method

Using SSA terminology, a care pathway corresponds to a sequence during which successive ''states'', i.e. the value taken by the variable at a time t,[23] are encountered. The term ''sequence'' will hence be used to refer to the entire care pathway. In our case study, states are defined as levels of care consumption during each period of time. Thus a sequence may start with a period of high care contact due to disease activity for instance, followed by a period of lower consumption with HCPs related to a period of stable disease. The general principle of SSA is to compare sequences (i.e. individual pathways of patients) with respect to the succession of their component states. In general, SSA can be broken down into three main steps: the coding of data, the measure of similarity (or dissimilarity) between sequences, and the clustering of sequences. In this work, we focused on dissimilarity-based methods, especially the one that uses optimal matching (OM) algorithms as introduced by Abbott in social sciences.[24]

2.3.1 The first step of SSA: identification of states and sequences

The first step of SSA needs to address preliminary questions. First, the study period and the time unit (i.e. day, month, trimester, year, etc.) have to be set. These two items will define the onset and the end of the sequences and the maximum number of states constituting a sequence. Dlouhy et al. recommended using sequences with at least 25 states. They also showed that the number of sequences only has a small effect on the final typology.[25] It is known that missing values can occur at the beginning of some sequences, due to data availability and the data source. Depending on the research question, it can be worth deleting these initial missing states so that the sequences become left-aligned. If the user does not choose to left-align the sequences, some sequences will have a specific state at their beginning coding for the missing value. In this case, the user may opt to split the analysis according to the length of sequences (e.g. short vs. long, or sequences starting in the same year) or simply keep all the sequences in a single analysis. However, the risk of not splitting the sequences into several groups with a homogeneous onset could lead to the creation of a cluster devoted to patients with initial missing states, since their

sequences will be more similar. This last scenario may not be meaningful if the main characteristic of this cluster is linked to a delayed start and is not linked to specific patterns of care.

In the present application to MS, inclusion criteria offered the opportunity to identify prevalent cases. Patients were thus not at the same stage of the disease at the start of the study, and there was not a clear indication of the time of the clinical onset of MS. As the disease duration was unknown, the calendar year was considered as the time unit for the analysis. The individual seven-year care pathway was hence obtained by placing the seven annual values of the variable characterizing individual care consumption end-to-end, i.e. one value by year. Sequences could not be left-aligned for individuals entering the study after 2007. To avoid the creation of a cluster devoted to patients with a delayed entry in the cohort and to facilitate understanding of the SSA, we chose to focus the present analysis on patients entering the cohort in 2007 ($n = 648$). The other patients entering the cohort in 2008 or later were analysed separately ($n = 352$) (data not shown).

The next step is to identify a discrete list of values that each state will have to fall into. This is referred as the alphabet in SSA terminology. Choosing the length of the alphabet is important, as it should remain relatively limited. Indeed, the more states there are, the more complicated the final interpretation is likely to be.[20] In our example, a variable characterizing individual annual care consumption was created. It corresponded to the sum of the annual number of consultations with HCPs (i.e. GPs, private neurologists and PMR, and outpatient consultations with neurologists) and MS-related hospital admissions. These annual values of care consumption followed a Poisson distribution with a wide range. We chose to categorize these into five states, using annual quartiles ([0-Q1], [Q1-Q2], [Q2-Q3] and >Q3) plus an extra group with no consumption at all (0). In general (not in our case study due to the characteristics of the database), missing values inside sequences can be considered using an additional state, coded as non-attributed (NA). However, it is recommended to exclude sequences when more than 30% of the states are missing.[25]

A sequence ends when the follow-up ends, which induces sequences of different lengths and right-truncated sequences. Depending on the setting, it may be worth considering death as a state of the alphabet. In this case, creating an additional devoted state (''Death'') could be an option. However, the use of such a state may favour a high level of similarity between sequences ending with this specific state. In our study, we opted not to create a specific state to code for death, and we decided to stop the sequence of a patient the year after their death. This created right-truncated care pathways and avoided the creation of a specific cluster for patients who had died. Indeed, transitions from any state of care to death were rare in the database. The corresponding substitution costs (described thereafter) would have been higher than the others, and thus explained the gathering of the sequences ending with death in a specific cluster.

2.3.2 The second step of SSA: measuring dissimilarity between sequences

Once sequences are created, the main component of SSA is the two-by-two comparison of sequences, which requires choosing a dissimilarity measure. Several measures of dissimilarity can be used in SSA. Some were derived from bioinformatics, such as OM, the Levenshtein II distance, and the Hamming distance. Others were developed in social sciences, such as the longest common subsequence, and the number of matching subsequences.[26,27] In this paper we describe OM, which is the most commonly used method for measuring dissimilarity,[28,29] since it can be adjusted for different research issues. The principle of OM is to assign a value to the number of operations necessary to render two sequences strictly similar. As an example, one can consider two sequences made of three consecutive states and that only the last state differs between them. A simple substitution of this state in one sequence will suffice to make the two similar. More operations would be necessary when there is more of a difference between the sequences. OM uses three elementary operations, which are insertion, deletion, and substitution. Insertion corresponds to the addition of a state in a sequence, whereas deletion is the opposite, i.e. the removal of a state. These two operations are joined into a single concept called an ''indel'', since both operations are warping time. Substitution consists of replacing a state with another one at a time t. To quantify the dissimilarity between two sequences, each of those three operations is assigned a specific cost, which can be constant or which can vary according to the states. The setting of each cost constitutes a crucial step of SSA.[30] Indeed, costs are a major parameter in optimal matching and hence affect clustering results.[31] The dissimilarity between two sequences is then defined as the minimal cost to transform one sequence into the other. This measure therefore depends on the number and the type of elementary operations to transform one sequence into the other and on their costs.[29] To test all of the combinations and to compute the dissimilarity, the Needleman–Wunsch algorithm[32] is used in the main statistical software.

Substitution costs can be chosen either theoretically or empirically. Although in both cases they must respect triangle inequality to ensure that the OM distance and subsequent clustering methods can be computed.[29]

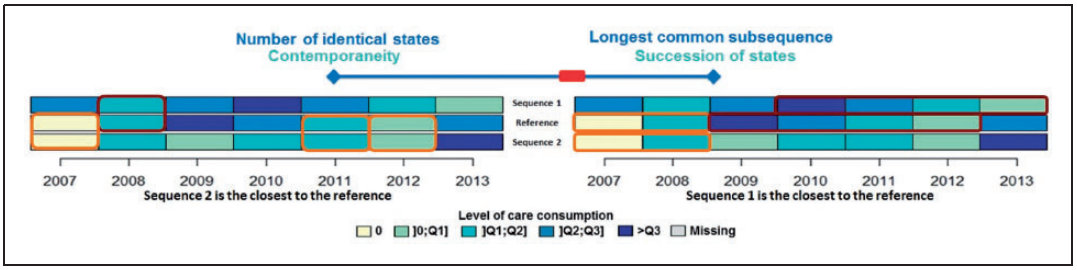


Figure 1. Patterns identified by optimal matching (adapted from Lesnard[31]).

The matrix of substitution costs, which is usually symmetrical, contains a value for each substitution of a state by another one and zeroes on the diagonal. Indeed, substituting a state by itself does not change the sequence. If no theoretical costs are available, substitution costs can be estimated empirically using the transition rates observed in the dataset.[20,33] In this case, the cost $c_{i->j/j->i}$ for a substitution of the state s_i by the state s_j, and inversely, equals

$$c_{i->j/j->i} = 2 - p(s_i|s_j) - p(s_j|s_i)$$

where $p(s_i|s_j)$ and $p(s_j|s_i)$ are the probability of observing the transition s_j towards s_i and s_j towards s_i in the dataset, respectively.[33] In regard to the cost of substituting a missing state with another state, it is usually set at 2 when using transition-based costs. The indel cost is set by the user, and most of the time it is fixed to a specific value, although it can also be state-dependant,[29] i.e. one value of indel cost for each state of the alphabet. Attention should be paid to the choice of the indel cost, since indels favour the order of events whereas substitutions preserve contemporaneity,[31] as presented in Figure 1. Moreover, indels warp time, while substitutions distort the succession of states in the sequence.[31] Using only indels by setting a very low cost for this operation – this measure is then called the Hamming distance – makes two sequences similar according to their identical states. In contrast, using only substitutions by setting an indel cost greater than the maximum substitution cost – this measure is called the Levenshtein II distance – makes two sequences similar according to their longest common subsequence.[31] (For a full explanation of the process whereby the distance matrix is calculated, see Pollock[34]).

As no theoretical knowledge regarding costs was available for our example, substitution costs were estimated empirically using the observed transition rates. Moreover, we aimed to compare care pathways in terms of the long-term care consumption and we did not want to focus on outstanding events but rather on the whole sequence.

2.3.3 The third step of SSA: clustering sequences

The dissimilarity matrix is used to cluster sequences and to create a typology of patterns according to care consumption. Two main methods can be used for this: agglomerative hierarchical clustering and k-medoids algorithms (similar to k-means). Depending on the predefined costs, the measure used to compare sequences can either be a dissimilarity or a distance (to view axioms of these measures, see Elzinga[35]). It can therefore influence the choice of the criterion to perform agglomerative hierarchical clustering. Ward's criterion is the most used in social sciences[25] with dissimilarity measures,[36,37] and it produces the best results in clustering.[25] Unweighted pair-group using arithmetic averages (UPGMA) and weighted pair-group using arithmetic averages (WPGMA) methods can also be considered for agglomerative clustering.[25] In regard to the k-medoids algorithm, the one that we favour is the partitioning around medoids (PAM) algorithm.[38] In order to identify the best clustering, two methods can be applied successively, such as an agglomerative hierarchical algorithm followed by k-medoids.[38]

The optimal number of clusters can be chosen either theoretically or empirically. When the Ward's criterion is used, the fall in inertia can permit to set the number of clusters, otherwise quality criteria can be used. We propose to use a combination of the two main criteria[38]: the weighted average silhouette width (ASWw)[38,39] and the Hubert's C index.[40] The ASWw measures the overall consistency of the clusters. It fluctuates between −1 and 1, a value close to 1 indicating high inter-cluster distances and strong intra-cluster homogeneity. The Hubert's C index, which varies between 0 and 1, reflects the difference between the obtained clustering and the best theoretically possible clustering. Hence a value close to 0 reflects a good clustering. Maximisation of the ASWw and minimisation of Hubert's C index allow for the optimal number of clusters to be set. The choice of these two parameters is in line with a recent social sciences paper, in which three criteria were selected for the partition,[41] namely the ASWw, Hubert's C index, and the point biserial correlation (PBC). We chose to forego the PBC as it usually led to the same choice of the number of clusters as the two other criteria.

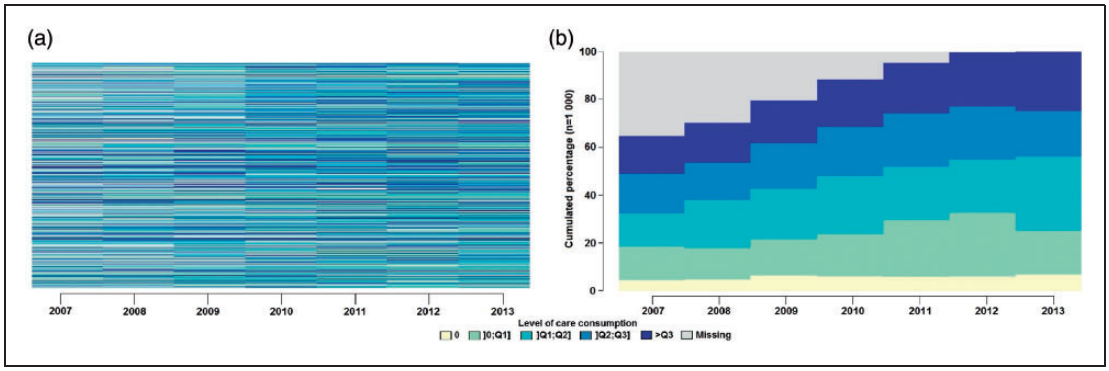


Figure 2. Index plot (A) and chronogram (B) of the set of care pathways ($n = 1000$).

Here, we used agglomerative hierarchical clustering analysis with Ward's criterion on the dissimilarity matrix to create homogeneous groups. The optimal number of clusters was set using quality parameters. Once the clusters were obtained, each group was described according to individual characteristics and care consumption.

2.3.4 Representing sequences

One major advantage of SSA is the wide option of graphical displays available to describe and communicate results of clustering. The most frequently used graphs are the index plot and the chronogram (also referred to as the state distribution plot) (Figure 2). The index plot is the superposition of each sequence of the set which allows the entire individual longitudinal succession of states of each patient to be seen at once.[33] Hence, the x-axis represents the time (either calendar-based or the duration, depending on the previous choice) and the y-axis represents the id of the sequence, which is usually not represented. By contrast, the individual information level is lost when using the chronogram, synthesizing the information. It displays the general pattern of the whole set of sequences,[33] for each time unit, on the x-axis, while the cumulative proportion of patients in the different states is presented on the y-axis. Other types of graphs can be used to represent sequences and results of SSA, such as the mean duration in each state, the sequence composed of the modal state (i.e. the most common state at each time unit), or the most frequent sequences (Supplementary Figure 1).

All computational and statistical analyses were performed with R v3.2.3 software.[42] The sequence analysis and the clustering used the *TraMineR* library v1.8-11, the *WeightedCluster* library v1.2, and the *cluster* library v2.0.3.[33,38,43]

3 Results

3.1 Population characteristics

Based on our selection criteria, one thousand MS patients were identified during the whole study period, of whom 648 were identified in 2007.Their characteristics are summarized in Table 1. The majority of these patients were women (71.1%) and the median year of birth was 1963 (ranging from 1914 to1997); that is to say, the patients had a median age of about 44 years at the start of the study.

The 648 patients identified in 2007 were older than the 352 identified in 2008 or later ($p < 0.001$), as they had a median age of 46 years in 2007. Of these patients, 511 (78.9%) had LDD status for MS at the beginning of the follow-up, with a median LDD duration of 6.9 years (ranging from 0.0 to 29.2 years). In total, 55 deaths (8.5%) occurred over the follow-up period, at a median of 63.5 years of age. Over the seven-year study period, 639 patients (98.6%) consulted at least once with a GP and 362 patients (55.9%) consulted at least once with a private neurologist. The number of GP consultations per person-year was 8.0 (7.1for men and 8.3 for women) while the number of consultations with a neurologists was 1.1 (1.0 for men and 1.1 for women).

3.2 Complete care pathway clustering

To compute the dissimilarity matrix between care pathways, the costs of substitution were based on observed transitions in the dataset of the one thousand patients. Using the complete dataset permitted to take into account all available information regarding care consumption (Table 2). Indel costs were set at 0.9 to approach the

Table 1. Population characteristics according to the year of identification ($n = 1000$).

