HAL
open science

Etude de l'effet des stratégies thérapeutiques en conditions réelles d'utilisation sur la progression du handicap chez les patients atteints de sclérose en plaques rémittente

Mathilde Lefort

To cite this version:

Mathilde Lefort. Etude de l'effet des stratégies thérapeutiques en conditions réelles d'utilisation sur la progression du handicap chez les patients atteints de sclérose en plaques rémittente. Médecine humaine et pathologie. Université de Rennes, 2019. Français. NNT : 2019REN1B050 . tel-03510267

HAL Id: tel-03510267

https://theses.hal.science/tel-03510267

Submitted on 4 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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
Mathilde LEFORT

Etude de l'effet des stratégies thérapeutiques en conditions réelles d'utilisation sur la progression du handicap chez les patients atteints de sclérose en plaques rémittente

Thèse présentée et soutenue à Rennes le 03 décembre 2019

Unité de recherche :
Univ Rennes, EHESP, REPERES - EA 7449
Inserm CIC-P 1414, Centre Hospitalier Universitaire de Rennes

Rapporteurs avant soutenance :

Florence Tubach — Professeur des Universités – Praticien Hospitalier, Hôpital Pitié-Salpêtrière, Paris
Eric Thouvenot — Professeur des Universités – Praticien Hospitalier, CHU de Nîmes

Composition du Jury :

Président : Francis Guillemin — Professeur des Universités – Praticien Hospitalier, CHU de Nancy

Rapporteurs :
Florence Tubach — Professeur des Universités – Praticien Hospitalier, Hôpital Pitié-Salpêtrière, Paris
Eric Thouvenot — Professeur des Universités – Praticien Hospitalier, CHU de Nîmes

Examinateurs :
Francis Guillemin — Professeur des Universités – Praticien Hospitalier, CHU de Nancy
Mar Tintoré — Professeur de Neurologie, Hôpital universitaire Vall d'Hebron, Barcelone
Nathalie Costet — Ingénieur de recherche, IRSET, Rennes

Dir. de thèse : Emmanuelle Leray — Professeur de l'EHESP, EA 7449 REPERES, Rennes
Co-dir. de thèse : Gilles Edan — Professeur des Universités – Praticien Hospitalier, CHU de Rennes

Cette thèse a été réalisée dans le cadre du Réseau Doctoral en Santé Publique animé par l'Ecole des Hautes Etudes en Santé Publique

A ma famille

That's what real love amounts to - letting a person be what he really is.
Jim Morrison

Quoi que vous puissiez faire, quoi que vous rêviez, commencez-le.
La hardiesse a du génie, de la force et de la magie.

J Wolfgang con Goethe

Remerciements

Cette thèse n'aurait pu aboutir sans ces personnes que je souhaite vivement remercier, que ce soit sur le plan professionnel ou personnel, sans ces rencontres et ces échanges, ce travail aurait été tout autre.

Mes premiers et plus chaleureux remerciements vont à mes encadrants : Docteure Emmanuelle Leray et Professeur Gilles Edan.

Emmanuelle, c'est déjà avec beaucoup d'émotions que j'écris ces mots. Après quelques mois au sein du département Métis, alors que nous nous croisions seulement pour les moments conviviaux, tu m'as fait confiance en m'offrant une proposition de CDD de deux ans. Notre première rencontre scientifique et nos échanges m'ont rapidement conduite à pousser la porte de ton bureau pour te formuler mon envie d'aller plus loin qu'un CDD de deux ans. Les trois années suivantes n'ont fait que me confirmer que j'avais fait le bon choix. Merci pour ton soutien quotidien sur le plan scientifique évidemment, je n'aurais pas pu mieux découvrir l'épidémiologie de la SEP, sur le plan humain, dans les moments plus difficiles et plus stressants de cette thèse où tu m'as toujours soutenue et encouragée.

Monsieur Edan, échanger avec vous a toujours été riche et me permettait de prendre du recul sur mon travail. Je souhaite vous remercier pour vos conseils et le partage de votre expérience clinique qui a contribué à ma compréhension de la prise en charge de cette pathologie complexe. Je vous remercie également de m'avoir accueillie à l'Association Neuro Bretagne.

Je remercie les membres de mon jury qui ont accepté de juger cette thèse : au professeure Florence Tubach et au professeur Eric Thouvenot qui me font l'honneur d'être rapporteurs de ce travail ; au professeur Francis Guillemin, au professeure Mar Tintoré et au Docteure Nathalie Costet pour leur participation en tant qu'examinateurs.

Ce travail n'aurait pas été le même si plusieurs rencontres n'avaient pas eu lieu. L'une des plus marquantes est ma rencontre avec la professeure Sandra Vukusic. Sandra, ton accompagnement sur l'ensemble des travaux, tes remarques toujours utiles ont largement contribué à la qualité des productions. Je te remercie pour la confiance que tu m'as accordée en me proposant de participer au projet collaboratif faisant intervenir deux autres équipes

internationales. Cette mobilité a été une belle et riche expérience qui m'a définitivement marquée sur le plan professionnel comme personnel, alors merci d'avoir pensé à moi. Je te remercie également d'avoir accepté de faire partie de mon comité de suivi de thèse.

Yohann Foucher, je te remercie pour le temps que tu m'as accordé, pour ta patience. Bénéficier de ton expertise a été une vraie plus-value pour cette thèse.

Tomas Kalincik, je te remercie de m'avoir accueillie dans ton équipe à l'université de Melbourne.

Ronan Garlantezec, je vous remercie d'avoir accepté de faire partie de mon comité de suivi et d'avoir été mon tuteur auprès de l'école doctorale biologie santé.

Ont également contribué à ces travaux : dans le cadre du projet INTERIMS : Rémi Lenain et Damien Le Port, dans le cadre du projet FraDenCORe : David Laplaud, Sifat Sharmin, Johanna Andersen, Melinda Magyari, dans le cadre des projets ACTION-MS et STRATEGY-MS : Romain Casey.

Enfin, pour les conseils, relecture et discussions enrichissantes, je souhaite remercier Professeur Emmanuel Oger, Docteure Laure Michel et Docteure Emmanuelle Le Page.

Bien sûr des remerciements particuliers pour le Professeur Emmanuel Oger et le Docteur Olivier Grimaud en tant que co-directeurs de l'équipe d'accueil REPERES. Aux membres de l'équipe : merci pour les échanges qu'ils se soient passés dans nos salles de réunions, en congrès ou autour d'un verre à Prague ou à Philadelphie.

Je souhaite remercier l'ensemble des membres de l'association Neuro Bretagne, en particulier merci à Claire Gallien, l'assistante de l'association qui est d'une aide précieuse dans les démarches administratives. Merci également à Marie-Pierre Chaudet pour son accompagnement sur le plan administratif lors de mon déplacement en Australie.

Je remercie l'Ecole des Hautes Etudes en Santé Publique (EHESP) pour m'avoir hébergée dans leurs locaux durant ces trois ans. En particulier, de nouveau, merci au Docteur Olivier Grimaud d'avoir accepté de me faire une place au sein de son département Métis de l'EHESP. J'ai une pensée pour l'ensemble des membres du département. Je souhaite remercier en particulier Nadine Migné, Séverine Salaud et Hélène Mignot qui m'ont toujours accompagnée dans les démarches administratives. Un grand et chaleureux merci à mes collègues proches qui rendent ou ont rendu le quotidien agréable, me donnant même parfois l'impression de venir rencontrer des ami.e.s plutôt que de me rendre au travail : Alice, Aurélie, Faustine, Hélène, Jonathan, Lola, Maxime, Michael, Pascal, Yann.

Pour finir, je souhaite remercier le Réseau Doctoral en santé publique, et plus particulièrement Karine Laboux et Emmanuelle Guevara. Je remercie aussi Christophe Le Rat pour son accompagnement dans les projets auxquels j'ai souhaité collaborer dans ce cadre. Nos échanges sont toujours ponctués par ta bonne humeur et ta gentillesse. Parmi les doctorants que j'ai eu l'occasion de rencontrer dans le cadre du Réseau, je souhaite formuler de vive remerciements aux collègues avec lesquels j'ai organisé le séminaire interdisciplinaire : Lola Neufcourt, Iris Cervenka, Vincent Schlegel. Le Réseau Doctoral, c'est aussi de belles rencontres : Noriane Cognez, Anne-Claire Binter, Samuel Thomas, Abigail Bourguignon.

Remerciements

Si sur le plan professionnel, il est certain que j'ai été accompagnée en témoigne la longue liste de premiers remerciements, sur le plan personnel, je ne saurais formuler à quel point votre présence à mes côtés a été et restera importante et primordiale. Merci à mes ami.e.s et à ma famille pour votre soutien, vos sourires et votre amour.

Vincent, merci pour ton soutien et tes mots toujours justes, m'encourageant à rester celle que je veux. Fin 2020, quand toi aussi, tu en seras à l'écriture de tes remerciements, on pourra chanter en cœur (et en boucle) une chanson que j'aurais pris soin de choisir entêtante.

Chloé et Catinus, nos rencontres, toujours ressourçantes, ont été un souffle, en particulier ces derniers mois. Les sourires de votre fille Abigaïl contribuent évidemment à ces moments de joie.

Anne-Hélène, nous arrivons toutes les deux à la concrétisation de nos projets professionnels qui ont beaucoup occupés nos vies ces derniers mois. J'ai hâte de nous retrouver.

Emilie, I'm back! De nouveau disponible pour les virées sur la côte, les retrouvailles du dimanche soir anti-déprime et les samedis soirs où l'esprit sera enfin libéré du lundi. Merci pour ta présence au quotidien, je sais toujours où sonner et j'espère que toi aussi.

Camille, je te remercie pour tes nombreux messages d'encouragements, ta compréhension et ta patience. Nous avons du temps à rattraper.

A la famille Pticrabes : Aurélie, Baptiste et Bruno, sans oublier Noa, nos week-ends sont toujours vivifiants (et rien à voir avec l'air marin vannetais). Ils se font plus rares mais, au risque de me répéter, on va rectifier le tir très vite. Rendez-vous au Paddy !

A la team « Galettes (sans) saucisses » : Mathilde, Pascal, Mandy, Marc, Emilie, merci pour les moments partagés, ces virées au marché des Lices et autres after-work aux Grands-Gamins (et parfois un peu plus loin…). Ils ont définitivement positivement marqué ces dernières années et la rennaise qui sommeille désormais en moi.

Mes pensées vont également à Ousseynou, Mathilde, et leur fille Marième, Jérôme, Solène, Yann et Faustine.

Evidemment, je n'oublie pas mes collègues devenus de « vrais » amis : Lola, Alice, Aurélie, Jonathan.

Lola, notre rencontre nourrit mon optimisme prouvant que derrière chaque évènement du positif se cache. Il aurait été bien dommage de ne pas partager cette aventure avec toi. Ta présence sur le bateau a été très importante, et ne t'en fais pas je reste présente à la barre pour les quelques vagues qu'il te reste à naviguer.

Alice, je me rappelle encore de ton arrivée au sein de la MS team, espérant qu'on s'entendrait bien. C'est plus que ce que j'avais imaginé. Hâte de retrouver le temps des « after work », ils nous ont donné l'occasion de partager des moments « tisanes » privilégiés qui piquent un peu les yeux le lendemain mais nous permettent de mieux nous connaître.

Jonathan, à notre longue histoire de cohabitation. J'ai été contente de partager les premières années de ma thèse avec Yann et toi dans le W39. Plus agréables encore sont les discussions voyage et shopping quand les couloirs se font vides et le partage d'un DBK tant que le mien est à l'orge fermenté (Donne-moi une Bière, je suis Ko).

Aurélie, définitivement élue Miss Belle Rencontre 2018, mon cœur de Bisounours a toujours plaisir à rencontrer le tien.

Les derniers mots vont à ma famille.

Merci à mes sœurs, Claire-Marie et Mélanie et à mon beau-frère Sébastien pour votre soutien, votre compréhension, votre attention et vos encouragements. En particulier, Claire-Marie, merci pour le temps que tu as passé à relire la plupart des chapitres de ce manuscrit.

A mes nièces, Agnès et Louise : vos sourires, vos éclats de rires, vos simples mots et vos regards d'enfants sont un souffle, une bulle de sérénité.

Enfin, je remercie très fort mes parents pour tout et en particulier leur soutien et leurs encouragements même si je ne suis pas toujours facile à suivre. Vous avez fait de moi une femme battante et m'avez donné les clés pour trouver la force de toujours réussir à m'épanouir que ce soit professionnellement ou personnellement. Aussi grand soit le merci que je veux vous adresser, il n'est rien à côté de ce que vous me donnez.

Financements

Cette thèse a été majoritairement financée grâce à un don de ROCHE SAS dans le cadre d'un mécénat, sans lien avec le contenu scientifique de ce manuscrit.

Un financement complémentaire a été octroyé en réponse à un appel à projet de la Fondation Eugène Devic EDMUS et de la fondation pour l'aide à la recherche sur la sclérose en plaques (ARSEP).

Dans le cadre de la mobilité internationale, une bourse de mobilité a été délivrée par la fondation pour l'aide à la recherche sur la sclérose en plaques (ARSEP).

Valorisations scientifiques

Articles

Articles en lien avec la thèse

Lefort M, Foucher Y, Lenain R, Vukusic S, Edan G, Leray E. Time-dependent propensity score matching: a novel method that warrants consideration in comparative effectiveness studies, *en révision dans Pharmacoepidemiology and Drug Safety depuis le 6 septembre 2019*

Andersen J[1], Sharmin S[1], Lefort M[1], Vukusic S, Butzkueven H, Leray E, Koch-Henriksen N, Kalincik T[2]; Laplaud DA[2], Magyari M[2], on behalf of MSBase, Danish MS Registry, and OFSEP. Understanding heterogeneity in comparative studies of treatment effectiveness in multiple sclerosis - Part 1: Validation of the comparisons of natalizumab and fingolimod, *en préparation, soumission prévue premier trimestre 2020*

Lefort M[1], Sharmin S[1], Andersen J[1], Butzkueven H, Laplaud DA, Koch-Henriksen N, Vukusic S, Magyari M[2], Kalincik T[2], Leray E[2], on behalf of MSBase, Danish MS Registry, and OFSEP. Understanding heterogeneity in comparative studies of treatment effectiveness in multiple sclerosis - Part 2: Effect of analytical methodology, *en préparation, soumission prévue premier trimestre 2020*

Sharmin S[1], Lefort M[1], Andersen J[1], Butzkueven H, Laplaud DA, Koch-Henriksen N, Leray E, Magyari M[2], Vukusic S[2], Kalincik T[2], on behalf of MSBase, Danish MS Registry, and OFSEP. Understanding heterogeneity in comparative studies of treatment effectiveness in multiple sclerosis - Part 3: Effect of cohort characteristics, *en préparation, soumission prévue premier trimestre 2020*

Lefort M, Vukusic S, Casey R, Edan G, Leray E. Use of therapeutic strategies in multiple sclerosis expert centers in France, *en préparation, soumission prévue fin d'année 2019*

[1]*contribution équivalente en tant que premier auteur*

[2]*contribution équivalente en tant que dernier auteur*

Articles sans lien avec la thèse

Kammerer-Jacquet S-F, Brunot A, Bensalah K, Campillo-Gimenez B, Lefort M, Bayat S, et al (2017). Hilar fat infiltration : A new prognostic factor in metastatic clear cell renal cell carcinoma with first-line sunitinib treatment. Urologic Oncology, 1–8.

Lefort M[1], Neufcourt L[1], Pannier B, Vaïsse B, Bayat S, Grimaud O, & Girerd X (2018). Sex differences in adherence to antihypertensive treatment in patients aged above 55: The French League Against Hypertension Survey (FLAHS). Journal of Clinical Hypertension, (April), 1496–1503.

Kammerer-Jacquet S-F., Brunot A, Lefort M, Bayat S, Peyronnet B, Verhoest G, et al (2019). Metastatic Clear-cell Renal Cell Carcinoma With a Long-term Response to Sunitinib: A Distinct Phenotype

Independently Associated With Low PD-L1 Expression. Clinical Genitourinary Cancer, 17(3), 169–176.

Le Meur N, Vigneau C, Lefort M, Lebbah S, Jais J. P, Daugas E, Bayat, S (2019). Categorical state sequence analysis and regression tree to identify determinants of care trajectory in chronic disease: Example of end-stage renal disease. Statistical Methods in Medical Research, 28(6), 1731–1740.