	Identified in 2007 $N = 648$	Identified in 2008 or later $n = 352$	Total $n = 1000$	p-Value[a]
No. of women (%)	473 (73.0%)	238 (67.6%)	711 (71.1%)	0.086
Birth year[b]	1961 (1914–1991)	1966 (1924–1997)	1963 (1914–1997)	<0.001
No. of LDD for MS[c]	589 (90.9%)	249 (70.7%)	838 (83.8%)	<0.001
Time since MS-LDD beginning[b,c,d] (years)	12.7 (0.0–34.6)	6.5 (0.2–33.8)	11.6 (0.0–34.6)	<0.001
At least one other LDD[c,e] (%)	186 (28.7%)	92 (26.1%)	278 (27.8%)	0.429
Follow-up[b] (years)	6.9 (0.0–7.0)	3.5 (0.0–6.0)	6.8 (0.0–7.0)	<0.001
No. of deaths (%)	55 (8.5%)	16 (4.5%)	71 (7.1%)	0.029
No. of deaths per person-year (/1000)	12.9	12.9	12.9	
At least one DMT prescription[f] (%)	368 (56.8%)	171 (48.6%)	539 (53.9%)	0.015
Proportion of follow-up under treatment[b,g] (%)	61.6 (3.3–100.0)	66.0 (7.6–100.0)	63.0 (3.3–100.0)	0.218
Care consumption				
Consultations with a GP[b,h]	43.0 (0.0–350.0)	17.0 (0.0–266.0)	32.0 (0.0–350.0)	<0.001
Consultations with a GP per person-year	8.0	6.5	7.7	
Consultations with a neurologist[b,i]	6.0 (0.0–48.0)	2.0 (0.0–19.0)	4.0 (0.0–48.0)	<0.001
Consultations with a neurologist per person-year	1.1	1.0	1.1	
At least one MS-relatedhospital admission[j] (%)	307 (47.3%)	166 (47.2%)	473 (47.3%)	1.000
MS-related hospital admissions per person-year	0.24	0.31	0.26	

[a]$P < 0.05$ comparing patients identified in 2007 and patients identified in 2008 or later using the Kruskal-Wallis, Pearson's chi-square, or Fisher's exact test if needed.
[b]Median (minimum–maximum).
[c]LDD: long disease duration.
[d]Calculated at the date of the most recent information (i.e. Dec. 31, 2013 or the date of the patient's death).
[e]LDD other than MS.
[f]DMT: disease-modifying therapy.
[g]For patients with at least one DMT use.
[h]GPs: general practitioners, Number of consultations and home visits, total per patient from 2007 to 2013.
[i]Number of consultations, home visits, and imputed outpatient consultations, total per patient from 2007 to 2013.
[j]MS-related hospital admissions, total per patient from 2007 to 2013 (except monthly DMT injections).

Table 2. Matrix of observed transition rates in the dataset ($n = 1000$).

	->0	->]0;Q_1]	->]Q_1;Q_2]	->]Q_2;Q_3]	->>Q_3
0 ->	0.490	0.232	0.105	0.108	0.064
]0;Q_1] ->	0.069	0.453	0.313	0.111	0.054
]Q_1;Q_2] ->	0.015	0.245	0.402	0.258	0.080
]Q_2;Q_3] ->	0.009	0.104	0.304	0.345	0.238
>Q_3 ->	0.007	0.020	0.091	0.223	0.658

Levenshtein II distance, and thereby compare pathways according to the longest common subsequence. The partitioning of the 648 care pathways led to a typology of five clusters (Figure 3); for the best-quality partitions, the ASWw and HC were equal to 0.238 and 0.102, respectively. This partition can be considered of weak quality.

Among the five clusters, clusters 4 and 5 were predominant, with 180 (27.8%) and 188 (29.0%) patients, respectively (Table 3). Based on the mean duration in each state, these two clusters corresponded to patients with overall high care consumption. Cluster 3 comprised 128 patients (19.8%) with medium consumption. By contrast, patients in cluster 2 ($n = 124$ (19.1%)) had low care consumption, and the 28 patients (4.3%) in cluster 1 had no consumption during nearly half of the follow-up period.

3.3 Characteristics of the final clustering

The patient characteristics and the care consumption according to the clusters obtained are summarized in Table 4.

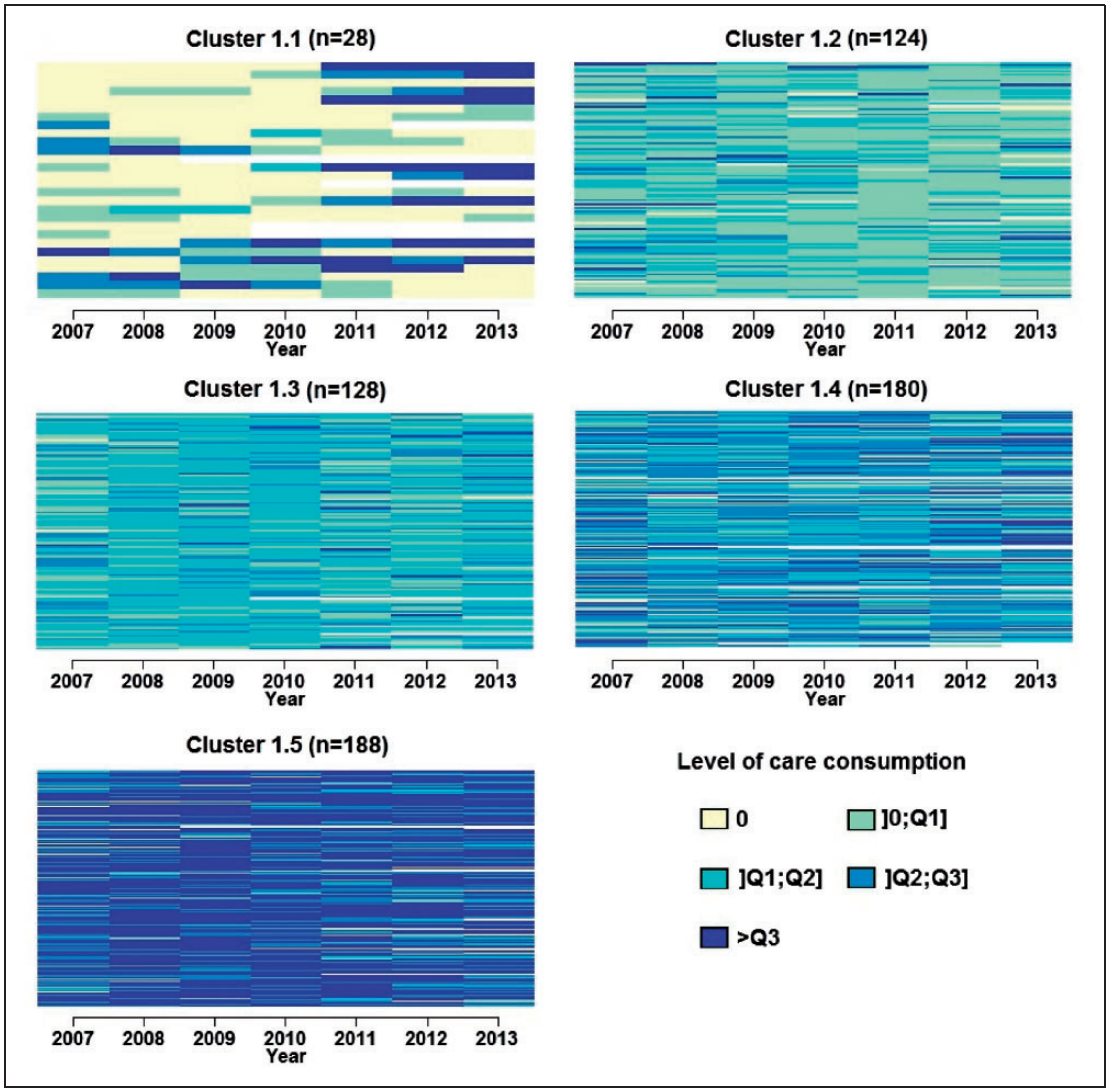


Figure 3. Index plots of clusters obtained for the care pathways of patients identified in 2007 after clustering with indel costs fixed at 0.9 and transition-based substitution costs ($n = 648$).

Table 3. The mean duration in years (part of total duration) in each state among the clustering of patients identified in 2007 ($n = 648$).

Cluster state	*n* (%)	0	[0–Q1]	[Q1–Q2]	[Q2–Q3]	> Q3
1.1	28 (4.3%)	*3.46 (49.4%)*	1.18 (16.9%)	0.14 (2.0%)	0.64 (9.1%)	0.93 (13.3%)
1.2	124 (19.1%)	0.31 (4.4%)	*4.06 (58.0%)*	1.86 (26.6%)	0.55 (7.9%)	0.19 (2.7%)
1.3	128 (19.8%)	0.09 (1.3%)	1.39 (19.9%)	*4.07 (58.1%)*	1.20 (17.1%)	0.16 (2.3%)
1.4	180 (27.8%)	0.13 (1.9%)	0.57 (8.1%)	1.76 (25.1%)	*2.95 (42.1%)*	0.95 (13.6%)
1.5	188 (29.0%)	0.10 (1.4%)	0.24 (3.4%)	0.45 (6.4%)	1.37 (19.6%)	*4.71 (67.3%)*

Note: The predominant state for each cluster is indicated in italics. The yearly consumption was classified into five groups by quartile of distribution: no consumption (0), less than the first quartile ([0−Q1]), between the first quartile and the median included ([Q1−Q2]), between the median and the third quartile included ([Q2−Q3]), and greater than the third quartile (> Q3).

As expected, care consumption differed from one cluster to another, and it increased as the cluster number went up. The number of consultations with a GP and a neurologist increased from 2.9 to 14.5 over the study period and from 0.2 to 1.3 per patient-year (Table 4). In cluster 1, the follow-up duration was the shortest ($p < 0.001$), which was probably related to the higher number of deaths. Moreover, the number of patients who used at least one DMT was

Table 4. Characteristics of patients identified in 2007 by clusters ($n = 648$).

Characteristics	Cluster 1.1 $n = 28$	1.2 $n = 124$	1.3 $n = 128$	1.4 $n = 180$	1.5 $n = 188$	p-Value[a]
No. of women (%)	23 (82.1%)	72 (58.1%)	92 (71.9%)	142 (78.9%)	144 (76.6%)	<0.001
Birth year[b]	1958.5 (1930–1989)	1965.5 (1929–1991)	1960 (1923–1988)	1957.5 (1914–1986)	1959 (1916–1987)	<0.001
No. of LDD for MS[c]	24 (85.7%)	110 (88.7%)	120 (93.8%)	167 (92.8%)	168 (89.4%)	0.373
Time since MS-LDD beginning[b,c,d] (years)	12.4 (1.0–26.2)	12.6 (1.0–28.6)	13.0 (3.0–34.6)	12.1 (0.6–34.0)	12.8 (0.0–34.3)	0.629
At least one other LDD[c,e] (%)	9 (32.1%)	19 (15.3%)	32 (25.0%)	48 (26.7%)	78 (41.5%)	<0.001
Follow-up[b] (years)	6.5 (0.0–7.0)	6.9 (0.0–7.0)	6.9 (0.0–7.0)	7.0 (0.0–7.0)	7.0 (0.0–7.0)	<0.001
No. of deaths (%)	6 (21.4%)	4 (3.2%)	4 (3.1%)	26 (14.4%)	15 (8.0%)	<0.001
No. of deaths per person-year (/1000)	37.4	4.7	4.6	23.4	11.8	
At least one DMT prescription[f] (%)	6 (21.4%)	78 (62.9%)	74 (57.8%)	95 (52.8%)	115 (61.2%)	0.001
Proportion of the follow-up under treatment[b,g] (%)	14.3 (3.6–59.0)	75.5 (3.3–100.0)	76.5 (4.4–100.0)	60.2 (3.3–100.0)	51.8 (4.4–100.0)	<0.001
Care consumption						
Consultations with a GP[b,h]	19.0 (0.0–73.0)	18.5 (0.0–51.0)	32.0 (10.0–54.0)	48.0 (0.0–96.0)	89.0 (31.0–350.0)	<0.001
Consultations with a GP per person-year	4.1	2.9	4.7	7.5	14.5	
Consultations with a neurologist[b,i]	0.0 (0.0–8.0)	6.0 (0.0–23.0)	7.0 (0.0–35.0)	5.0 (0.0–38.0)	7.0 (0.0–48.0)	<0.001
Consultations with a neurologist per person-year	0.2	0.9	1.0	1.1	1.3	
At least one MS-related hospital admission[j]	8 (28.6%)	61 (49.2%)	44 (34.3%)	84 (46.7%)	110 (58.5%)	<0.001
MS-related hospital admission per person-year	0.14	0.23	0.09	0.25	0.35	

[a]$P < 0.05$ comparing the five groups by the Kruskal–Wallis, Pearson's chi-square test, or Fisher's exact test.
[b]Median (minimum–maximum).
[c]LDD: long disease duration.
[d]Calculated at the date of the most recent information (i.e. Dec. 31, 2013 or date of the patient's death).
[e]LDD other than MS.
[f]DMT: disease-modifying therapy.
[g]For patients with at least one DMT use.
[h]GPs: general practitioners, Number of consultations and home visits, total per patient from 2007 to 2013.
[i]Number of consultations, home visits, and imputed outpatient consultations, total per patient from 2007 to 2013.
[j]MS-related hospital admissions, total per patient from 2007 to 2013 (except monthly DMT injections).

lower in this cluster than in the other clusters ($p = 0.001$). Patients in cluster 2 were significantly younger than those in the other clusters ($p < 0.001$). They were the most frequently treated for MS, with a median treatment duration that lasted about three-quarters of the follow-up period. Furthermore, this cluster comprised a nearly equal proportion of men and women, as opposed to the other clusters which comprised more women. Cluster 3 included a large portion of patients who were prescribed a DMT for most of the duration of the follow-up. Cluster 4 was the oldest one. In cluster 5 – which had the highest consumption – the patients were older, with significantly more comorbidities than those in other clusters ($p < 0.001$). These patients, who went to the hospital more often than the others, mostly received treatment for MS, although they were treated for a shorter period than those in clusters 2 to 4.

4 Discussion

The aim of the present study was to provide a reference process and recommendations regarding the use of state sequence analysis to analyse longitudinal care pathways. This process was illustrated by a case study of the care

provided to MS patients in France. Application of the proposed methodology revealed five patterns of consumption for the pathways starting in 2007. A similar clustering was obtained for pathways corresponding to patients with a delayed entry in the cohort (data not shown). All of the clusters were mostly driven by a particular dominating state.