Roux J, Guilleux A, Lefort M, Leray E (2019) Use of health care services from patients with multiple sclerosis in France over 2010-2015: a nationwide population-based study using health administrative data. Multiple Sclerosis Journal – Experimental, Translational and Clinical (Accepté)

[1]*contribution équivalente en tant que premier auteur*

Communications orales

Lefort M, Edan G, Leray E. Propensity score methods: The study of the effect of β-interferon and glatiramer acetate on multiple sclerosis disability. 34th International Conference on Pharmacoepidemiology & Therapeutic Risk Management (ICPE). Prague, Czech Republic, August 22-26, 2018.

Lefort M, Edan G, Leray E. Etude des stratégies thérapeutiques en conditions réelles d'utilisation sur la progression du handicap chez les patients atteints de sclérose en plaques. 10ème Rencontres Scientifiques du Réseau Doctoral en Santé Publique. Bordeaux, France, Mars, 2019.

Lefort M, Foucher Y, Lenain R, Vukusic S, Edan G, Leray E Long-term effect of interferon-β and glatiramer acetate in real-world settings use on disability progression: input of time-dependent propensity score. 35th International Conference on Pharmacoepidemiology & Therapeutic Risk Management (ICPE). Philadelphia, Pennsylvania, United States of America, August 24-28, 2019.

Communications affichées

Lefort M, Goetz M, Edan G, Leray E. Long-term effect of interferon in real-world settings on disability progression in relapsing-remitting multiple sclerosis patients: A series of 1308 patients, Rennes, France. 24th Annual Meeting of the European Charcot Foundation, Baveno, Italy, Novembre 2016

Lefort M, Edan G, Leray E. Effet des interférons à long terme en conditions réelles d'utilisation sur la progression du handicap chez les patients atteints de rémittente : étude de 1308 patients, CHU de Rennes, France. Journées de Neurologie de Langue Française. Toulouse, France, Mars, 2017

Lefort M, Edan G, Leray E. Long-term effect of interferon-β in real-world settings on disability progression: A series of 2,451 relapsing-remitting multiple sclerosis patients, Rennes, France. ECTRIMS, Paris, France, Octobre, 2017

Lefort M, Edan G, Leray E. Effect of early treatment initiation on disability progression. 25th Annual Meeting of the European Charcot Foundation, Baveno, Italie, Décembre, 2017

Lefort M, Edan G, Leray E. Effet des interféron-β et de l'acétate de glatiramère sur la progression du handicap chez les patients atteints de sclérose en plaques. 9ème Rencontres Scientifiques du Réseau Doctoral en Santé Publique. Bordeaux, France, Mars, 2018.

Lefort M, Foucher Y, Vukusic S, Edan G, Leray E. Long-term effect of interferon-β and glatiramer acetate in real-world settings on disability progression: input of time-dependent propensity score. 26th annual meeting of the European Charcot Foundation. Baveno, Italie, Novembre, 2018.

Lefort M, Foucher Y, Vukusic S, Edan G, Leray E. Long-term effect of interferon-β and glatiramer acetate in real-world settings on disability progression: input of time-dependent propensity score. Journées scientifiques de l'école Doctorale Biologie Santé. Rennes, France, Décembre, 2018

Lefort M[1], Sharmin S[1], Andersen J[1], Butzkueven H, Laplaud DA, Koch-Henriksen N, Vukusic S, Magyari M[1], Kalincik T[1], Leray E[1]; on behalf of MSBase, Danish MS Registry, and OFSEP. Comparative effectiveness of natalizumab and fingolimod in different subgroups of patients with relapsing-remitting multiple sclerosis, Assises de l'OFSEP. Lyon, France, Mars, 2019

Lefort M, Vukusic S, Casey R, Edan G, Leray E. Effet de l'initiation précoce d'un traitement de fond sur la progression du handicap en sclérose en plaques rémittente : une étude observationnelle des données de l'OFSEP. Journées de Neurologie de Langue Française. Lille, France, Avril, 2019

Sharmin S[1], Lefort M[1], Andersen J[1], Butzkueven H, Laplaud DA, Koch-Henriksen N, Leray E, Magyari M[1], Vukusic S[1], Kalincik T[1], on behalf of MSBase, Danish MS Registry, and OFSEP. Comparative effectiveness of natalizumab and fingolimod in different subgroups of patients with relapsing-remitting multiple sclerosis. ECTRIMS, Stockholm, Suède, Septembre, 2019

Réseau Doctoral en Santé Publique animé par l'EHESP

Formations et séminaires suivis dans le cadre du Réseau Doctoral

Formations

- Design, Concept and Methods in Epidemiology, EHESP, Parisa Tehranifar, Paris, France, décembre 2016
- Spatial statistical analysis, EHESP, Cindy Padilla, Paris, France, janvier 2017
- Cours International D'Epidémiologie Appliquée (IDEA), Pascal Crepey, Rennes, France, mars 2017
- Critical windows of exposures and vulnerability, Robert Barouki and Aurore Gely-Pernot, Paris, France, février 2018
- Introduction à l'épidémiologie sociale, Université Pierre et Marie Curie, Pierre Chauvin, Paris, France, février 2018
- Introduction aux méthodes qualitatives, EHESP, Virginie Muniglia et Marie-Renée Guével, Rennes, France, février 2019

Séminaires

- La causalité en Santé Publique, Doctorants du Réseau Doctoral en Santé publique, Paris, France, décembre 2017
- Les inégalités sociales de santé au prisme au genre, Doctorants du Réseau Doctoral en Santé publique, Paris, France, mars 2019

Autres investissements au sein du Réseau Doctoral

Organisation d'évènements

- Organisation d'un séminaire interdisciplinaire intitulé La trajectoire des perturbateurs endocriniens du laboratoire à l'espace public : identification, (non-)politisation et réponses, Paris, France, Septembre 2018
- Organisation de (Pars en thèse), évènement de valorisation du doctorat à l'EHESP, Rennes, France, Novembre 2018

L'association PH'Doc, association de doctorants, post-doctorants et d'anciens doctorants du Réseau doctoral de l'EHESP

- 2017-2018 : secrétaire de l'association
- 2018-2019 : présidente de l'association

Mandats dans les instances du Réseau Doctoral en Santé publique :

- 2017-2018 : représentante des doctorant.e.s au Conseil des Formations de l'EHESP
- 2018-2019 : représentante des doctorant.e.s au Conseil scientifique de l'EHESP
- 2018-2019 : représentante des doctorant.e.s au Réseau Doctoral en Santé Publique

Sommaire

I. Introduction 2

A. La sclérose en plaques 4

1. Du mécanisme à l'expression de la SEP 4

1. 1 Le système nerveux central, cible du processus inflammatoire 4

1.2 Maladie démyélinisante responsable de la dégénérescence axonale 5

1.3 Diagnostic 7

1.4 Expression de la maladie 7

2. Phénotypes et activité de la SEP 8

2.1 Les formes de SEP 8

2.2 Activité IRM 9

2.3 Activité clinique : le handicap 10

3. Etiologie de la SEP 11

3.1 Epidémiologie de la SEP 11

3.2 Génétique et environnement, quels sont les facteurs associés à la SEP ? 12

4. Prise en charge de la SEP 12

B. Les traitements de fond de la SEP rémittente 15

1. Généralités sur les traitements de la SEP rémittente 15

2. Zoom sur les traitements de première ligne 17

3. Zoom sur les traitements de deuxième et troisième ligne 19

4. Stratégies thérapeutiques 20

C. Efficacité d'un traitement en vie réelle 22

1. Processus de développement d'un médicament 23

2. Evaluation de l'effet d'un médicament 24

2.1 Evaluation en conditions contrôlées 25

2.2 Evaluation en conditions réelles d'utilisation, en vie réelle 26

3. Quelles données observationnelles disponibles en SEP ? 28

D. Problématique et objectifs de la thèse 32

1. Problématique 32

2. Objectifs 33

II. Méthodes 36

A. Inférence causale 38

1. Approche graphique 39

1.1 Principe 39

1.2 Définitions 40
1.3 DAG en pharmacoépidémiologie 43
2. Approche contrefactuelle 44
B. Score de propension 46
1. Principe 46
2. Méthodes utilisant le score de propension 46
3. Appariement 48
4. Pondération 49
5. Sélection des variables pour spécifier le score de propension 50
6. Les hypothèses en pratique 50
C. Analyse de survie 52
1. Délai et fonctions : définitions et notations 52
2. Notion de censure 53
3. Temps de participation 54
4. Estimation de la fonction de survie 55
4.1 Définition et notation 55
4.2 Utilisation et interprétation 56
4.3 Comparaison de deux groupes à partir des courbes de survie 57
4.4 Pondération de la courbe de survie 57
3. Modèle à risques proportionnels de Cox 57
5.1 Hypothèses du modèle de Cox 58
5.2 Les modèles dérivés du modèle de Cox original 59
III. Efficacité d'un traitement en vie réelle : l'exemple des interférons-β et de l'acétate de glatiramère 62
A. Contexte 64
1. Une méta-analyse tendant à montrer le bénéfice de ces deux traitements à long terme 64
2. Quatre études à la loupe 65
B. Une réflexion méthodologique étape par étape 69
1. Population d'étude : date d'inclusion, groupe contrôle 69
2. Analyse statistique 70
C. Un score de propension dépendant du temps 71
1. Principe de la méthode 72
2. Adaptation du score dépendant du temps au contexte de la SEP rémittente 74
IV. Comparaison d'efficacité de traitements en vie réelle : l'exemple du natalizumab et du fingolimod 104
A. Contexte général de l'étude 106

1. Contexte scientifique 106
2. Contexte logistique 107
B. Les études originales 108
1. Des différences méthodologiques 108
2. Des résultats discordants 111
C. Une comparaison par étape 111
1. Les critères d'inclusion 111
2. L'hétérogénéité des pratiques 112
3. Les méthodes d'équilibre et d'analyse des données 113
4. Les critères de jugement 114
5. Déroulement de l'étude 115
D. Résultats issus de la mobilité 115
V. Evaluation des stratégies thérapeutiques 164
A. Contexte 166
1. Contexte scientifique 166
1.1 Le choix de la stratégie thérapeutique 166
1.2 Le moment de l'initiation du traitement 167
2. Contexte logistique 169
B. Description des stratégies thérapeutiques 170
C. Efficacité de l'initiation précoce d'un traitement : résultats préliminaires 187
1. Contexte 187
2. Patients et méthodes 188
2.1 Population d'étude 188
2.2 Exposition 189
2.3 Critères de jugement 189
2.4 Analyse statistique 189
3. Résultats 190
4. Discussion et perspectives 194
D. Comparaison de l'efficacité des stratégies thérapeutiques : perspectives 197
1. Contexte et objectif 197
2. Protocole d'étude 198
VI. Discussion générale 200
A. Les résultats principaux 202
1. Efficacité d'un traitement en vie réelle 203
2. Comparaison d'efficacité de traitements en vie réelle 203
3. Evaluation des stratégies thérapeutiques 204

B. La matière première : les connaissances, les données et les méthodes 205

1. Les données observationnelles : une richesse non exhaustive 206

1.1 OFSEP : des données longitudinales, rapportant l'état clinique du patient..... 206

1.2 OFSEP : des données collectées essentiellement en centre expert..... 207

1.3 OFSEP : des données non exhaustives..... 210

2. L'utilisation des méthodes..... 212

2.1 DAG de la relation entre les traitements de fond sur la progression du handicap dans la SEP rémittente..... 212

2.2 Les critères d'inclusion 215

2.3 Utilisation d'un score de propension..... 216

C. Les résultats et conclusions..... 219

1. L'exposition, une variable dépendante du temps 219

2. Les analyses en intention de traiter et en per protocol 222

3. Question d'efficacité à long terme en vie réelle : entre volonté et faisabilité 223

D. Un travail au carrefour de deux disciplines..... 224

Références 228

Annexes 246

A. Systèmes fonctionnels et échelle du score EDSS..... 248

B. Prise en charge des patients traités par natalizumab vis-à-vis du risque de LEMP 252

C. Exemple de l'interface du logiciel EDMUS pour un patient 254

D. Caractéristiques de 19 cohortes spécifiques SEP en disponible dans le monde..... 256

E. Sharmin S et al, Communication affichée, ECTRIMS, 2019 258

F. Evaluation de l'équilibre de l'étude de l'efficacité de l'initiation précoce du traitement, critère de jugement EDSS 6..... 260

G. Représentation de 1000 valeurs de Hazard Ratio obtenus après un appariement aléatoire 262

H. Exemple de graphique de conception d'une étude par REPEAT..... 264

Table des Figures

Figure 1 : Page de garde de la restitution des propos de J.M Charcot lors de ses leçons inaugurales en 1868, recueillis par D.M. Bourneville 4
Figure 2 : Description d'un neurone 5
Figure 3 : Attaque de la myéline par les lymphocytes dans la SEP 6
Figure 4 : Formes cliniques de SEP 8
Figure 5 : Imagerie cérébrale dans la SEP 9
Figure 6 : Disponibilité des traitements de fond de la SEP rémittente au cours du temps 16
Figure 7 : Utilisation des traitements de fond de la SEP rémittente entre 2010 et 2015 17
Figure 8 : Proposition d'algorithme thérapeutique dans la prise en charge médicamenteuse 21
Figure 9 : Nombre de publications mentionnant « propensity score » dans Pubmed jusqu'en 2018 27
Figure 10 : Exemple de graphe orienté acyclique ou « directed acyclic graph » (DAG) 40
Figure 11 : Variable « collider » et « non-collider » 40
Figure 12 : règle 1 de « D-séparation » 41
Figure 13 : règle 2 de « D-séparation » 41
Figure 14 : règle 3 de « D-séparation » 42
Figure 15 : règle 4 de « D-séparation » 42
Figure 16 : DAG en situation observationnelle, notion de chemin "porte-arrière" 43
Figure 17 : DAG d'un essai randomisé 44
Figure 18: Etudes portant sur l'efficacité des traitements de première ligne injectables incluses dans la méta-analyse de Signori et al, 2016 65
Figure 19 : Schématisation de l'appariement sur le score de propension dépendant du temps 73
Figure 20 : Contrainte d'appariement sur le score de propension dépendant du temps 75
Figure 21 : Distribution des scores de propension relatifs aux critères de jugement EDSS 3 et de poussées pour chaque définition de groupes 191
Figure 22 : Différences standardisées entre les groupes selon les trois définitions (score de propension associée aux critères de jugement EDSS 3 et poussées) 193
Figure 23 : Risque de survenue du critère de jugement associé à l'initiation précoce du traitement 194
Figure 24 : Distribution des stratégies thérapeutiques initiées par centre sur la période 2007-2013 198
Figure 25 : Evaluation des scores de propension dépendant du temps au cours du suivi 209
Figure 26 : DAG de la relation du traitement de fond sur la progression du handicap irréversible dans la SEP rémittente 213
Figure 27 : Relation causale du traitement de la SEP avec trois temps d'évaluation proposée par Karim et al, 2014 221

Tables des tableaux

Tableau 1 : Caractéristiques des études observationnelles portant sur l'efficacité des traitements de première ligne injectables 68
Tableau 2 : Comparaison des méthodologies utilisées dans chacune des trois études comparant le natalizumab et le fingolimod considérées dans l'étude collaborative 110
Tableau 3 : Critères d'inclusion pour chaque population cible étudiée dans le projet collaboratif portant sur l'efficacité des traitements de deuxième ligne 112
Tableau 4 : Analyses statistiques conduites dans le cadre du projet collaboratif portant sur l'efficacité des traitements de deuxième ligne 114
Tableau 5 : Durées de suivi selon la définition d'un traitement initié tôt et des critères de jugements considérés dans l'analyse de l'efficacité d'un traitement précoce 190
Tableau 6 : Caractéristiques des patients selon les définitions d'initiation précoce 192