As expected, patients with MS overall appear to be more likely than the general population to consult with a HCP. Indeed, when only considering GPs and neurologists, we observed 8.8 consultations per person-year in our sample. On the opposite, the Organisation for Economic Co-operation and Development reported an average of 6.7 consultations per person-year from 2007 to 2013 in the general population when considering all specialities.[44] Compared to the general population in France, with 4.0 GP consultations per patient-year over the study period, patients with MS had 7.7 consultations per person-year.[45] In accordance with the Canadian studies,[46,47] this underscores the importance of GPs in MS care. Overall, about half of the patients had "low" care consumption (clusters 1 to 3), about one quarter had "medium" consumption (cluster 4), and another quarter (cluster 5) had "high" consumption. Based on the detailed results obtained from the clustering, it appears that the cluster with the highest number of patients (i.e. cluster 5, with 29.0 % of the patients) had the highest level of care consumption. This was probably related to the MS itself (i.e. a high number of consultations with a neurologist and a high frequency of DMT use) as well as comorbid conditions (e.g. LDD other than MS). Clusters 2, 3, and 4 appear to reflect care associated with MS, and can be considered to be low to medium levels of care. Cluster 1 was the smallest one with only 4.3% of the patients, and it comprised older patients with a higher proportion of deaths.

In our opinion, SSA is an innovative method that warrants consideration in the field of care pathways in light of several favourable features. First, this method allows for a holistic approach to pathways, and it therefore considers them as a meaningful conceptual unit.[48] Indeed such an approach permits the coordinated succession of HCPs to be taken into account, which is a characteristic of care pathways. It is opposed to conventional approaches in epidemiology that consider each state independently from one another. The comparison of SSA with other methods, which can be appropriate for the study of a care pathway, such as latent class analysis (LCA), revealed that both methodologies yield similar clustering results.[41,49,50] These studies have highlighted that SSA is easier to use and to compute than LCA and that, unlike LCA, it does not require specification of a model. Another major limitation of LCA is the hypothesis of local independence,[41] which cannot be assumed in care pathways since the order in the succession of care consumption is primordial. To overcome this limitation, Mikolai and Lyons-Amos proposed the use of latent class growth models (LCGM) that takes into account the ordering of events, and they reached a conclusion regarding the specifications for the use of each model.[50] However, further work is needed to better evaluate the merits of SSA compared to the latent class model in the field of care pathways.

Second, we demonstrated that SSA is a flexible methodology. It can be adapted to different epidemiological contexts in light of the possibility of allowing specification of the measure between sequences amongst others. To transfer this method from social sciences to areas of health care, we used choices based on empirical experiments, namely the effect of costs and death coding. Thus, the choice of splitting and then analysing sequences according to their length avoids the creation of a cluster for patients with delayed entry into the cohort. Similarly, with the choice of terminating the sequence after the death of a patient, we can account for the death of patients without having to create a specific group, which may have occurred with a devoted state.

The data were based on a representative random sample of the French population, which reduces selection bias and which ensures external validity of the results. Moreover, the characteristics (e.g. gender, age, and LDD status) of our study population are close to those of MS patients in France in general.[7,11] The identification of MS patients was based on several criteria, which allows for the detection of low care consumers who may only be identified according to one criterion. In addition, by using administrative databases, we accessed all reimbursed care consumption by patients, independently of self-reporting. This minimised memory bias and also represented a measure of care consumption as close as possible to the true consumption of the study population.

However, some limitations are due to the database itself. Indeed, EGB does not allow for clear identification of outpatient consultations with hospital-based neurologists, even though we know that MS expert centres are mainly located at French university hospitals. Furthermore, 31% of the hospital consultations in France are with a specialist,[51] and only 21.3% of French neurologists were in private practice in 2013.[52] Therefore, such hospital-based consultations with neurologists were imputed, although the true level of consumption is certainly greater than the assigned level. Indeed, we probably missed the close follow-up at DMT initiation, the monitoring of potential adverse events of DMT, and the follow-up consultations for untreated patients.

Unfortunately, only a few individual characteristics were available to describe the patterns, as the administrative databases do not contain clinical data. A discrepancy analysis (a generalization of the principles

of analysis of variance to study the relationship between state sequences and covariates) was performed.[37] The results were not shown because the pseudo-R^2 was too small (<0.05) regarding the available variables. In addition, the quality of the two typologies we obtained can be considered to be quite weak, which does not allow for assessment of a strong structure in the data. However, this limitation is probably due to the fact that the patients were not at the same stage of the disease during the 2007–2013 period. Moreover, care consumption may depend on several factors such as MS relapses or DMT use, for example. To build the states of sequences, we chose to sum the different consultations with health professionals based on equal weights. Even if this operation allowed for an easy to understand variable, specific work has to be undertaken to define weights for each type of consultation. Our work was based on calendar years, and we had no information in regard to the onset of the MS for these patients. This does not allow for left-aligned sequences, and clusters are hence driven by a particular state. These specific characteristics of the EGB database (particularly the lack of clinical data) led to some limitations in our study, although this does not detract from the relevance of SSA in achieving our objective.

This work is only a preliminary draft. It will be extended to the entire French population of patients with MS through use of the French national health insurance information system, i.e. the entire source population from which the EGB is extracted (these analyses are ongoing). It will make for a more accurate examination of the care pathways in a larger population of about 110,000 patients. It will mainly analyse outpatient consultations as well as biological tests, medical imaging examinations, and paramedical and other specialist care. A new method developed in sequence analysis, called multichannel sequence analysis,[34,53] will be used. This method is an extension of the conventional SSA presented in this article, and it allows several thematic sequences for one patient to be considered and for them to be studied simultaneously. Moreover, we are also currently developing algorithms to estimate relevant clinical parameters that affect care consumption, such as motor disability or adherence to treatments, by using parameters described by Hess et al.[54] Indeed, motor disability is a well-known factor involved in the increase in health care use and costs[55,56] while, by contrast, reliance on DMTs reduces care consumption, such as hospitalizations.[57,58] Comorbidities will also be explored more closely, through the recommendations made by Marrie et al.[59,60] These parameters may offer opportunities to better characterize the different clusters of the typologies obtained. To overcome the problem of quality of the partition, we will try to identify incident MS cases with an algorithm developed for Canadian administrative databases,[61] and thereby study the impact of the onset of MS on care consumption.

In conclusion, SSA is a promising way to study care pathways. For the next stage of this investigation, we intend to compare state sequence analysis with latent class models in the field of epidemiology so as to better determine its potential, strengths, and weaknesses in the area of health care. Further work based on SSA use is needed in the context of care pathways.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the French National Agency for Medicines and Health Product Safety (ANSM) through the PEPS platform (Pharmacoepidemiology of Health Products).

Supplementary material

Supplementary material is available for this article online.

ORCID iD

Jonathan Roux http://orcid.org/0000-0002-0158-2837

References

1. World Health Organization. *Innovative care for chronic conditions: building blocks for actions: global report*. WHO, 2002.

2. Vanhaecht K. *The impact of clinical pathways on the organisation of care processes*. Katholieke Universiteit Leuven, 2007.
3. Zander KS, Bower KA and Etheredge MLS. *Nursing case management: blueprints for transformation*. Boston, Mass.: New England Medical Center Hospitals, 1987.
4. Schrijvers G, Hoorn A van and Huiskes N. The care pathway: concepts and theories: an introduction. *Int J Integrated Care* 2012; **12**: e192.
5. Féry-Lemonnier E. Les parcours, une nécessité. *Actualité et dossier en santé publique* 2014; **88**: 12–15.
6. Haute Autorité de Santé (HAS). *Parcours de soins. Questions/Réponses*. HAS, May 2012.
7. Foulon S, Maura G, Dalichampt M, et al. Prevalence and mortality of patients with multiple sclerosis in France in 2012: a study based on French health insurance data. *J Neurol* 2017; **264**: 1185–1192.
8. MS International Federation. Atlas of MS, http://www.msif.org/about-us/advocacy/atlas/ (2015).
9. Leray E, Vukusic S, Debouverie M, et al. Excess Mortality in patients with multiple sclerosis starts at 20 years from clinical onset: data from a large-scale french observational study. *PLoS ONE* 2015; **10**: e0132033.
10. Direction de la stratégie, des études et des statistiques (DSES). *Personnes prises en charge pour sclérose en plaques (SEP) en 2015*. Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS), 2017.
11. Lefeuvre D, Rudant J, Foulon S, et al. Healthcare expenditure of multiple sclerosis patients in 2013: A nationwide study based on French health administrative databases. *Multiple Sclerosis J – Experiment Translat Clinical* 2017; **3**: 2055217317730421.
12. Fromont A, Lehanneur M-N, Rollot F, et al. Cost of multiple sclerosis in France. *Revue Neurologique* 2014; **170**: 432–439.
13. Cordesse V, Jametal T, Guy C, et al. Analysis of clinical pathway in changing and disabling neurological diseases. *Revue Neurologique* 2013; **4085**: 1–529.
14. Tuppin P, Samson S, Fagot-Campagna A, et al. Care pathways and healthcare use of stroke survivors six months after admission to an acute-care hospital in France in 2012. *Revue Neurologique* 2016; **172**: 295–306.
15. Egho E, Jay N, Raïssi C, et al. An approach for mining care trajectories for chronic diseases. In: Peek N, Marín Morales R and Peleg M (eds) *Artificial intelligence in medicine*. Heidelberg: Springer, 2013, pp.258–267.
16. Nuemi G, Afonso F, Roussot A, et al. Classification of hospital pathways in the management of cancer: application to lung cancer in the region of burgundy. *Cancer Epidemiol* 2013; **37**: 688–696.
17. Brzinsky-Fay C. Lost in transition? Labour market entry sequences of school leavers in Europe. *Eur Sociol Rev* 2007; **23**: 409–422.
18. McVicar D and Anyadike-Danes M. Predicting successful and unsuccessful transitions from school to work by using sequence methods. *J Royal Stat Soc: Ser A (Stat Soc)* 2002; **165**: 317–334.
19. Anyadike-Danes M and McVicar D. You'll never walk alone: childhood influences and male career path clusters. *Labour Econom* 2005; **12**: 511–530.
20. Biemann T and Datta DK. Analyzing sequence data: optimal matching in management research. *Organizational Res Meth* 2014; **17**: 51–76.
21. Le Meur N, Gao F and Bayat S. Mining care trajectories using health administrative information systems: the use of state sequence analysis to assess disparities in prenatal care consumption. *BMC Health Serv Res* 2015; **15**: 200.
22. World Health Organization. ICD-10 Version: 2016, http://apps.who.int/classifications/icd10/browse/2016/en (accessed 31 May 2016).
23. Studer M. *Étude des inégalités de genre en début de carrière académique à l'aide de méthodes innovatrices d'analyse de données séquentielles*. Geneva: University of Geneva, 2012.
24. Abbott A and Forrest J. Optimal matching methods for historical sequences. *J Interdiscipl History* 1986; **16**: 471.
25. Dlouhy K and Biemann T. Optimal matching analysis in career research: a review and some best-practice recommendations. *J Vocation Behav* 2015; **90**: 163–173.
26. Elzinga CH. Combinatorial representations of token sequences. *J Classification* 2005; **22**: 87–118.
27. Elzinga CH. Sequence similarity. *Sociol Meth Res* 2003; **32**: 3–29.
28. Abbott A and Tsay A. Sequence analysis and optimal matching methods in sociology: review and prospect. *Sociol Meth Res* 2000; **29**: 3–33.
29. Studer M and Ritschard G. What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *J Royal Stat Soc: Ser A (Stat Soc)* 2015; **179**: 481–511.
30. Robette N and Bry X. Harpoon or bait? A comparison of various metrics in fishing for sequence patterns. *BMS: Bull Sociol Methodol/Bulletin de Méthodologie Sociologique* 2012; **116**: 5–24.
31. Lesnard L. Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns. *Sociol Meth Res* 2010; **38**: 389–419.
32. Needleman SB and Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 1970; **48**: 443–453.
33. Gabadinho A, Ritschard G, Müller NS, et al. Analyzing and visualizing state sequences in R with TraMineR. *J Stat Software* 2011; **40**: 1–37.
34. Pollock G. Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis. *J Royal Stat Soc: Ser A (Stat Soc)* 2007; **170**: 167–183.