Table des abréviations

AHSCT	Transplantation de cellules souches hématopoïétiques autologues ou « Autologous Hematopoietic Stem Cell Transplantation »
AMM	Autorisation de Mise sur le Marché
ANSM	Agence Nationale de Sécurité du Médicament et des produits de santé
ARSEP	Fondation pour l'Aide à la Recherche sur la SEP
ATE	Effet moyen du traitement dans la population entière ou « Average Treatment effect in the Entire population »
ATT	Effet moyen du traitement dans la population traitée ou « Average Treatment effect on the Treated »
ATU	Autorisation Temporaires d'Utilisation
BCMS	« British Columbia Multiple Sclerosis »
CIS	Syndrome Cliniquement Isolé ou « Clinical Isolated Syndrom »
CRC SEP	Centre de Ressources et de Compétences Sclérose en plaques
DAG	Graphes orientés acycliques ou « Directed Acyclic Graphs »
ECTRIMS	« European Commitee for Treatment and Research in Multiple Sclerosis »
EDMUS	« European Database for multiple Sclerosis »
EDSS	« Expanded Disability Status Scale »
HAS	Haute Autorité de Santé
HR	Rapport de Risque ou « Hazard Ratio »
IPTW	Pondération inverse sur les probabilités d'être traité ou « Inverse Probability of Treatment Weighting »
IRM	Imagerie à Résonnance Magnétique
IRR	Rapport du taux d'incidence ou « Incidence Rate Ratio »
LEMP	LEucoencéphalite Multifocale Progressive
MPR	Médecin de spécialité Physique et Réadaptation
OFSEP	Observatoire Française de la Sclérose En Plaques
RelSEP	Registre Lorrain des Scléroses En Plaques
RMST	Bénéfice de survie moyen restreint ou « Restricted Mean Survival Time »
SEP	Sclérose En Plaques
sIPTW	Pondération stabilisée inverse sur les probabilités d'être traité ou « Stabilized Inverse Probability of Treatment Weighting »
SNDS	Système National Des Données de Santé

Chapitre 1

I. Introduction

A. La sclérose en plaques 4
1. Du mécanisme à l'expression de la SEP 4
1. 1 Le système nerveux central, cible du processus inflammatoire 4
1.2 Maladie démyélinisante responsable de la dégénérescence axonale 5
1.3 Diagnostic 7
1.4 Expression de la maladie 7
2. Phénotypes et activité de la SEP 8
2.1 Les formes de SEP 8
2.2 Activité IRM 9
2.3 Activité clinique : le handicap 10
3. Etiologie de la SEP 11
3.1 Epidémiologie de la SEP 11
3.2 Génétique et environnement, quels sont les facteurs associés à la SEP ? 12
4. Prise en charge de la SEP 12
B. Les traitements de fond de la SEP rémittente 15
1. Généralités sur les traitements de la SEP rémittente 15
2. Zoom sur les traitements de première ligne 17
3. Zoom sur les traitements de deuxième et troisième ligne 19
4. Stratégies thérapeutiques 20
C. Efficacité d'un traitement en vie réelle 22
1. Processus de développement d'un médicament 23
2. Evaluation de l'effet d'un médicament 24
2.1 Evaluation en conditions contrôlées 25
2.2 Evaluation en conditions réelles d'utilisation, en vie réelle 26
3. Quelles données observationnelles disponibles en SEP ? 28
D. Problématique et objectifs de la thèse 32
1. Problématique 32
2. Objectifs 33

A. La sclérose en plaques

La sclérose en plaques, communément appelée SEP, est une affection du système nerveux central chronique marquée par des expressions cliniques diverses exposant chacune des fonctions assurées par le système nerveux central à des déficits[1]. Depuis la description à la fois clinique et anatomique de cas évocateurs de SEP de J.M. Charcot au XIX siècle (Figure 1), la connaissance de cette pathologie ne cesse de progresser (Hautecoeur, 2019).

HISTOLOGIE

DE LA

SCLÉROSE EN PLAQUES

LEÇON FAITE A L'HOSPICE DE LA SALPÊTRIÈRE PAR M. CHARCOT

Et recueillie par M. Bourneville

PARIS

IMPRIMERIE L. POUPART-DAVYL

30, RUE DU BAC, 30

1869

Figure 1 : Page de garde de la restitution des propos de J.M Charcot lors de ses leçons inaugurales en 1868, recueillis par D.M. Bourneville

1. Du mécanisme à l'expression de la SEP

1. 1 Le système nerveux central, cible du processus inflammatoire

Au cœur du fonctionnement de l'organisme, deux systèmes nerveux peuvent être distingués : le système nerveux central et le système nerveux périphérique. La fonction principale du système nerveux périphérique est d'assurer la communication entre les organes et le système nerveux central. Dans le système nerveux central, on distingue deux éléments jouant un rôle dans le cas de la SEP : les neurones et les cellules gliales.

[1] *https://www.inserm.fr/information-en-sante/dossiers-information/sclerose-en-plaques-sep*

La fonction principale des neurones est de traiter et propager un message nerveux. Les trois principales composantes des neurones sont : le corps cellulaire, les dendrites qui sont des prolongements récepteurs et un prolongement particulier appelé axone (Figure 2). Chacune de ces composantes contribue à la transmission du message nerveux afin que celui-ci se propage dans l'ensemble du système pour transmettre des informations diverses afin d'organiser, contrôler et réguler les fonctions essentielles de notre organisme (motricité, perception, émotions, fonctionnement de certains organes, etc.).

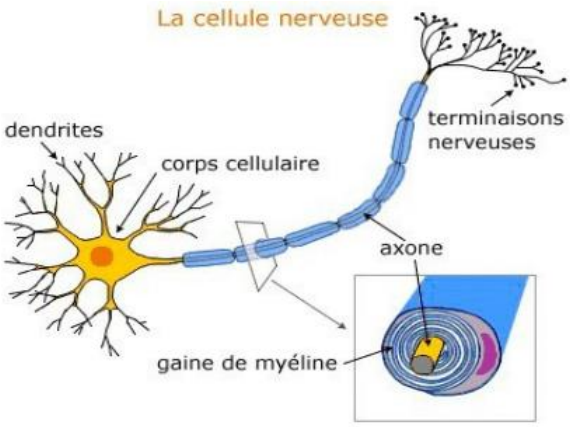


Figure 2 : Description d'un neurone
source : ARSEP, le système nerveux

Les cellules gliales entretiennent, nourrissent et communiquent avec les neurones. En particulier, on s'intéresse aux oligodendrocytes, cellules gliales qui synthétisent la myéline, substance protectrice de l'axone. La myéline est une substance de couleur blanche constituée à 70% de graisses et 30% de protéines. Formant une gaine autour de l'axone, elle a une action isolante, protectrice et nourricière. Ses fonctions la rendent indispensable dans le processus de transmission du message nerveux, d'autant plus qu'elle permet une diffusion rapide et efficace du message. La myéline peut se dégrader et se restaurer. Nous y reviendrons dans les sections suivantes.

Ces deux acteurs de la SEP simplement définis, nous pouvons désormais préciser le mécanisme de la SEP.

1.2 Maladie démyélinisante responsable de la dégénérescence axonale

Dans la SEP, des lymphocytes, sans qu'on ne sache encore pourquoi, pénètrent le système nerveux central. Ces lymphocytes, assurant la défense de système immunitaire en le protégeant des agents pathogènes, sont supposés rester à l'écart du système nerveux central possédant son propre système de défense. Ainsi, lorsqu'ils pénètrent dans le système, zone inconnue, ils attaquent celui-ci provoquant une inflammation et qualifiant ainsi la SEP de maladie inflammatoire auto-immune. Ces attaques détruisent petit à petit les constituants de la myéline. Cette disparition ou destruction est appelée démyélinisation (Figure 3). Ce phénomène

de destruction partielle cause alors des lésions à différents endroits du système nerveux central. L'aspect scléreux des lésions, aussi appelées plaques, donne d'ailleurs son nom à la maladie.