35. Elzinga CH. Distance, similarity and sequence comparison. In: Blanchard P, Bühlmann F, Gauthier J-A (eds) *Advances in sequence analysis: theory, method, applications*. Vol. 2. Berlin: Springer, 2014, pp.51–73.
36. Batagelj V. Generalized ward and related clustering problems. In: Bock HH (ed.) *Classification and related methods of data analysis*. Amsterdam, the Netherlands: North-Holland, 1988, pp.67–74.
37. Studer M, Ritschard G, Gabadinho A, et al. Discrepancy analysis of state sequences. *Sociol Meth Res* 2011; **40**: 471–510.
38. Studer M. *WeightedCluster library manual: a practical guide to creating typologies of trajectories in the social sciences with R*. LIVES Working Papers 24, 2013.
39. Kaufman L and Rousseeuw PJ. *Finding groups in data: an introduction to cluster analysis*. New York, NY: Wiley, 1990.
40. Hubert L and Levin J. A general statistical framework for assesing categorical clustering in free recall. *Psychol Bull* 1976; **83**: 1072–1080.
41. Han Y, Liefbroer A and Elzinga C. Comparing methods of classifying life courses: sequence analysis and latent class analysis. *Longitudinal and Life Course Studies* 2017; **8**: 319–341.
42. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Core Team, http://www.R-project.org/ (accessed 22 July 2015).
43. Maechler M, Rousseeuw P, Struyf A, et al. *cluster: cluster analysis basics and extensions*. R package version 2.0.3, 2015.
44. OECD Health Statistics. *Health care utilisation*. Organisation for Economic Co-operation and Development (OECD), http://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics/oecd-health-data-health-care-utilisation_data-00542-en (accessed 31 May 2016).
45. Institut de Recherche et Documentation en Economie de la Santé (IRDES). Bases de données Eco-Santé, www.ecosante.fr (accessed 29 July 2016).
46. Marrie RA, Yu N, Wei Y, et al. High rates of physician services utilization at least five years before multiple sclerosis diagnosis. *Multiple Sclerosis J* 2013; **19**: 1113–1119.
47. Pohar SL, Jones CA, Warren S, et al. Health status and health care utilization of multiple sclerosis in Canada. *Can J Neurol Sci* 2007; **34**: 167–74.
48. Billari FC. Sequence analysis in demographic research. *Can Studies Populat* 2001; **28**: 439–458.
49. Barban N and Billari FC. Classifying life course trajectories: a comparison of latent class and sequence analysis. *J Royal Stat Soc Ser C (Appl Stat)* 2012; **61**: 765–784.
50. Mikolai J and Lyons-Amos M. Longitudinal methods for life course research: A comparison of sequence analysis, latent class growth models, and multi-state event history models for studying partnership transitions. *Longitudinal Life Course Studies* 2017; **8**: 191–208.
51. Tellier S, De Peretti C and Boisguérin B. *Qui sont les patients des consultations externes hospitalières?* Direction de la recherche, des études, de l'évaluation et des statistiques (Drees), April 2002.
52. Le Breton-Lerouvillois G. *Atlas de la démographie médicale en France - Situation au 1er Janvier 2013*. Conseil National de l'Ordre des Médecins (CNOM), 2013.
53. Gauthier J-A, Widmer ED, Bucher P, et al. Multichannel sequence analysis applied to social science data. *Sociol Methodol* 2010; **40**: 1–38.
54. Hess LM, Raebel MA, Conner DA, et al. Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. *Ann Pharmacother* 2006; **40**: 1280–88.
55. Jones E, Pike J, Marshall T, et al. Quantifying the relationship between increased disability and health care resource utilization, quality of life, work productivity, health care costs in patients with multiple sclerosis in the US. *BMC Health Service Res* 2016; **16**: 1–9.
56. Kobelt G, Berg J, Lindgren P, et al. Costs and quality of life of patients with multiple sclerosis in Europe. *J Neurol Neurosurg Psychiatr* 2006; **77**: 918–926.
57. Halpern R, Agarwal S, Dembek C, et al. Comparison of adherence and persistence among multiple sclerosis patients treated with disease-modifying therapies: a retrospective administrative claims analysis. *Patient Prefer Adher* 2011; **5**: 73–84.
58. Steinberg SC, Faris RJ, Chang CF, et al. Impact of adherence to interferons in the treatment of multiple sclerosis. *Clin Drug Investigat* 2010; **30**: 89–100.
59. Marrie RA, Fisk JD, Stadnyk KJ, et al. Performance of administrative case definitions for comorbidity in multiple sclerosis in Manitoba and Nova Scotia. *Chronic Dis Injuries Canada* 2014; **34**: 145–153.
60. Marrie RA, Miller A, Sormani MP, et al. Recommendations for observational studies of comorbidity in multiple sclerosis. *Neurology* 2016; **86**: 1446–1453.
61. Marrie RA, Yu N, Blanchard J, et al. The rising prevalence and changing age distribution of multiple sclerosis in Manitoba. *Neurology* 2010; **74**: 465–471.

Annexe G-Résultats de l'application de la MCSA sur les données canadiennes

Table 39-Caractéristiques de la cohorte 1 (N=1 048) et de la cohorte 2 (N=3 180)

	Cohort 1 N=1048	**Cohort 2** N=3180
No. of women (%)	796 (76.0%)	2334 (73.4%)
Birth year [a]	1962 (1928-1987)	1951 (1913-1986)
Age at inclusion [a]	36.6 (11.3-71.0)	44.2 (9.1-82.8)
Year of MS onset [a]	1999 (1996-2004)	1985 (1942-1995)
Age at MS onset [a]	36.6 (11.3-70.9)	30.5 (3.4-75.2)
Age at MS onset categorized (%)		
≤20	39 (3.7%)	344 (10.8%)
]20;30]	254 (24.2%)	1187 (37.3%)
]30;40]	367 (35.0%)	1021 (32.1%)
]40;50]	287 (27.4%)	484 (15.2%)
>50	101 (9.6%)	144 (4.5%)
Duration since MS onset at inclusion (years) [a]	-	11.0 (0.1-54.0)
Follow-up (years) [a]	9.6 (4.0-13.0)	13.0 (4.0-13.0)
No. of deaths (%)	25 (2.4%)	400 (12.6%)
No. of deaths per person-year (/1000)	2.6	10.3
No. of patients lost to follow-up (%)	73 (7.0%)	556 (17.5%)
Socioeconomic quantile at inclusion		
1 (low)	185 (17.7%)	526 (16.5%)
2	190 (18.1%)	568 (17.9%)
3	217 (20.7%)	645 (20.3%)
4	202 (19.3%)	701 (22.0%)
5 (high)	232 (22.1%)	674 (21.2%)
Missing	22 (2.1%)	66 (2.1%)
EDSS score at inclusion		
Mild	1048 (100.0%)	998 (31.4%)
Moderate	0	656 (20.6%)
Severe	0	676 (21.3%)
Missing	0	850 (26.7%)

	Cohort 1 N=1048	Cohort 2 N=3180
EDSS score at the end of follow-up		
Mild	611 (58.3%)	977 (30.7%)
Moderate	261 (24.9%)	821 (25.8%)
Severe	176 (16.8%)	1382 (43.5%)
Global ARMSS score at inclusion [a]	0.6 (0.0-7.0)	5.4 (0.0-10.0)
Global ARMSS score at the end of follow-up [a]	3.9 (0.1-9.9)	4.9 (0.1-10.0)
Number of relapses [a*]	1.0 (0.0-13.0)	0.0 (0.0-13.0)
Annualized relapses rate [a*]	0.2 (0.0-1.3)	0.0 (0.0-2.2)
Initial course of the disease		
Relapsing-remitting MS (%)	989 (94.4%)	2920 (91.8%)
Primary progressive MS (%)	59 (5.6%)	260 (8.2%)
Secondary progressive course at the end of follow-up (%)	105/989 (10.0%)	1380/2920 (43.4%)
Care consumption since inclusion		
Number of visits to GPs per person-year [b]	10.1 (7.1)	10.2 (7.1)
Number of visits to neurologists per person-year [b]	1.8 (1.2)	1.0 (0.9)
Number of visits to internal medicine specialists per person-year [b]	0.5 (1.0)	0.7 (2.6)
Number of visits to a physician specialized in most frequent MS symptoms [b, c]	0.9 (1.1)	0.9 (1.2)
Number of visits to psychiatrists per person-year [b]	0.6 (2.9)	0.4 (1.7)
At least one hospital admission (%)	717 (68.4%)	2559 (80.5%)
Number of hospital admissions per person-year [b, d]	0.4 (0.5)	0.6 (0.9)
Length of stay [b, d]	4.7 (17.4)	9.6 (46.4)
Number of visits to one of the considered physicians [b, e]	13.9 (9.6)	13.2 (9.1)
Number of visits to any physician [b]	21.6 (13.9)	22.0 (14.3)

[a] Median (Minimum-Maximum). [b] Mean (Standard deviation). [c] PMR physician, ophthalmologist or urologist. [d] Only for patients having at least one hospital admission. [e] GP, neurologist, internal medicine specialist, physician specialized in most frequent MS symptoms, psychiatrist.* Excluding onset attack (relapse date of first relapse=MS onset date) and only for people having a RRMS initial disease course.

† Data presented are restricted to prevent direct or residual disclosure of identifiable data.

Table 40-Descriptif des comorbidités présentes des populations de la cohorte 1 (N=1 048) et de la cohorte 2 (N=3 180)

	Cohort 1 N=1048		**Cohort 2** N=3180	
	Using algorithm	**At least one H or P**	**Using algorithm**	**At least one H or P**
Hypertension				
Prior to inclusion (%)	32 (3.1%)	45 (4.3%)	101 (3.2%)	147 (4.6%)
Prior or after the inclusion (%)	147 (14.0%)	251 (24.0%)	763 (24.0%)	1117 (35.1%)
Hyperlipidemia				
Prior to inclusion (%)	5 (0.5%)	9 (0.9%)	12 (0.4%)	28 (0.9%)
Prior or after the inclusion (%)	72 (6.9%)	145 (13.8%)	300 (9.4%)	647 (20.3%)
Diabetes				
Prior to inclusion (%)	5 (0.5%)	11 (1.0%)	42 (1.3%)	48 (1.5%)
Prior or after the inclusion (%)	52 (5.0%)	107 (10.2%)	287 (9.0%)	486 (15.3%)
Chronic lung disease				
Prior to inclusion (%)	32 (3.1%)	42 (4.0%)	44 (1.4%)	64 (2.0%)
Prior or after the inclusion (%)	103 (9.8%)	167 (15.9%)	355 (11.2%)	602 (18.9%)
Mood or Anxiety disorder				
Prior to inclusion (%)	155 (14.8%)	234 (22.3%)	559 (17.6%)	862 (27.1%)
Prior or after the inclusion (%)	468 (44.7%)	819 (78.1%)	1379 (43.4%)	2503 (78.7%)
No. of comorbidities prior to inclusion (%)				
0	>834 (>79.6%)†	>741 (>70.7%)†	2467 (77.6%)	>2125 (66.8%)†
1	192 (18.3%)	269 (25.7%)	668 (21.0%)	960 (30.2%)
2	17 (1.6%)	33 (3.1%)	45 (1.4%)	90 (2.8%)
>2	<5 (<0.5%)†	<5 (<0.5%)†	0 (0.0%)	<5 (<0.2%)†
No. of comorbidities prior or after the inclusion (%)				
0	436 (41.6%)	145 (13.8%)	1105 (34.7%)	308 (9.7%)
1	436 (41.6%)	501 (47.8%)	1319 (41.5%)	1244 (39.1%)
2	133 (12.7%)	270 (25.8%)	549 (17.3%)	972 (30.6%)
>2	43 (4.1%)	132 (12.6%)	207 (6.5%)	656 (0.6%)
At least one comorbidity prior to inclusion (%)	210 (20.0%)	304 (29.0%)	713 (22.4%)	1053 (33.1%)
At least one comorbidity prior or after to inclusion (%)	612 (58.4%)	903 (86.2%)	2075 (65.3%)	2872 (90.3%)

H: Hospital admission; P: Physician claim. † Data presented are restricted to prevent direct or residual disclosure of identifiable data.

Table 41-Descriptif des comorbidités présentes des populations de la cohorte 1 (N=1 048) et de la cohorte 2 (N=3 180)

	Cohort 1 N=1048	**Cohort 2** N=3180
At least one prescription of Avonex® (%)	160 (15.3%)	208 (6.5%)
At least one prescription of Betaseron® (%)	227 (21.7%)	506 (15.9%)
At least one prescription of Rebif® (%)	355 (33.9%)	422 (13.3%)
At least one prescription of Copaxone® (%)	214 (20.4%)	235 (7.4%)
At least one prescription of Tysabri® (%)	16 (1.5%)	19 (0.6%)
At least one prescription of an interferon-β (%)	614 (58.6%)	933 (29.3%)
At least one prescription of a DMT excluding Tysabri® (%)	698 (66.6%)	1009 (31.7%)
At least one prescription of a DMT including Tysabri® (%)	699 (66.7%)	1010 (31.8%)
Number of DMTs prescribed over follow-up excluding Tysabri® (%)		
0	350 (33.4%)	2171 (68.3%)
1	476 (45.4%)	715 (22.5%)
2	189 (18.0%)	230 (7.2%)
>2	33 (3.1%)	64 (2.0%)
Number of DMTs prescribed over follow-up including Tysabri® (%)		
0	349 (33.3%)	2170 (68.2%)
1	468 (44.7%)	708 (22.3%)
2	193 (18.4%)	232 (7.3%)
>2	38 (3.6%)	70 (2.2%)

DMT: Disease-Modifying Therapy

Table 42-Caractéristiques des clusters de la partition en douze groupes obtenue suite à la multichannel sequence analysis appliquée sur la cohorte 1 (n=1 048)

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**
	n=221	n=63	n=59	n=200	n=105	n=58	n=88
	(21.1%)	(6.0%)	(5.6%)	(19.1%)	(10.0%)	(5.5%)	(8.4%)
No. of women (%)	173 (78.3%)	49 (77.8%)	45 (76.3%)	151 (75.5%)	72 (68.6%)	49 (84.5%)	61 (69.3%)
Birth year [b]	1959	1962	1969	1961	1966	1959	1966.5
	(1931-1981)	(1932-1983)	(1942-1984)	(1939-1985)	(1936-1980)	(1928-1978)	(1942-1983)
Age at inclusion [b]	37.3	36.6	33.7	37.0	33.5	38.5	35.2
	(15.3-64.5)	(16.0-66.1)	(17.2-58.6)	(11.3-58.3)	(19.5-63.8)	(17.9-67.2)	(18.6-60.1)
Year of MS onset [b]	1997	1999	2002	1998	2000	1997	2002
	(1996-1998)	(1996-2000)	(1996-2003)	(1996-1999)	(1996-2001)	(1996-1999)	(1996-2003)
Age at MS onset categorized (%)							
≤20	7 (3.2%)	5 (7.9%)	<5 (<8.5%)†	8 (4.0%)	<5 (<4.8%)†	<5 (<8.6%)†	<5 (<5.7%)†
]20;30]	42 (19.0%)	17 (27.0%)	>16 (>27.1%)†	41 (20.5%)	34 (32.4%)	11 (19.0%)	21 (23.9%)
]30;40]	88 (39.8%)	17 (27.0%)	17 (28.8%)	79 (39.5%)	>33 (>31.4%)†	19 (32.8%)	>34 (>38.6%)†
]40;50]	62 (28.1%)	18 (28.6%)	15 (25.4%)	54 (27.0%)	20 (19.0%)	>17 (>29.3%)†	23 (26.1%)
>50	22 (10.0%)	6 (9.5%)	6 (10.2%)	18 (9.0%)	13 (12.4%)	6 (10.3%)	5 (5.7%)
Follow-up (years) [b]	12.0 (10.7-13.0)	9.8 (8.1-11.0)	6.2 (5.0-7.6)	10.6 (9.0-12.0)	8.8 (7.6-10.0)	11.7 (9.4-13.0)	6.4 (5.2-7.6)
No. of deaths (%)	<5 (<2.3%)†	<5 (<7.9%)†	<5 (<8.5%)†	0 (0.0%)	<5 (<4.8%)†	<5 (<8.6%)†	<5 (<5.7%)†
No. of deaths per person-year (/1,000)	<1.9†	<8.1†	<13.7†	0.0	<5.4†	<7.4†	<8.9†

	Cluster 1 n=221 (21.1%)	**Cluster 2** n=63 (6.0%)	**Cluster 3** n=59 (5.6%)	**Cluster 4** n=200 (19.1%)	**Cluster 5** n=105 (10.0%)	**Cluster 6** n=58 (5.5%)	**Cluster 7** n=88 (8.4%)
Socioeconomic quantile at inclusion							
1 (low)	31 (14.0%)	10 (15.9%)	13 (22.0%)	40 (20.0%)	21 (20.0%)	12 (20.7%)	11 (12.5%)
2	38 (17.2%)	>16 (>25.4%)†	10 (16.9%)	38 (19.0%)	16 (15.2%)	12 (20.7%)	13 (14.8%)
3	40 (18.1%)	17 (27.0%)	>13 (>22.0%)†	38 (19.0%)	>26 (>24.8%)†	10 (17.2%)	14 (15.9%)
4	49 (22.2%)	8 (12.7%)	8 (13.6%)	35 (17.5%)	19 (18.1%)	10 (17.2%)	20 (22.7%)
5 (high)	57 (25.8%)	7 (11.1%)	10 (16.9%)	>44 (>22.0%)†	18 (17.1%)	14 (24.1%)	>25 (>28.4%)†
Missing	6 (2.7%)	<5 (<7.9%)†	<5 (<8.5%)†	<5 (<2.5%)†	<5 (<4.8%)†	0 (0.0%)	<5 (<5.7%)†
EDSS score at inclusion							
Mild	221 (100.0%)	63 (100.0%)	59 (100.0%)	200 (100.0%)	105 (100.0%)	58 (100.0%)	88 (100.0%)
Moderate	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Severe	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
EDSS score at the end of follow-up							
Mild	114 (51.6%)	30 (47.6%)	35 (59.3%)	130 (65.0%)	71 (67.6%)	23 (39.7%)	63 (71.6%)
Moderate	58 (26.2%)	19 (30.2%)	13 (22.0%)	45 (22.5%)	24 (22.9%)	18 (31.0%)	17 (19.3%)
Severe	49 (22.2%)	14 (22.2%)	11 (18.6%)	25 (12.5%)	10 (9.5%)	17 (29.3%)	8 (9.1%)
Global ARMSS score at inclusion [b]	0.6 (0.1-2.5)	0.6 (0.1-1.4)	0.7 (0.1-1.4)	0.6 (0.1-3.1)	0.8 (0.1-1.6)	0.5 (0.1-1.6)	0.7 (0.1-6.1)
Global ARMSS score at the end of follow-up [b]	3.8 (0.2-9.8)	4.1 (0.3-9.8)	4.6 (0.9-9.3)	3.2 (0.1-9.5)	3.7 (0.2-9.7)	3.9 (0.5-9.9)	3.2 (0.2-8.9)
Number of relapses reported in BC MS database [a]	3.0 (0.0-14.0)	2.0 (0.0-9.0)	2.0 (0.0-8.0)	3.0 (0.0-10.0)	2.0 (0.0-11.0)	2.0 (0.0-10.0)	2.0 (0.0-7.0)

	Cluster 1 n=221 (21.1%)	**Cluster 2** n=63 (6.0%)	**Cluster 3** n=59 (5.6%)	**Cluster 4** n=200 (19.1%)	**Cluster 5** n=105 (10.0%)	**Cluster 6** n=58 (5.5%)	**Cluster 7** n=88 (8.4%)
Initial course of the disease							
Relapsing-remitting MS (%)	207 (93.7%)	57 (90.5%)	>54 (>91.5%)†	191 (95.5%)	99 (94.3%)	>53 (>91.4%)†	>83 (>94.3%)†
Primary progressive MS (%)	14 (6.3%)	6 (9.5%)	<5 (<8.5%)†	9 (4.5%)	6 (5.7%)	<5 (<8.6%)†	<5 (<5.7%)†
Secondary progressive course at the end of follow-up (%)	42 (19.0%)	6 (9.5%)	<5 (<8.5%)†	14 (7.0%)	9 (8.6%)	8 (13.8%)	0 (0.0%)
Comorbidities							
Hypertension							
Using algorithms							
Prior to inclusion (%)	8 (3.6%)	<5 (<7.9%)†	<5 (<8.5%)†	7 (3.5%)	<5 (<4.8%)†	<5 (<8.6%)†	<5 (<5.7%)†
Prior or after the inclusion (%)	38 (17.2%)	9 (14.3%)	8 (13.6%)	26 (13.0%)	8 (7.6%)	17 (29.3%)	7 (8.0%)
At least one P or H [c]							
Prior to inclusion (%)	11 (5.0%)	<5 (<7.9%)†	<5 (<8.5%)†	9 (4.5%)	<5 (<4.8%)†	<5 (<8.6%)†	<5 (<5.7%)†
Prior or after the inclusion (%)	64 (29.0%)	17 (27.0%)	12 (20.3%)	46 (23.0%)	15 (14.3%)	27 (46.6%)	14 (15.9%)
Hyperlipidemia							
Using algorithms							
Prior to inclusion (%)	<5 (<2.3%)†	0 (0.0%)	0 (0.0%)	<5 (<2.5%)†	0 (0.0%)	0 (0.0%)	0 (0.0%)
Prior or after the inclusion (%)	14 (6.3%)	7 (11.1%)	<5 (<8.5%)†	11 (5.5%)	5 (4.8%)	9 (15.5%)	<5 (<5.7%)†

	Cluster 1 n=221 (21.1%)	**Cluster 2** n=63 (6.0%)	**Cluster 3** n=59 (5.6%)	**Cluster 4** n=200 (19.1%)	**Cluster 5** n=105 (10.0%)	**Cluster 6** n=58 (5.5%)	**Cluster 7** n=88 (8.4%)
At least one P or H [c]							
Prior to inclusion (%)	<5 (<2.3%)†	0 (0.0%)	<5 (<8.5%)†	<5 (<2.5%)†	0 (0.0%)	0 (0.0%)	0 (0.0%)
Prior or after the inclusion (%)	32 (14.5%)	14 (22.2%)	7 (11.9%)	22 (11.0%)	10 (9.5%)	18 (31.0%)	8 (9.1%)
Diabetes							
Using algorithms							
Prior to inclusion (%)	<5 (<2.3%)†	0 (0.0%)	0 (0.0%)	0 (0.0%)	<5 (<4.8%)†	<5 (<8.6%)†	0 (0.0%)
Prior or after the inclusion (%)	12 (5.4%)	6 (9.5%)	0 (0.0%)	8 (4.0%)	<5 (<4.8%)†	9 (15.5%)	<5 (<5.7%)†
At least one P or H [c]							
Prior to inclusion (%)	<5 (<2.3%)†	0 (0.0%)	<5 (<8.5%)†	<5 (<2.5%)†	<5 (<4.8%)†	<5 (8.6%)†	0 (0.0%)
Prior or after the inclusion (%)	26 (11.8%)	8 (12.7%)	<5 (<8.5%)†	15 (7.5%)	<5 (<4.8%)†	16 (27.6%)	6 (6.8%)
Chronic lung disease							
Using algorithms							
Prior to inclusion (%)	5 (2.3%)	<5 (<7.9%)†	<5 (<8.5%)†	5 (2.5%)	<5 (4.8%)†	5 (8.6%)	<5 (<5.7%)†
Prior or after the inclusion (%)	19 (8.6%)	7 (11.1%)	12 (20.3%)	13 (6.5%)	7 (6.7%)	18 (31.0%)	<5 (<5.7%)†
At least one P or H [c]							
Prior to inclusion (%)	6 (2.7%)	<5 (<7.9%)†	<5 (<8.5%)†	5 (2.5%)	6 (5.7%)	5 (8.6%)	<5 (<5.7%)†
Prior or after the inclusion (%)	31 (14.0%)	14 (22.2%)	13 (22.0%)	25 (12.5%)	16 (15.2%)	22 (37.9%)	8 (9.1%)

	Cluster 1 n=221 (21.1%)	**Cluster 2** n=63 (6.0%)	**Cluster 3** n=59 (5.6%)	**Cluster 4** n=200 (19.1%)	**Cluster 5** n=105 (10.0%)	**Cluster 6** n=58 (5.5%)	**Cluster 7** n=88 (8.4%)
Mood or Anxiety disorder							
Using algorithms							
Prior to inclusion (%)	17 (7.7%)	10 (15.9%)	24 (40.7%)	8 (4.0%)	9 (8.6%)	22 (37.9%)	6 (14.8%)
Prior or after the inclusion (%)	95 (43.0%)	36 (57.1%)	39 (66.1%)	70 (35.0%)	26 (24.8%)	46 (79.3%)	19 (21.6%)
At least one P or H [c]							
Prior to inclusion (%)	36 (16.3%)	15 (23.8%)	29 (49.2%)	26 (13.0%)	20 (19.0%)	25 (43.1%)	14 (15.9%)
Prior or after the inclusion (%)	181 (81.9%)	58 (92.1%)	53 (89.8%)	152 (76.0%)	63 (60.0%)	>53 (>91.4%)†	55 (62.5%)
No. of comorbidities prior to inclusion (%)							
Using algorithms							
0	>185 (>83.7%)†	>48 (>76.2%)†	>31 (>52.5%)†	179 (89.5%)	90 (85.7%)	>29 (>50.0%)†	80 (90.9%)
1	31 (14.0%)	10 (15.9%)	22 (37.3%)	21 (10.5%)	15 (14.3%)	24 (41.4%)	8 (9.1%)
2	<5 (<2.3%)†	<5 (<7.9%)†	<5 (<8.5%)†	0 (0.0%)	0 (0.0%)	<5 (<8.6%)†	0 (0.0%)
>2	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
At least one P or H [c]							
0	>165 (>74.7%)†	>44 (>69.8%)†	>24 (>40.7%)†	>153 (>76.5%)†	>78 (>74.3%)†	>26 (>44.8%)†	70 (79.5%)
1	51 (23.1%)	14 (22.2%)	23 (39.0%)	42 (21.0%)	22 (21.0%)	27 (46.6%)	18 (20.5%)
2	<5 (<2.3%)†	<5 (<7.9%)†	7 (11.9%)	<5 (<2.5%)†	<5 (<4.8%)†	<5 (<8.6%)†	0 (0.0%)
>2	0 (0.0%)	0 (0.0%)	<5 (<8.5%)†	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

	Cluster 1 n=221 (21.1%)	**Cluster 2** n=63 (6.0%)	**Cluster 3** n=59 (5.6%)	**Cluster 4** n=200 (19.1%)	**Cluster 5** n=105 (10.0%)	**Cluster 6** n=58 (5.5%)	**Cluster 7** n=88 (8.4%)
No. of comorbidities prior or after the inclusion (%)							
Using algorithms							
0	86 (38.9%)	16 (25.4%)	13 (22.0%)	104 (52.0%)	>60 (>57.1%)†	8 (13.8%)	>52 (>59.1%)†
1	99 (44.8%)	>33 (>52.4%)†	>28 (>47.5%)†	71 (35.5%)	35 (33.3%)	19 (32.8%)	26 (29.5%)
2	30 (13.6%)	9 (14.3%)	13 (22.0%)	18 (9.0%)	5 (4.8%)	18 (31.0%)	<5 (<5.7%)†
>2	6 (2.7%)	<5 (<7.9%)†	<5 (<8.5%)†	7 (3.5%)	<5 (<4.8%)†	13 (22.4%)	<5 (<5.7%)†
At least one P or H [c]							
0	23 (10.4%)	<5 (<7.9%)†	<5 (<8.5%)†	31 (15.5%)	28 (26.7%)	0 (0.0%)	26 (29.5%)
1	96 (43.4%)	>30 (47.6%)†	>28 (>47.5%)†	103 (51.5%)	>52 (>49.5%)†	18 (31.0%)	>36 (>40.9%)†
2	75 (33.9%)	12 (19.0%)	20 (33.9%)	46 (23.0%)	20 (19.0%)	12 (20.7%)	21 (23.9%)
>2	27 (12.2%)	16 (25.4%)	6 (10.2%)	20 (10.0%)	<5 (<4.8%)†	24 (41.4%)	<5 (<5.7%)†
At least one comorbidity prior to inclusion using algorithms (%)	32 (14.5%)	12 (19.0%)	26 (44.1%)	21 (10.5%)	15 (14.3%)	27 (46.6%)	8 (9.1%)
At least one comorbidity prior or after to inclusion using algorithms (%)	135 (61.1%)	47 (74.6%)	46 (78.0%)	96 (48.0%)	41 (39.0%)	50 (86.2%)	31 (35.2%)
At least one comorbidity prior to inclusion using at least one P or H [c] (%)	54 (24.4%)	17 (27.0%)	31 (52.5%)	43 (21.5%)	26 (24.8%)	31 (53.4%)	18 (20.5%)

	Cluster 1 n=221 (21.1%)	**Cluster 2** n=63 (6.0%)	**Cluster 3** n=59 (5.6%)	**Cluster 4** n=200 (19.1%)	**Cluster 5** n=105 (10.0%)	**Cluster 6** n=58 (5.5%)	**Cluster 7** n=88 (8.4%)
At least one comorbidity prior or after to inclusion using at least one P or H [c] (%)	198 (89.6%)	>58 (>92.1%)†	>54 (>91.5%)†	169 (84.5%)	77 (73.3%)	58 (100.0%)	62 (70.5%)
Prescription of DMTs							
At least one prescription of Avonex® (%)	23 (10.4%)	13 (20.6%)	13 (22.0%)	30 (15.0%)	15 (14.3%)	12 (20.7%)	16 (18.2%)
At least one prescription of Betaseron® (%)	46 (20.8%)	15 (23.8%)	19 (32.2%)	38 (19.0%)	18 (17.1%)	14 (24.1%)	23 (26.1%)
At least one prescription of Rebif® (%)	60 (27.1%)	19 (30.2%)	21 (35.6%)	65 (32.5%)	35 (33.3%)	20 (34.5%)	32 (36.4%)
At least one prescription of Copaxone® (%)	42 (19.0%)	17 (27.0%)	13 (22.0%)	34 (17.0%)	18 (17.1%)	10 (17.2%)	21 (23.9%)
At least one prescription of an interferon-β (%)	110 (49.8%)	38 (60.3%)	41 (69.5%)	113 (56.5%)	58 (55.2%)	37 (63.8%)	62 (70.5%)
At least one prescription of a DMT (%)	132 (59.7%)	43 (68.3%)	44 (74.6%)	127 (63.5%)	71 (67.6%)	39 (67.2%)	67 (76.1%)
Number of DMTs prescribed over follow-up (%)							
0	89 (40.3%)	20 (31.7%)	15 (25.4%)	73 (36.5%)	34 (32.4%)	19 (32.8%)	21 (23.9%)
1	99 (44.8%)	>23 (>36.5%)†	30 (50.8%)	>90 (45.0%)†	56 (53.3%)	>23 (>39.7%)†	>41 (>46.6%)†
2	27 (12.2%)	15 (23.8%)	8 (13.6%)	32 (16.0%)	15 (14.3%)	11 (19.0%)	21 (23.9%)
>2	6 (2.7%)	<5 (<7.9%)†	6 (10.2%)	<5 (<2.5%)†	0 (0.0%)	<5 (<8.6%)†	<5 (<5.7%)†