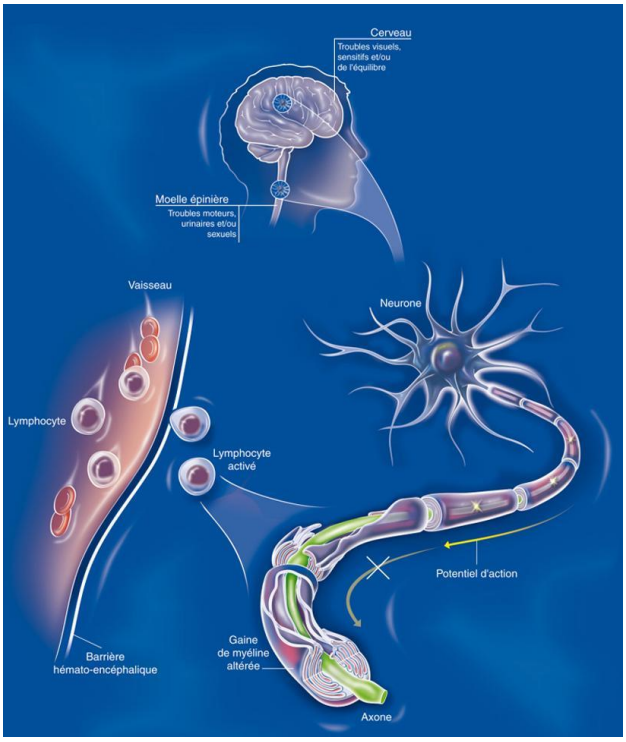


Figure 3 : Attaque de la myéline par les lymphocytes dans la SEP
Source : Dossier d'information de l'INSERM

La démyélinisation est très impactante puisque, la myéline fragilisée ne remplit plus son rôle (Lubetzki and Stankoff, 2017). La protection et l'isolation ne sont plus assurées, le message nerveux est moins vite et/ou moins bien transmis, l'axone est fragilisé et altéré par manque de ressources via la myéline et l'absence de protection. Bien qu'il existe un processus de remyélinisation naturel, ce phénomène de régénération est partiel et souvent insuffisant (Lubetzki and Stankoff, 2017). Ainsi, les modifications du système, bien que d'abord réversibles, deviennent à long terme irréversibles.

Notons que dans certains cas de SEP, l'atteinte peut être d'une nature indépendante de l'inflammation. Dans ce cas, les mécanismes causant l'atteinte axonale progressent plus doucement (Lubetzki and Stankoff, 2017).

Dans tous les cas, cette atteinte axonale conduit à des troubles de diverse nature car plusieurs fonctions sont assurées par le système nerveux central. Ces troubles seront décrits dans la partie 1.4.

1.3 Diagnostic

Le diagnostic de SEP s'appuie sur un faisceau d'arguments cliniques, radiologiques et biologiques. En particulier, il repose sur trois principes : la dissémination spatiale, la dissémination temporelle et l'inflammation[1]. La dissémination spatiale signifie que les manifestations de la maladie doivent être rapportées à au moins deux lésions différentes du système nerveux central. La dissémination temporelle est caractérisée par la nécessité que les évènements neurologiques doivent être au minimum de deux et espacés d'au moins un mois d'intervalle. Enfin, l'inflammation du système nerveux central est établie par l'analyse d'une ponction lombaire.

A partir de ces éléments, des critères ont été établis, d'abord par Poser (Poser *et al.*, 1983) en 1983 puis par Mcdonald (Mcdonald *et al.*, 2001) en 2001. Ce dernier, dans un contexte d'évolution technique, a donné une place plus importante aux critères basés sur l'imagerie. Les critères de Macdonald ont ensuite été révisés en 2005 (Polman *et al.*, 2005), 2010 (Polman *et al.*, 2011), puis 2017 (Thompson *et al.*, 2018), permettant d'identifier la maladie de plus en plus précocement.

1.4 Expression de la maladie

Puisque le système nerveux central assure des fonctions multiples, la SEP est caractérisée par une expression clinique diverse. En effet, l'ambulation, la vision, les activités manuelles, la cognition et le contrôle sphinctérien peuvent être impactés. Il en résulte un grand nombre de symptômes et une difficulté à identifier les premières manifestations cliniques, d'autant plus qu'elles peuvent parfois être discrètes, ne poussant pas à consulter un neurologue et retardant la prise en charge (Papeix, 2017).

Parmi ces diverses manifestations, nous citons quelques exemples (Papeix, 2017). Une atteinte d'ordre visuel peut être caractérisée par une baisse d'acuité visuelle, une altération du champ visuel, un trouble de la vision des contrastes, etc. Des symptômes au niveau des membres, comme une limitation du périmètre de marche ou une boiterie, peuvent être observés. Des patients sont également touchés par des troubles urinaires, comme le besoin absolu

[1] *La Sclérose en plaques, Collège des Enseignants de Neurologie, www.cen-neurologie.fr*

d'uriner. Parmi les fonctions cognitives, on peut notamment observer une atteinte de la vitesse à traiter l'information, de l'attention, de la mémoire à long terme ou du raisonnement conceptuel. La fatigue est également un symptôme caractéristique de la SEP.

2. Phénotypes et activité de la SEP

2.1 Les formes de SEP

Deux évènements peuvent être distingués en SEP : la poussée et la progression[1]. La poussée est définie comme l'apparition de nouveaux symptômes, la réapparition ou l'aggravation de symptômes déjà existants (Thompson *et al.*, 2018). Les poussées durent au minimum 24 heures. Après une poussée, le patient peut récupérer entièrement ou seulement partiellement. Pour être distinctes, deux poussées doivent débuter à plus de 30 jours d'intervalle. La progression est définie comme l'apparition continue des symptômes au long cours.

A partir de ces deux évènements, deux formes de SEP peuvent être distinguées (Lublin *et al.*, 2014) : la forme avec poussées ou rémittente et la forme progressive, soit d'emblée ou secondaire à un début rémittent appelée forme secondairement progressive (Figure 4).

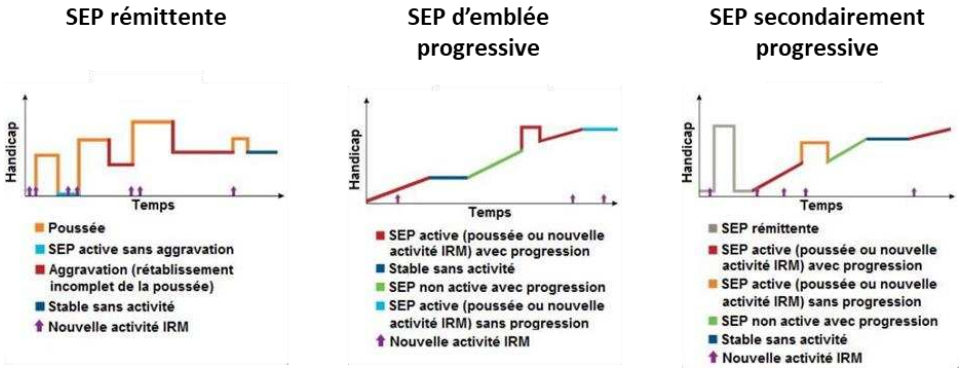


Figure 4 : Formes cliniques de SEP
Source : Roux, 2018 adapté de Lublin et al, 2014

La forme avec poussées ou rémittente est caractérisée par l'apparition de poussées successives, pouvant laisser des séquelles ou non, sans progression de handicap entre les poussées (Lublin *et al.*, 2014). Cette forme est la forme la plus répandue puisqu'elle représente

[1] *La Sclérose en plaques, Collège des Enseignants de Neurologie, www.cen-neurologie.fr*

85% à 90% des SEP. Elle débute chez le jeune adulte et prédomine chez la femme avec un sex-ratio de 3 femmes pour 1 homme[1].

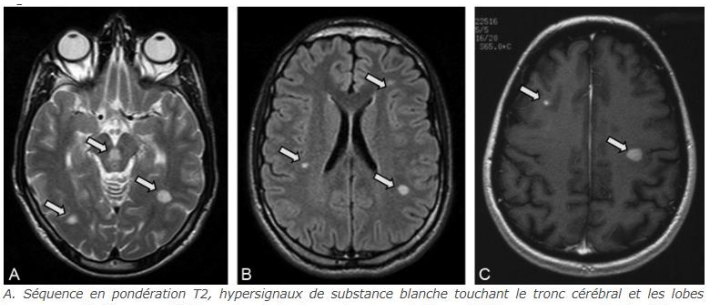
La forme d'emblée progressive est marquée par des symptômes insidieux, s'aggravant et conduisant ainsi à l'apparition progressive d'un handicap sans poussée surajoutée (Lublin *et al.*, 2014). Elle représente 10 à 15 % des SEP. Cette forme de SEP débute plus tard que la forme avec poussées et touche, dans une même mesure, les hommes et les femmes.