	Cluster 1 n=221 (21.1%)	**Cluster 2** n=63 (6.0%)	**Cluster 3** n=59 (5.6%)	**Cluster 4** n=200 (19.1%)	**Cluster 5** n=105 (10.0%)	**Cluster 6** n=58 (5.5%)	**Cluster 7** n=88 (8.4%)
Care consumption since inclusion							
Number of visits to GPs per person-year [d]	8.3 (3.5)	12.1 (4.9)	16.7 (7.5)	7.2 (3.6)	7.1 (3.6)	21.7 (8.1)	6.4 (2.3)
Number of visits to neurologists per person-year [d]	1.3 (0.6)	2.0 (0.9)	2.2 (1.1)	1.4 (0.5)	1.5 (0.7)	2.0 (1.1)	1.9 (0.9)
Number of visits to internal medicine specialists per person-year [d]	0.3 (0.5)	0.7 (1.1)	0.7 (1.1)	0.2 (0.4)	0.2 (0.3)	1.4 (2.2)	0.3 (0.5)
Number of visits to symptoms specialists per person-year [d, e]	0.8 (0.9)	1.6 (1.3)	1.2 (1.0)	0.5 (0.5)	0.6 (0.6)	1.7 (1.9)	0.8 (0.9)
Number of visits to psychiatrists per person-year [d]	0.1 (0.4)	0.1 (0.3)	0.5 (1.5)	0.0 (0.2)	0.1 (0.5)	0.0 (0.1)	0.1 (0.3)
At least one hospital admission (%)	169 (76.5%)	57 (90.5%)	41 (69.5%)	129 (64.5%)	57 (54.3%)	53 (91.4%)	42 (47.7%)
Number of hospital admissions per person-year [d, e]	0.2 (0.2)	0.4 (0.2)	0.6 (0.4)	0.2 (0.1)	0.2 (0.2)	0.6 (0.5)	0.4 (0.3)
Length of stay [d, f]	3.4 (8.9)	3.4 (11.7)	4.5 (9.5)	2.2 (5.7)	3.2 (10.5)	3.2 (5.6)	5.4 (9.9)
Number of visits to one of the considered physicians [d, g]	10.9 (4.1)	16.5 (4.9)	21.3 (8.5)	9.4 (3.8)	9.6 (3.8)	27.5 (9.3)	9.3 (3.2)
Number of visits to any physician [d]	16.8 (7.4)	24.0 (8.7)	29.5 (10.8)	14.7 (7.1)	13.9 (6.3)	38.8 (16.3)	14.4 (5.3)

[a] P-value of the comparison of the groups according to Kruskal-Wallis, Chi-squared test or Fisher exact test, if needed. [b] Mean (Standard deviation). [c] P: Physician claim or H: Hospital admission. [d] Mean (standard deviation). [e] Symptoms specialist: PMR physician, ophthalmologist or urologist. [f] Only for patients having at least one hospital admission. [g] GP, neurologist, internal medicine specialist, physician specialized in most frequent MS symptoms, psychiatrist. [†] Data presented are restricted to prevent direct or residual disclosure of identifiable data.

Table 42-Caractéristiques des clusters de la partition en douze groupes obtenue suite à la multichannel sequence analysis appliquée sur la cohorte 1 (n=1 048) (continued)

	Cluster 8	**Cluster 9**	**Cluster 10**	**Cluster 11**	**Cluster 12**	**p-value** [a]
	n=44	n=83	n=32	n=81	n=14	
	(4.2%)	(7.9%)	(3.1%)	(7.7%)	(1.3%)	
No. of women (%)	32 (72.7%)	71 (85.5%)	27 (84.4%)	53 (65.4%)	>9 (>64.3%)†	**0.035**
Birth year [b]	1970.5	1964	1959	1965	1965	**<0.001**
	(1939-1987)	(1934-1986)	(1938-1985)	(1928-1985)	(1937-1983)	
Age at inclusion [b]	31.3 (17.3-64.8)	36.8 (14.7-63.4)	39.6 (14.8-63.2)	36.8 (15.6-71.0)	37.3 (21.1-67.3)	0.595
Year of MS onset [b]	2004	2001	1998	2001	2003	**<0.001**
	(1997-2004)	(1996-2003)	(1996-2001)	(1996-2002)	(1996-2004)	
Age at MS onset categorized (%)						
≤20	<5 (<11.4%)†	<5 (<6.0%)†	<5 (<15.6%)†	6 (7.4%)	0 (0.0%)	0.301
]20;30]	>15 (>34.1%)†	22 (26.5%)	6 (18.8%)	20 (24.7%)	<5 (<35.7%)†	
]30;40]	12 (27.3%)	23 (27.7%)	9 (28.1%)	23 (28.4%)	>5 (>35.7%)†	
]40;50]	7 (15.9%)	>25 (>30.1%)†	>7 (>21.9%)†	26 (32.1%)	<5 (<35.7%)†	
>50	5 (11.4%)	8 (9.6%)	<5 (<15.6%)†	6 (7.4%)	<5 (<35.7%)†	
Follow-up (years) [b]	4.9 (4.0-5.5)	7.8 (4.9-9.0)	10.7 (6.7-13.0)	7.4 (6.6-9.0)	5.2 (4.2-6.4)	**<0.001**
No. of deaths (%)	<5 (<11.4%)†	6 (7.2%)	<5 (<15.6%)†	<5 (<6.2%)†	<5 (35.7%)†	**<0.001**
No. of deaths per person-year (/1,000)	<23.2†	9.4	<14.6†	<8.3†	<68.7†	-

	Cluster 8 n=44 (4.2%)	**Cluster 9** n=83 (7.9%)	**Cluster 10** n=32 (3.1%)	**Cluster 11** n=81 (7.7%)	**Cluster 12** n=14 (1.3%)	**p-value** [a]
Socioeconomic quantile at inclusion						0.569
1 (low)	9 (20.5%)	14 (16.9%)	>5 (>15.6%)†	14 (17.3%)	<5 (<35.7%)†	
2	7 (15.9%)	15 (18.1%)	5 (15.6%)	13 (16.0%)	<5 (<35.7%)†	
3	7 (15.9%)	>17 (>20.5%)†	7 (21.9%)	>15 (>18.5%)†	<5 (<35.7%)†	
4	9 (20.5%)	16 (19.3%)	7 (21.9%)	17 (21.0%)	<5 (<35.7%)†	
5 (high)	12 (27.3%)	16 (19.3%)	<5 (<15.6%)†	17 (21.0%)	<5 (<35.7%)†	
Missing	0 (0.0%)	<5 (<6.0%)†	<5 (<15.6%)†	<5 (<6.2%)†	0 (0.0%)	
EDSS score at inclusion						-
Mild	44 (100.0%)	83 (100.0%)	32 (100.0%)	81 (100.0%)	14 (100.0%)	
Moderate	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Severe	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
EDSS score at the end of follow-up						**<0.001**
Mild	>31 (>70.5%)†	37 (44.6%)	16 (50.0%)	57 (70.4%)	<5 (<35.7%)†	
Moderate	8 (18.2%)	29 (34.9%)	9 (28.1%)	13 (16.0%)	>5 (>35.7%)†	
Severe	<5 (<11.4%)†	17 (20.5%)	7 (21.9%)	11 (13.6%)	<5 (<35.7%)†	
Global ARMSS score at inclusion [b]	0.9 (0.1-1.6)	0.6 (0.1-1.4)	0.5 (0.1-7.0)	0.6 (0.0-1.6)	0.6 (0.1-1.4)	0.599
Global ARMSS score at the end of follow-up [b]	4.7 (0.5-8.8)	5.0 (0.5-9.9)	4.5 (0.3-9.8)	3.9 (0.1-9.5)	6.2 (0.5-8.9)	**<0.001**
Number of relapses reported in BC MS database [b]	1.0 (0.0-6.0)	2.0 (0.0-8.0)	3.0 (0.0-8.0)	2.0 (0.0-6.0)	1.0 (0.0-6.0)	**<0.001**

	Cluster 8 n=44 (4.2%)	**Cluster 9** n=83 (7.9%)	**Cluster 10** n=32 (3.1%)	**Cluster 11** n=81 (7.7%)	**Cluster 12** n=14 (1.3%)	**p-value** [a]
Initial course of the disease						0.409
Relapsing-remitting MS (%)	>39 (>88.6%)†	74 (89.2%)	32 (100.0%)	76 (93.8%)	14 (100.0%)	
Primary progressive MS (%)	<5 (<11.4%)†	9 (10.8%)	0 (0.0%)	5 (6.2%)	0 (0.0%)	
Secondary progressive course at the end of follow-up (%)	<5 (<11.4%)†	13 (15.7%)†	<5 (<15.6%)†	<5 (<6.2%)†	<5 (<35.7%)†	**<0.001**
Comorbidities						
Hypertension						
Using algorithms						
Prior to inclusion (%)	0 (0.0%)	<5 (<6.0%)†	<5 (<15.6%)†	<5 (<6.2%)†	<5 (<35.7%)†	0.655
Prior or after the inclusion (%)	5 (11.4%)	13 (15.7%)	5 (15.6%)	8 (9.9%)	<5 (<35.7%)†	**0.033**
At least one P or H [c]						
Prior to inclusion (%)	0 (0.0%)	<5 (<6.0%)†	<5 (<15.6%)†	<5 (<6.2%)†	<5 (<35.7%)†	0.681
Prior or after the inclusion (%)	7 (15.9%)	24 (28.9%)	7 (21.9%)	15 (18.5%)	<5 (<35.7%)†	**0.001**
Hyperlipidemia						
Using algorithms						
Prior to inclusion (%)	0 (0.0%)	0 (0.0%)	<5 (<15.6%)†	<5 (<6.2%)†	0 (0.0%)	0.780
Prior or after the inclusion (%)	0 (0.0%)	7 (8.4%)	5 (15.6%)	5 (6.2%)	<5 (35.7%)†	0.079

	Cluster 8 n=44 (4.2%)	**Cluster 9** n=83 (7.9%)	**Cluster 10** n=32 (3.1%)	**Cluster 11** n=81 (7.7%)	**Cluster 12** n=14 (1.3%)	**p-value** [a]
At least one P or H [c]						
Prior to inclusion (%)	0 (0.0%)	<5 (<6.0%)†	<5 (<15.6%)†	<5 (<6.2%)†	0 (0.0%)	0.363
Prior or after the inclusion (%)	<5 (<11.4%)†	16 (19.3%)	7 (21.9%)	9 (11.1%)	<5 (<35.7%)†	**0.001**
Diabetes						
Using algorithms						
Prior to inclusion (%)	<5 (<11.4%)†	0 (0.0%)	0 (0.0%)	<5 (<6.2%)†	0 (0.0%)	0.218
Prior or after the inclusion (%)	<5 (<11.4%)†	5 (6.0%)	<5 (<15.6%)†	<5 (<6.2%)†	<5 (<35.7%)†	**0.018**
At least one P or H [c]						
Prior to inclusion (%)	<5 (<11.4%)†	0 (0.0%)	<5 (<15.6%)†	<5 (<6.2%)†	0 (0.0%)	0.217
Prior or after the inclusion (%)	5 (11.4%)	10 (12.0%)	7 (21.9%)	5 (6.2%)	<5 (<35.7%)†	**<0.001**
Chronic lung disease						
Using algorithms						
Prior to inclusion (%)	0 (0.0%)	<5 (<6.0%)†	<5 (<15.6%)†	0 (0.0%)	<5 (<35.7%)†	0.713
Prior or after the inclusion (%)	<5 (<11.4%)†	10 (12.0%)	6 (18.8%)	<5 (<6.2%)†	<5 (<35.7%)†	**<0.001**
At least one P or H [c]						
Prior to inclusion (%)	0 (0.0%)	<5 (<6.0%)†	<5 (<15.6%)†	<5 (<6.2%)†	<5 (<35.7%)†	0.329
Prior or after the inclusion (%)	<5 (<11.4%)†	20 (24.1%)	7 (21.9%)	7 (8.6%)	<5 (<35.7%)†	**<0.001**

	Cluster 8 n=44 (4.2%)	**Cluster 9** n=83 (7.9%)	**Cluster 10** n=32 (3.1%)	**Cluster 11** n=81 (7.7%)	**Cluster 12** n=14 (1.3%)	**p-value** [a]
Mood or Anxiety disorder						
Using algorithms						
Prior to inclusion (%)	9 (20.5%)	20 (24.1%)	17 (53.1%)	6 (7.4%)	7 (50.0%)	**<0.001**
Prior or after the inclusion (%)	20 (45.5%)	53 (63.9%)	>27 (>84.4%)†	20 (24.7%)	14 (100.0%)	**<0.001**
At least one P or H [c]						
Prior to inclusion (%)	11 (25.0%)	24 (28.9%)	17 (53.1%)	10 (12.3%)	7 (50.0%)	**<0.001**
Prior or after the inclusion (%)	29 (65.9%)	76 (91.6%)	32 (100.0%)	49 (60.5%)	14 (100.0%)	**<0.001**
No. of comorbidities prior to inclusion (%)						
Using algorithms						**<0.001**
0	34 (77.3%)	>60 (>72.3%)†	13 (40.6%)	70 (86.4%)	>5 (>35.7%)†	
1	10 (22.7%)	18 (21.7%)	>9 (>28.1%)†	11 (13.6%)	>5 (>35.7%)†	
2	0 (0.0%)	<5 (<6.0%)†	<5 (<15.6%)†	0 (0.0%)	<5 (<35.7%)†	
>2	0 (0.0%)	0 (0.0%)	<5 (<15.6%)†	0 (0.0%)	0 (0.0%)	
At least one P or H [c]						**<0.001**
0	32 (72.7%)	>55 (>66.3%)†	12 (37.5%)	>60 (>74.1%)†	>5 (35.7%)†	
1	12 (27.3%)	23 (27.7%)	>10 (>31.3%)†	16 (19.8%)	>5 (35.7%)†	
2	0 (0.0%)	<5 (<6.0%)†	<5 (<15.6%)†	<5 (<6.2%)†	<5 (<35.7%)†	
>2	0 (0.0%)	0 (0.0%)	<5 (<15.6%)†	0 (0.0%)	0 (0.0%)	