La forme secondairement progressive est caractérisée par une phase initiale de poussées. Lors d'une seconde phase de la maladie, on observe une progression du handicap, parfois marquée par des poussées surajoutées suivies de rémissions minimes, puis une phase de plateau (Lublin *et al.*, 2014).

Pour évaluer l'activité de la SEP, l'imagerie à résonnance magnétique (IRM) est utilisée pour décrire les lésions et le handicap est mesuré. Dans la SEP évoluant par poussée, la description des poussées, que ce soit en termes de fréquence, de sévérité ou de phénotype est également utilisée.

2.2 Activité IRM

L'IRM permet de localiser les lésions responsables des symptômes. Grâce à l'IRM, on mesure l'ampleur de la maladie par l'évaluation de la charge ou du volume lésionnel (Granziera *et al.*, 2017). Les lésions sont caractérisées par leur localisation, leur taille et par l'intensité de leur signal. Elles peuvent parfois être mises en évidence par l'utilisation d'un produit de contraste, comme le gadolinium, qui permettra d'identifier les lésions récentes (Figure 5). L'IRM fournit également des éléments concernant la perte ou la réparation tissulaire.



A. Séquence en pondération T2, hypersignaux de substance blanche touchant le tronc cérébral et les lobes temporaux. B. Séquence FLAIR montrant des hypersignaux nodulaires. C. Séquence en pondération T1 avec injection de gadolinium montrant deux lésions rehaussées par le produit de contraste.

Figure 5 : Imagerie cérébrale dans la SEP
Source : Collège des enseignants de Neurologie

[1] *https://www.inserm.fr/information-en-sante/dossiers-information/sclerose-en-plaques-sep*

Des IRM cérébrales ou encéphaliques et médullaires peuvent être pratiquées. Elles fournissent des informations nécessaires, voire primordiales, à la prise en charge d'un patient que ce soit au moment du diagnostic ou au cours de son suivi notamment pour évaluer l'évolution de l'activité de la maladie et la réponse aux traitements (Tintore *et al.*, 2015; Papeix, 2017) (Papeix, 2017).

2.3 Activité clinique : le handicap

Les évolutions cliniques spécifiques de la SEP ont imposé un langage particulier pour décrire le handicap. L'échelle de Kurtzke (Kurtzke, 1983) est communément utilisée par les neurologues pour coter le handicap. Avec cette échelle, appelée EDSS pour « Expanded Disability Status Scale », le niveau de handicap est ainsi résumé par un score global synthétisant l'évaluation individuelle de divers systèmes fonctionnels. Ce score global évalue l'importance du handicap de 0 (examen neurologique normal) à 10 (décès lié à la SEP).

Les paramètres fonctionnels évalués sont les fonctions pyramidales (i.e. la motricité), cérébelleuses (i.e. équilibre et coordination des mouvements), sensorielles ou sensitives, cérébrales ou mentales, sphinctériennes et visuelles. Le détail de leur évaluation est disponible en annexe A. L'échelle de cotation globale y est également présentée. Selon le niveau du handicap, on notera que son évaluation est dominée par certains paramètres fonctionnels (Edan, 2017). Elle est d'ailleurs très influencée par le périmètre de marche à partir du score 5. Parmi les scores marquants et fréquemment utilisés dans les études, on peut citer le niveau d'EDSS 3 qui correspond à une atteinte minime de 3 paramètres fonctionnels (3 fonctions à 2 points, les autres à 0 ou 1 ou bien une fonction à 3 points les autres à 0 ou 1). Ce niveau est décrit comme le début d'un handicap sans conséquence sur la vie quotidienne (Leray *et al.*, 2010). A ce niveau de handicap, le patient n'est pas limité dans son périmètre de marche mais ne peut en revanche plus courir. Toujours à titre d'exemple, le score EDSS 6 correspond à la nécessité d'utiliser une aide unilatérale, de façon quasiment constante, pour parcourir 100 m sans s'arrêter ou le besoin d'une aide bilatérale pour parcourir plus de 100 m.

Le handicap est ainsi mesuré de manière longitudinale au cours du suivi du patient. Il peut être évalué de manière ponctuelle en notant, le cas échéant, la proximité d'une poussée qui peut accentuer les troubles et donc surévaluer le handicap réel du patient.

Des scores de handicap confirmés (i.e. mesurés à distance les uns des autres) et maintenus (i.e. persistants jusqu'aux dernières nouvelles) fourniront une évaluation de l'accumulation du handicap ou du handicap résiduel. Notamment, un score de handicap confirmé à 6 mois et maintenu jusqu'aux dernières nouvelles est appelé irréversible (Confavreux *et al.*, 2003).

Cette échelle présente l'avantage de fournir une évaluation chiffrée du handicap permettant d'étudier la progression du handicap d'un patient mais également de comparer les patients entre eux.

3. Etiologie de la SEP

3.1 Epidémiologie de la SEP

Depuis les années 90, le nombre de cas de SEP en France n'a cessé d'évoluer. La première étude, menée à partir d'un questionnaire que les personnes étaient invitées à remplir suite à un appel donné à la télévision, a conduit à établir une première estimation de prévalence entre 30 et 40 pour 100 000 (Lhermitte *et al.*, 1988). Finalement, l'étude la plus récente, à partir des données de la Caisse Nationale d'Assurance Maladie des Travailleurs Salariés, a utilisé un algorithme multicritère basé sur le statut d'affection longue durée et les remboursements d'un traitement ou d'un acte lié à la SEP, comprenant les hospitalisations. Cette étude a conduit à une estimation de 151.2 pour 100 000 correspondant à 99 123 cas de SEP en France (Foulon *et al.*, 2017). L'évolution de la prévalence au cours du temps peut certainement être le résultat d'un meilleur comptage mais également d'une mortalité moins importante. On ne peut pour autant exclure une effective augmentation de l'incidence de la SEP (Browne *et al.*, 2014). La dernière étude d'incidence en France, datant de 2012, a permis d'estimer l'incidence nationale entre 2001 et 2007 à 7.6 pour 100 000 (Fromont *et al.*, 2012).

Au niveau européen, la prévalence s'élève à 700 000 cas de SEP. A l'échelle du monde, elle est estimée à 2.5 millions (Browne *et al.*, 2014). Sa répartition est marquée par un gradient augmentant avec l'éloignement de l'équateur[1].

[1] *https://www.inserm.fr/information-en-sante/dossiers-information/sclerose-en-plaques-sep*

L'étude de la mortalité a conduit à montrer que les patients atteints de SEP avaient un risque similaire de décès pendant 20 ans suivant le début de la maladie en comparaison à la population générale. Cette même étude a montré que l'espérance de vie serait réduite de 7 ans chez les patients atteints de SEP, toujours en comparaison à la population générale (Leray *et al.*, 2015). Dans 50% des cas, le décès était lié à une complication de la maladie. Les autres causes sont principalement des maladies cardiovasculaires, des cancers et des suicides.

3.2 Génétique et environnement, quels sont les facteurs associés à la SEP ?

La SEP est une maladie multifactorielle. Sa survenue est inexpliquée mais la génétique et l'environnement sont souvent évoqués comme des causes potentielles (McKay *et al.*, 2017).

Sur le plan génétique, certains gènes semblent être liés à la SEP en particulier présents dans la région HLA (antigène des leucocytes humains), en charge de la gestion de l'immunité (Brassat, 2017). De plus, des études ont montré un risque plus marqué d'avoir une SEP dans les familles dont un des membres est atteint. On note tout de même que le risque d'avoir une SEP chez les jumeaux monozygotes, lorsque l'un est atteint, est de 30%. Cet argument a tendance à contester l'argument de l'hérédité de la SEP.

Outre la prédisposition génétique qu'on ne peut exclure, les disparités géographiques de prévalence et de risque laissent penser que des facteurs environnementaux peuvent également être associés (Pantazou *et al.*, 2017). Les habitudes de vie, l'ensoleillement, et certaines maladies sont évoqués comme potentiels facteurs de risque d'avoir une SEP. En particulier, on retrouve la consommation de tabac active ou passive, la vitamine D, l'obésité pendant l'enfance et l'adolescence. On notera également que 100% des patients atteints de SEP présentent une séropositivité au virus Epstein Barr.