	Cluster 8 n=44 (4.2%)	**Cluster 9** n=83 (7.9%)	**Cluster 10** n=32 (3.1%)	**Cluster 11** n=81 (7.7%)	**Cluster 12** n=14 (1.3%)	**p-value** [a]
No. of comorbidities prior or after the inclusion (%)						
Using algorithms						**<0.001**
0	18 (40.9%)	20 (24.1%)	<5 (<15.6%)†	>44 (>54.3%)†	0 (0.0%)	
1	>21 (>47.7%)†	>42 (>50.6%)†	>13 (>40.6%)†	27 (33.3%)	>5 (>35.7%)†	
2	<5 (<11.4%)†	16 (19.3%)	9 (28.1%)	<5 (<6.2%)†	5 (35.7%)	
>2	0 (0.0%)	<5 (<6.0%)†	<5 (<15.6%)†	<5 (<6.2%)†	<5 (<35.7%)†	
At least one P or H [c]						**<0.001**
0	9 (20.5%)	<5 (<6.0%)†	0 (0.0%)	21 (25.9%)	0 (0.0%)	
1	26 (59.1%)	>39 (>47.0%)†	12 (37.5%)	42 (51.9%)	>5 (>35.7%)†	
2	9 (20.5%)	23 (27.7%)	15 (46.9%)	13 (16.0%)	<5 (<35.7%)†	
>2	0 (0.0%)	16 (19.3%)	5 (15.6%)	5 (6.2%)	<5 (<35.7%)†	
At least one comorbidity prior to inclusion using algorithms (%)	10 (22.7%)	21 (25.3%)	19 (59.4%)	11 (13.6%)	8 (57.1%)	**<0.001**
At least one comorbidity prior or after to inclusion using algorithms (%)	26 (59.1%)	63 (75.9%)	>27 (>84.4%)†	32 (39.5%)	14 (100.0%)	**<0.001**
At least one comorbidity prior to inclusion using at least one P or H [c] (%)	12 (27.3%)	27 (32.5%)	20 (62.5%)	17 (21.0%)	8 (57.1%)	**<0.001**

	Cluster 8 n=44 (4.2%)	**Cluster 9** n=83 (7.9%)	**Cluster 10** n=32 (3.1%)	**Cluster 11** n=81 (7.7%)	**Cluster 12** n=14 (1.3%)	**p-value** [a]
At least one comorbidity prior or after to inclusion using at least one P or H [c] (%)	35 (79.5%)	>78 (>94.0%)†	32 (100.0%)	60 (74.1%)	14 (100.0%)	**<0.001**
Prescription of DMTs						
At least one prescription of Avonex® (%)	6 (13.6%)	12 (14.5%)	<5 (<15.6%)†	16 (19.8%)	<5 (<35.7%)†	0.129
At least one prescription of Betaseron® (%)	10 (22.7%)	23 (27.7%)	<5 (<15.6%)†	14 (17.3%)	<5 (<35.7%)†	0.386
At least one prescription of Rebif® (%)	16 (36.4%)	39 (47.0%)	12 (37.5%)	30 (37.0%)	6 (42.9%)	0.288
At least one prescription of Copaxone® (%)	5 (11.4%)	23 (27.7%)	11 (34.4%)	65 (19.8%)	<5 (<35.7%)†	0.201
At least one prescription of an interferon-β (%)	28 (63.6%)	54 (65.1%)	16 (50.0%)	48 (59.3%)	9 (64.3%)	**0.049**
At least one prescription of a DMT (%)	30 (68.2%)	62 (74.7%)	19 (59.4%)	54 (66.7%)	>9 (>64.3%)†	0.216
Number of DMTs prescribed over follow-up (%)						**0.009**
0	14 (31.8%)	21 (25.3%)	13 (40.6%)	27 (33.3%)	<5 (<35.7%)†	
1	23 (44.8%)	33 (39.8%)	10 (31.3%)	>33 (>40.7%)†	<5 (<35.7%)†	
2	7 (15.9%)	23 (27.7%)	9 (28.1%)	15 (18.5%)	>5 (>35.7%)†	
>2	0 (0.0%)	6 (7.2%)	0 (0.0%)	<5 (<6.2%)†	0 (0.0%)	

	Cluster 8 n=44 (4.2%)	**Cluster 9** n=83 (7.9%)	**Cluster 10** n=32 (3.1%)	**Cluster 11** n=81 (7.7%)	**Cluster 12** n=14 (1.3%)	**p-value** [a]
Care consumption since inclusion						
Number of visits to GPs per person-year [d]	9.8 (4.9)	15.3 (6.8)	12.3 (7.5)	6.0 (2.8)	28.7 (21.0)	**<0.001**
Number of visits to neurologists per person-year [d]	2.8 (1.3)	2.3 (2.1)	2.3 (1.2)	1.7 (0.8)	5.1 (4.5)	**<0.001**
Number of visits to internal medicine specialists per person-year [d]	0.5 (0.7)	0.8 (1.1)	1.3 (2.1)	0.3 (0.7)	1.7 (1.7)	**<0.001**
Number of visits to symptoms specialists per person-year [d, e]	1.2 (1.1)	0.9 (0.8)	1.5 (1.6)	1.0 (1.1)	2.6 (2.3)	**<0.001**
Number of visits to psychiatrists per person-year [d]	0.1 (0.3)	1.1 (2.8)	9.8 (9.1)	0.1 (0.2)	9.9 (8.9)	**<0.001**
At least one hospital admission (%)	23 (52.3%)	63 (75.9%)	>27 (>84.4%)†	40 (49.4%)	>9 (>64.3%)†	**<0.001**
Number of hospital admissions per person-year [d, f]	0.6 (0.6)	0.6 (0.6)	0.7 (0.7)	0.2 (0.2)	1.8 (2.3)	**<0.001**
Length of stay [d, f]	5.7 (13.1)	6.0 (17.1)	7.1 (14.5)	2.7 (5.5)	24.8 (75.0)	**<0.001**
Number of visits to one of the considered physicians [d, g]	14.3 (6.0)	20.5 (7.7)	27.2 (16.7)	9.1 (3.9)	48.0 (27.3)	**<0.001**
Number of visits to any physician [d]	20.3 (10.0)	28.1 (10.5)	34.3 (28.1)	13.4 (6.8)	55.5 (31.3)	**<0.001**

[a] P-value of the comparison of the groups according to Kruskal-Wallis, Chi-squared test or Fisher exact test, if needed. [b] Mean (Standard deviation). [c] P: Physician claim or H: Hospital admission. [d] Mean (standard deviation). [e] Symptoms specialist: PMR physician, ophthalmologist or urologist. [f] Only for patients having at least one hospital admission. [g] GP, neurologist, internal medicine specialist, physician specialized in most frequent MS symptoms, psychiatrist. † Data presented are restricted to prevent direct or residual disclosure of identifiable data.

Table 43-Caractéristiques des clusters de la partition en six groupes obtenue suite à la multichannel sequence analysis appliquée sur la cohort 2 (N=3 180)

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**p-value** [a]
	n=121 (4.8%)	n=1622 (51.0%)	n=251 (7.9%)	n=352 (11.1%)	n=629 (19.8%)	n=205 (6.4%)	
No. of women (%)	103 (85.1%)	1163 (71.7%)	171 (68.1%)	255 (72.4%)	512 (81.4%)	130 (63.4%)	**<0.001**
Birth year [b]	1953 (1928-1970)	1953 (1920-1986)	1948 (1918-1978)	1948 (1921-1977)	1951 (1913-1978)	1947 (1914-1985)	**<0.001**
Age at inclusion [b]	42.1 (25.4-68.0)	42.8 (9.1-76.0)	47.4 (17.9-77.8)	47.8 (18.6-74.3)	44.2 (17.3-82.8)	48.6 (11.0-81.5)	**<0.001**
Year of MS onset [b]	1984 (1953-1995)	1986 (1942-1995)	1983 (1944-1995)	1983 (1942-1995)	1984 (1954-1995)	1980 (1942-1995)	**<0.001**
Age at MS onset [b]	28.1 (4.4-57.9)	30.0 (3.4-67.8)	30.6 (5.7-75.2)	32.6 (6.6-64.9)	30.4 (6.0-65.4)	31.4 (8.3-63.5)	**0.003**
Age at MS onset categorized (%)							**<0.001**
≤20	20 (16.5%)	158 (9.7%)	23 (9.2%)	32 (9.1%)	81 (12.9%)	30 (14.6%)	
]20;30]	>45 (>37.2%)	649 (40.0%)	96 (38.2%)	111 (31.5%)	223 (35.4%)	62 (30.2%)	
]30;40]	41 (33.9%)	540 (33.3%)	72 (28.9%)	113 (32.1%)	190 (30.2%)	65 (31.7%)	
]40;50]	10 (8.3%)	224 (13.8%)	38 (15.1%)	74 (21.0%)	110 (17.5%)	28 (13.7%)	
>50	<5 (<4.1%)	51 (3.1%)	22 (8.8%)	22 (6.3%)	25 (4.0%)	20 (9.8%)	
Duration since MS onset at inclusion (years) [b]	12.0 (0.2-43.0)	10.0 (0.1-54.0)	13.0 (0.1-52.0)	13.0 (0.1-54.0)	12.0 (0.1-42.0)	16.0 (0.1-54.0)	**<0.001**
Follow-up (years) [b]	13.0 (9.1-13.0)	13.0 (10.3-13.0)	9.7 (6.5-11.8)	13.0 (9.5-13.0)	13.0 (9.0-13.0)	5.7 (4.0-8.2)	**<0.001**
No. of deaths (%)	<5 (<4.1%)†	36 (2.2%)	177 (70.5%)	15 (4.3%)	33 (5.2%)	136 (66.3%)	**<0.001**
No. of deaths per person-year (/1000)	<3.2†	1.7	74.8	3.3	4.1	116.2	-

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**p-value** [a]
	n=121 (4.8%)	n=1622 (51.0%)	n=251 (7.9%)	n=352 (11.1%)	n=629 (19.8%)	n=205 (6.4%)	
Socioeconomic quantile at inclusion							0.716
1 (low)	22 (18.2%)	250 (15.4%)	50 (19.9%)	58 (16.5%)	106 (16.9%)	40 (19.5%)	
2	23 (19.0%)	288 (17.8%)	40 (15.9%)	59 (16.8%)	118 (18.8%)	40 (19.5%)	
3	15 (12.4%)	341 (21.0%)	47 (18.7%)	76 (21.6%)	124 (19.7%)	42 (20.5%)	
4	27 (22.3%)	369 (22.7%)	46 (18.3%)	77 (21.9%)	141 (22.4%)	41 (20.0%)	
5 (high)	28 (23.1%)	343 (21.1%)	62 (21.1%)	77 (21.9%)	127 (20.2%)	37 (18.0%)	
Missing	6 (5.0%)	31 (1.9%)	6 (2.4%)	5 (1.4%)	13 (2.1%)	5 (2.4%)	
EDSS score at inclusion							**<0.001**
Mild	37 (30.6%)	562 (34.6%)	48 (19.1%)	95 (27.0%)	208 (33.1%)	48 (23.4%)	
Moderate	28 (23.1%)	313 (19.3%)	40 (15.9%)	71 (20.2%)	155 (24.6%)	49 (23.9%)	
Severe	17 (14.0%)	254 (15.7%)	110 (43.8%)	87 (24.7%)	120 (19.1%)	88 (42.9%)	
Missing	39 (32.2%)	493 (30.4%)	53 (21.1%)	99 (28.1%)	146 (23.2%)	20 (9.8%)	
EDSS score at the end of follow-up							**<0.001**
Mild	44 (36.4%)	580 (35.8%)	52 (20.7%)	78 (22.2%)	186 (29.6%)	37 (18.0%)	
Moderate	38 (31.4%)	427 (26.3%)	38 (15.1%)	91 (25.9%)	190 (30.2%)	37 (18.0%)	
Severe	39 (32.2%)	615 (37.9%)	161 (64.1%)	183 (52.0%)	253 (40.2%)	131 (63.9%)	
Global ARMSS score at inclusion [b]	5.3 (0.5-9.8)	4.8 (0.2-10.0)	7.1 (0.3-9.9)	5.5 (0.2-9.9)	5.3 (0.0-9.9)	6.2 (0.5-10.0)	**<0.001**
Global ARMSS score at the end of follow-up [b]	3.9 (0.1-10.0)	4.4 (0.1-10.0)	7.4 (0.3-10.0)	5.4 (0.1-10.0)	4.8 (0.1-10.0)	7.5 (0.3-10.0)	**<0.001**

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**p-value** [a]
	n=121 (4.8%)	n=1622 (51.0%)	n=251 (7.9%)	n=352 (11.1%)	n=629 (19.8%)	n=205 (6.4%)	
Number of relapses reported in BC MS database [b]	1.0 (0.0-8.0)	0.0 (0.0-13.0)	0.0 (0.0-11.0)	0.0 (0.0-11.0)	1.0 (0.0-12.0)	0.0 (0.0-11.0)	**<0.001**
Initial course of the disease							**0.003**
Relapsing-remitting MS (%)	>116 (>95.9%)†	1496 (92.2%)	223 (88.8%)	321 (91.2%)	585 (93.0%)	177 (86.3%)	
Primary progressive MS (%)	<5 (<4.1%)†	126 (7.8%)	28 (11.2%)	31 (8.8%)	44 (7.0%)	28 (13.7%)	
Secondary progressive course at the end of follow-up (%)	46 (38.0%)	629 (38.8%)	140 (55.8%)	180 (51.1%)	262 (41.7%)	123 (60.0%)	**<0.001**
Comorbidities							
Hypertension							
Using algorithms							
Prior to inclusion (%)	<5 (<4.1%)†	32 (2.0%)	12 (4.8%)	18 (5.1%)	27 (4.3%)	9 (4.4%)	**0.007**
Prior or after the inclusion (%)	34 (28.1%)	323 (19.9%)	52 (20.7%)	130 (36.9%)	197 (31.3%)	27 (13.2%)	**<0.001**
At least one P or H [c]							
Prior to inclusion (%)	<5 (<4.1%)†	53 (3.3%)	15 (6.0%)	25 (7.1%)	36 (5.7%)	14 (6.8%)	**0.001**
Prior or after the inclusion (%)	56 (46.3%)	497 (30.6%)	75 (29.9%)	171 (48.6%)	275 (43.7%)	43 (21.0%)	**<0.001**
Hyperlipidemia							
Using algorithms							
Prior to inclusion (%)	0 (0.0%)	6 (0.4%)	0 (0.0%)	<5 (<1.4%)†	<5 (<0.8%)†	<5 (<2.4%)†	0.426
Prior or after the inclusion (%)	17 (14.0%)	131 (8.1%)	18 (7.2%)	46 (13.1%)	84 (13.4%)	<5 (<2.4%)†	**<0.001**

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**p-value** [a]
	n=121 (4.8%)	n=1622 (51.0%)	n=251 (7.9%)	n=352 (11.1%)	n=629 (19.8%)	n=205 (6.4%)	
At least one P or H [c]							
Prior to inclusion (%)	<5 (<4.1%)†	11 (0.7%)	0 (0.0%)	<5 (<1.4%)†	7 (1.1%)	<5 (<2.4%)†	**0.050**
Prior or after the inclusion (%)	30 (24.8%)	317 (19.5%)	32 (12.7%)	80 (22.7%)	172 (27.3%)	16 (7.8%)	**<0.001**
Diabetes							
Using algorithms							
Prior to inclusion (%)	<5 (<4.1%)†	10 (0.6%)	7 (2.8%)	5 (1.4%)	14 (2.2%)	<5 (<2.4%)†	0.297
Prior or after the inclusion (%)	15 (12.4%)	101 (6.2%)	32 (12.7%)	39 (11.1%)	86 (13.7%)	14 (6.8%)	**<0.001**
At least one P or H [c]							
Prior to inclusion (%)	<5 (<4.1%)†	11 (0.7%)	7 (2.8%)	5 (1.4%)	16 (2.5%)	6 (2.9%)	0.043
Prior or after the inclusion (%)	31 (25.6%)	184 (11.3%)	40 (15.9%)	61 (17.3%)	144 (22.9%)	26 (12.7%)	**<0.001**
Chronic lung disease							
Using algorithms							
Prior to inclusion (%)	<5 (<4.1%)†	16 (1.0%)	6 (2.4%)	<5 (<1.4%)†	16 (2.5%)	<5 (<2.4%)†	0.742
Prior or after the inclusion (%)	17 (14.0%)	116 (7.2%)	39 (15.5%)	40 (11.4%)	123 (19.6%)	20 (9.8%)	**<0.001**
At least one P or H [c]							
Prior to inclusion (%)	6 (5.0%)	25 (1.5%)	8 (3.2%)	<5 (<1.4%)†	18 (2.9%)	<5 (<2.4%)†	0.368
Prior or after the inclusion (%)	32 (26.4%)	222 (13.7%)	58 (23.1%)	73 (20.7%)	187 (29.7%)	30 (14.6%)	**<0.001**

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**p-value** [a]
	n=121 (4.8%)	n=1622 (51.0%)	n=251 (7.9%)	n=352 (11.1%)	n=629 (19.8%)	n=205 (6.4%)	
Mood or Anxiety disorder							
Using algorithms							
Prior to inclusion (%)	77 (63.6%)	160 (9.9%)	34 (13.5%)	59 (16.8%)	179 (28.5%)	50 (24.4%)	**<0.001**
Prior or after the inclusion (%)	>116 (>95.9%)†	504 (31.1%)	99 (39.4%)	151 (42.9%)	413 (65.7%)	95 (46.3%)	**<0.001**
At least one P or H [c]							
Prior to inclusion (%)	82 (67.8%)	310 (19.1%)	61 (24.3%)	90 (25.6%)	247 (39.3%)	72 (35.1%)	**<0.001**
Prior or after the inclusion (%)	>116 (>95.9%)†	1161 (71.6%)	197 (78.5%)	294 (83.5%)	579 (92.1%)	152 (74.1%)	**<0.001**
No. of comorbidities prior to inclusion (%)							
Using algorithms							**<0.001**
0	42 (34.7%)	1406 (86.7%)	>193 (>76.9%)†	275 (78.1%)	406 (64.5%)	>142 (>69.3%)†	
1	74 (61.2%)	208 (12.8%)	53 (21.1%)	68 (19.3%)	207 (32.9%)	58 (28.3%)	
2	5 (4.1%)	8 (0.5%)	<5 (<2.0%)†	9 (2.6%)	16 (2.5%)	<5 (<2.4%)†	
>2	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
At least one P or H [c]							**<0.001**
0	34 (28.1%)	1237 (76.3%)	>159 (>63.3%)†	>235 (>66.8%)†	>333 (>52.9%)†	115 (56.1%)	
1	76 (62.8%)	360 (22.2%)	82 (32.7%)	99 (28.1%)	261 (41.5%)	82 (40.0%)	
2	11 (9.1%)	25 (1.5%)	<5 (<2.0%)	13 (3.7%)	30 (4.8%)	8 (3.9%)	
>2	0 (0.0%)	0 (0.0%)	<5 (<2.0%)†	<5 (<1.4%)†	<5 (<0.8%)†	0 (0.0%)	

	Cluster 1 n=121 (4.8%)	**Cluster 2** n=1622 (51.0%)	**Cluster 3** n=251 (7.9%)	**Cluster 4** n=352 (11.1%)	**Cluster 5** n=629 (19.8%)	**Cluster 6** n=205 (6.4%)	**p-value** [a]
No. of comorbidities prior or after the inclusion (%)							
Using algorithms							**<0.001**
0	<5 (<4.1%)†	739 (45.6%)	90 (35.9%)	93 (26.4%)	98 (15.6%)	81 (39.5%)	
1	>57 (>47.1%)†	637 (39.3%)	107 (42.6%)	152 (43.2%)	270 (42.9%)	95 (46.3%)	
2	38 (31.4%)	209 (12.9%)	33 (13.1%)	73 (20.7%)	174 (27.7%)	22 (10.7%)	
>2	21 (17.4%)	37 (2.3%)	21 (8.4%)	34 (9.7%)	87 (13.8%)	7 (3.4%)	
At least one P or H [c]							**<0.001**
0	0 (0.0%)	214 (13.2%)	31 (12.4%)	20 (5.7%)	13 (2.1%)	30 (14.6%)	
1	38 (31.4%)	714 (44.0%)	99 (39.4%)	112 (31.8%)	173 (27.5%)	108 (52.7%)	
2	33 (27.3%)	464 (28.6%)	75 (29.9%)	127 (36.1%)	225 (35.8%)	48 (23.4%)	
>2	50 (41.3%)	230 (14.2%)	46 (18.3%)	93 (26.4%)	218 (34.7%)	19 (9.3%)	
At least one comorbidity prior to inclusion using algorithms (%)	79 (65.3%)	216 (13.3%)	56 (22.3%)	77 (21.9%)	223 (35.5%)	62 (30.2%)	**<0.001**
At least one comorbidity prior or after to inclusion using algorithms (%)	>116 (>95.9%)†	883 (54.4%)	161 (64.1%)	259 (73.6%)	531 (84.4%)	124 (60.5%)	**<0.001**
At least one comorbidity prior to inclusion using at least one P or H [c] (%)	87 (71.9%)	385 (23.7%)	86 (34.3%)	113 (32.1%)	292 (46.4%)	90 (43.9%)	**<0.001**
At least one comorbidity prior or after to inclusion using at least one P or H [c] (%)	121 (100.0%)	1408 (86.8%)	220 (87.6%)	332 (94.3%)	616 (97.9%)	175 (85.4%)	**<0.001**

	Cluster 1 n=121 (4.8%)	**Cluster 2** n=1622 (51.0%)	**Cluster 3** n=251 (7.9%)	**Cluster 4** n=352 (11.1%)	**Cluster 5** n=629 (19.8%)	**Cluster 6** n=205 (6.4%)	**p-value** [a]
Prescription of DMTs							
At least one prescription of Avonex® (%)	18 (14.9%)	95 (5.9%)	5 (2.0%)	25 (7.1%)	63 (10.0%)	<5 (<2.4%)†	**<0.001**
At least one prescription of Betaseron® (%)	24 (19.8%)	258 (15.9%)	23 (9.2%)	60 (17.0%)	124 (19.7%)	17 (8.3%)	**<0.001**
At least one prescription of Rebif® (%)	15 (12.4%)	212 (13.1%)	18 (7.2%)	58 (16.5%)	106 (16.9%)	13 (6.3%)	**<0.001**
At least one prescription of Copaxone® (%)	23 (19.0%)	101 (6.2%)	9 (3.6%)	31 (8.8%)	67 (10.7%)	<5 (<2.4%)†	**<0.001**
At least one prescription of an interferon-β (%)	42 (34.7%)	483 (29.8%)	40 (15.9%)	112 (31.8%)	232 (36.9%)	24 (11.7%)	**<0.001**
At least one prescription of a DMT (%)	48 (39.7%)	520 (32.1%)	43 (17.1%)	121 (34.4%)	251 (39.9%)	26 (12.7%)	**<0.001**
Number of DMTs prescribed over follow-up (%)							**<0.001**
0	73 (60.3%)	1102 (67.9%)	>204 (>81.3%)†	231 (65.6%)	378 (60.1%)	>175 (>85.4%)†	
1	22 (18.2%)	396 (24.4%)	32 (12.7%)	79 (22.4%)	169 (26.9%)	17 (8.3%)	
2	21 (17.4%)	102 (6.3%)	10 (4.0%)	31 (8.8%)	58 (9.2%)	8 (3.9%)	
>2	5 (4.1%)	22 (1.4%)	<5 (<2.0%)†	11 (3.1%)	24 (3.8%)	<5 (<2.4%)†	

	Cluster 1 n=121 (4.8%)	**Cluster 2** n=1622 (51.0%)	**Cluster 3** n=251 (7.9%)	**Cluster 4** n=352 (11.1%)	**Cluster 5** n=629 (19.8%)	**Cluster 6** n=205 (6.4%)	**p-value** [a]
Care consumption since inclusion							
Number of visits to GPs per person-year [d]	14.8 (8.3)	6.3 (2.9)	11.2 (7.4)	10.3 (3.9)	17.9 (8.0)	12.6 (8.6)	**<0.001**
Number of visits to neurologists per person-year [d]	1.5 (0.9)	0.9 (0.6)	0.8 (0.7)	1.1 (0.7)	1.2 (0.9)	1.2 (1.8)	**<0.001**
Number of visits to internal medicine specialists per person-year [d]	0.8 (0.9)	0.4 (0.7)	1.0 (1.9)	0.9 (2.6)	1.1 (2.1)	1.9 (8.4)	**<0.001**
Number of visits to symptoms specialists per person-year [d, e]	1.3 (1.3)	0.5 (0.6)	0.9 (1.3)	1.9 (1.3)	1.4 (1.4)	1.4 (1.8)	**<0.001**
Number of visits to psychiatrists per person-year [d]	6.1 (5.6)	0.1 (0.4)	0.2 (1.1)	0.2 (0.4)	0.2 (0.8)	0.5 (1.7)	**<0.001**
At least one hospital admission (%)	106 (87.6%)	1158 (71.4%)	215 (85.7%)	325 (92.3%)	586 (93.2%)	169 (82.4%)	**<0.001**
Number of hospital admissions per person-year [d, f]	0.6 (0.5)	0.3 (0.3)	0.9 (0.9)	0.5 (0.5)	0.6 (0.8)	1.8 (2.2)	**<0.001**
Length of stay [d, f]	7.3 (17.4)	6.2 (21.8)	16.3 (51.3)	7.0 (28.7)	7.6 (46.0)	27.9 (105.9)	**<0.001**
Number of visits to one of the considered physicians [d, g]	24.5 (11.8)	8.1 (3.4)	14.2 (8.7)	14.4 (5.1)	21.8 (9.4)	17.6 (13.4)	**<0.001**
Number of visits to any physician [d]	37.1 (17.8)	14.3 (7.1)	22.7 (12.9)	24.4 (9.4)	34.7 (14.8)	29.8 (19.7)	**<0.001**

[a] P-value of the comparison of the groups according to Kruskal-Wallis, Chi-squared test or Fisher exact test, if needed. [b] Mean (Standard deviation). [c] P: Physician claim or H: Hospital admission. [d] Mean (standard deviation). [e] Symptoms specialist: PMR physician, ophthalmologist or urologist. [f] Only for patients having at least one hospital admission. [g] GP, neurologist, internal medicine specialist, physician specialized in most frequent MS symptoms, psychiatrist. [†] Data presented are restricted to prevent direct or residual disclosure of identifiable data.

UNIVERSITE BRETAGNE LOIRE / BIOLOGIE SANTE

UNIVERSITÉ DE RENNES 1

Titre : Parcours de soins des patients atteints de sclérose en plaques à partir des données médico-administratives en France

Mots clés : Sclérose en plaques ; Parcours de soins ; Analyse de séquences ; Système National des Données de Santé

Résumé : La sclérose en plaques (SEP) est une maladie neurologique chronique du jeune adulte affectant environ 100 000 personnes en France. Le parcours de soins dans la SEP implique plusieurs professionnels de santé médicaux (généralistes, neurologues) et paramédicaux (kinésithérapeutes, infirmiers). Actuellement, aucune recommandation sur ces parcours n'a été définie ; peu de données existent sur le recours aux soins des individus ayant une SEP ou l'utilisation des traitements spécifiques à la SEP en France. L'objectif principal de cette thèse était d'étudier les parcours de soins des personnes ayant une SEP en France, à partir du SNDS (Système National des Données de Santé), afin de décrire les consommations de soins et de mettre en évidence une typologie de parcours. Les objectifs secondaires étaient d'étudier l'utilisation des traitements de fond de la SEP en France, et d'étudier la faisabilité de mesurer le niveau de handicap moteur dans ces bases. Au total sur la période de suivi de 2010 à 2015, 112 745 patients ont été identifiés, dont 47,4% avaient reçu au moins une délivrance d'un traitement de fond spécifique à la SEP. Une typologie a été obtenue identifiant cinq groupes cliniquement distincts. Un indicateur de quantification du niveau de handicap moteur, pouvant être répliqué dans d'autres études, a été proposé afin d'enrichir les données du SNDS. Au travers des différentes analyses réalisées et des questions méthodologiques soulevées, des éléments clés permettant l'utilisation des méthodes d'analyse de séquences en santé, notamment la multichannel sequence analysis, ont pu être mis en évidence.

Title : Care pathways of persons with multiple sclerosis in France using administrative data

Keywords : Multiple sclerosis; Care pathways; State sequence analysis; French National Health Data System

Abstract : Multiple sclerosis (MS) is a chronic neurological disease starting in young adulthood and affecting about 100,000 persons in France. Care pathways in MS involve both medical (general practitioners, neurologists) and paramedical (physiotherapists and nurses) health care professionals. So far, no recommendation on care pathways in MS exists in France. Moreover, few data are available on care-seeking of persons with MS (PwMS) and the utilization of MS disease-modifying therapies (DMTs) in France. The main objective of this PhD thesis was to study the care pathways of PwMS in France using data from SNDS (French National Health Data System), in order to describe care-seeking and to create a typology of pathways. The secondary objectives were to study MS DMTs utilization in France, and the feasibility to measure the level of motor disability in SNDS. Over the 2010-2015 study period, 112,745 PwMS were identified. Amongst them, 47.4% had at least one delivery of a MS-specific DMT. A typology was obtained allowing the identification of five clinically distinct groups of patients. A parameter quantifying the level of motor disability in SNDS, which could be replicated in other studies, was defined. Thanks to the different analyses and raised methodological questions, key-elements allowing the use of state sequence analysis, especially multichannel sequence analysis, in health field were highlighted.

Univ Rennes, EHESP, REPERES – EA 7449
15 avenue du Professeur Léon-Bernard – CS 74312, 35043 RENNES Cedex