4. Prise en charge de la SEP

Les multiples expressions de la SEP conduisent à une prise en charge pluridisciplinaire des patients. Bien que plusieurs professionnels interviennent dans la prise en charge du patient, l'un des principaux acteurs est le neurologue, expert de cette pathologie neuro- dégénérative (Papeix, 2017). Son travail est aussi complété par celui du médecin généraliste. Ils interviennent

dès le début de la maladie. Le médecin généraliste observe souvent les premiers symptômes, le neurologue formule le diagnostic. Au cours du suivi, le médecin généraliste aura en charge le suivi régulier et rapproché du patient, notamment vis-à-vis des symptômes liés à la SEP. Le neurologue, quant à lui, sera consulté moins souvent, une fois par an étant recommandé (Lublin *et al.*, 2014). Il est le seul à pouvoir prescrire les traitements de fond de la SEP que nous présenterons dans la section suivante.

Le neurologue peut exercer en cabinet libéral ou dans le secteur hospitalier. Il existe des centres experts, aussi appelés CRC SEP pour Centre de Ressources et de Compétences SEP, dans les Centres Hospitalo-Universitaires. Les centres experts proposent une prise en charge d'abord médicale, mais également médico-sociale. Ils sont aussi reconnus pour leurs travaux de recherche et d'innovation scientifiques (Derache *et al.*, 2018).

Certains patients sont suivis exclusivement ou régulièrement dans ces centres. En revanche, les patients peuvent aussi venir consulter le centre expert ponctuellement. En effet, de par l'expertise du personnel de centre, les patients ou les neurologues non experts de la SEP sont amenés à consulter le centre pour un avis ponctuel, pour une confirmation de diagnostic ou un avis sur l'instauration d'un traitement de fond par exemple. Dans ce cas, les neurologues des centres experts recevront le patient et la suite de la prise en charge pourra être poursuivie à l'extérieur du centre. Le niveau d'expertise aux centres et leur rattachement à des centres hospitaliers universitaires, peuvent marquer la patientèle de ces services. De fait, la consultation d'un expert peut parfois témoigner d'une forte activité ou d'un avancement de maladie. Pour autant, ce n'est pas le seul profil de SEP dont sont atteintes les personnes consultant le centre expert.

Comme nous l'évoquions précédemment, d'autres professionnels de santé peuvent être impliqués dans la prise en charge des patients. Ils interviennent dans le cadre des traitements des symptômes liés à la SEP et pour accompagner les patients à vivre avec leur pathologie.

On citera notamment les médecins de spécialité physique et réadaptation (MPR), les kinésithérapeutes, les ophtalmologistes, les orthophonistes, les urologues, les psychiatres, psychologues ou neuropsychologues.

Enfin, les infirmiers, acteurs principaux de l'éducation thérapeutique, jouent un rôle majeur dans la prise en charge des patients. L'éducation thérapeutique du patient vise à aider les patients à gérer au mieux leur vie avec une maladie chronique. Ils peuvent, par conséquent, intervenir à différent moments. Lors du diagnostic, ils pourront apporter des explications

complémentaires à celles déjà fournies lors de la consultation avec le neurologue. Ils aideront aussi en informant les patients sur la reconnaissance des poussées et la gestion de la maladie au quotidien (Higue-van Steenbrugghe, 2016). Egalement, ils accompagneront les patients dans la prise du traitement que ce soit dans l'explication des différents traitements ou pour l'administration.

Dans cette première partie, nous avons évoqué, parfois simplement, les données qui nous apparaissaient importantes pour la suite de ce travail. La SEP est une maladie complexe dont la connaissance est en constante évolution. Qu'il s'agisse de connaissances fondamentales, cliniques ou techniques, elles modifient la prise en charge de la maladie, impactant la pratique des professionnels associés et marquant la vie des patients. Parmi ces évolutions, l'une des plus marquantes de la prise en charge concerne le développement de traitements spécifiques à la SEP. Au moment de ce travail, les traitements disponibles dans la SEP étaient adaptés seulement aux formes de SEP à début rémittent (une poussée ou plus). Dans la suite, nous nous y réfèrerons comme SEP rémittente. Pour rappel, cette forme est la plus répandue.

B. Les traitements de fond de la SEP rémittente

Le premier traitement de fond spécifique à la SEP rémittente est disponible depuis le milieu des années 90 en France et dans de nombreux pays du monde. Depuis, l'arsenal thérapeutique a évolué largement avec désormais une quinzaine de traitements de fond disponibles. La connaissance de la balance bénéfice/risque se clarifie et l'intérêt de traiter un patient atteint de SEP apparait essentiel. Ces avancements ont conduit à faire évoluer les stratégies de prise en charge des patients.

1. Généralités sur les traitements de la SEP rémittente

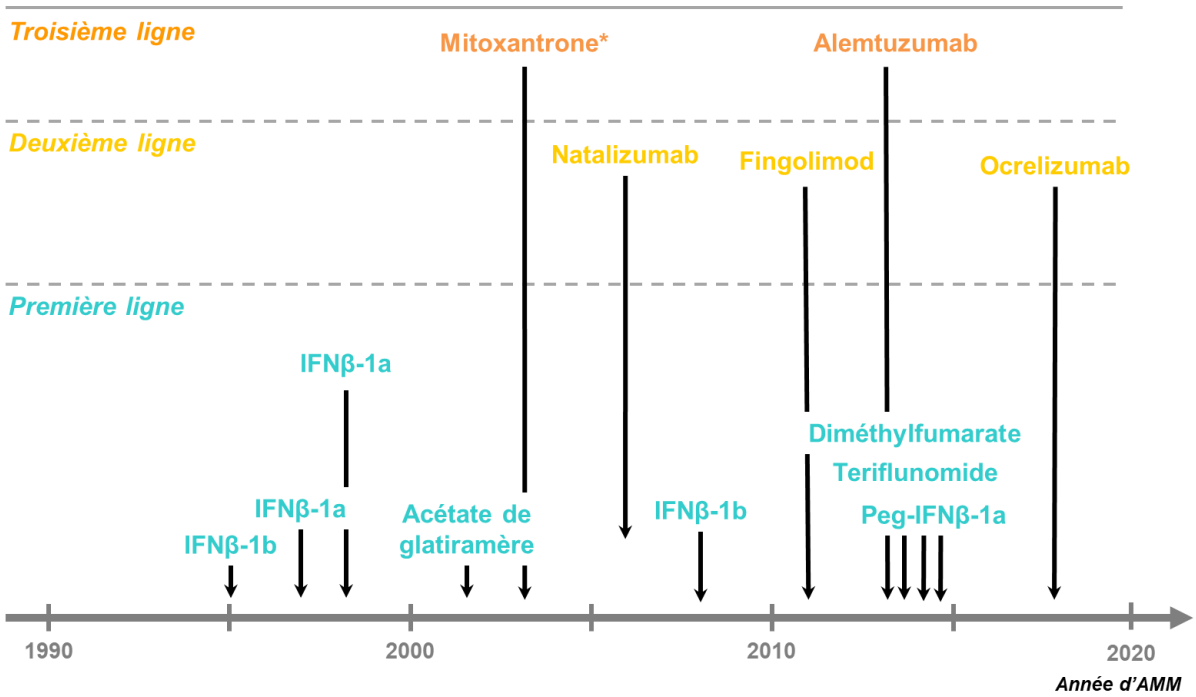
Il n'existe pas de traitement curatif de la SEP. Pendant plusieurs années, seuls les symptômes ont pu être traités. Ces traitements symptomatiques sont nombreux et variés du fait de l'expression multiple de la maladie. On peut par exemple citer la rééducation pour les troubles de la marche, mais également certains traitements médicamenteux comme l'amantadine ou le modafinil contre la fatigue (Clavelou *et al.*, 2017). Parmi les traitements proposés à un patient atteint de SEP, on retrouve également les corticoïdes proposés en cas de poussée[1]. Ils ne sont pas prescrits systématiquement mais ils permettent d'accélérer la récupération de la poussée. Ces traitements ne présentent pas d'effet sur la prévention du risque de survenue de poussée.

Après plusieurs années d'utilisation de divers traitements n'ayant pas démontré leur efficacité, l'histoire des thérapeutiques dans la SEP est marquée par le développement, au milieu des années 90, du premier traitement de fond spécifique, l'interféron-β (The IFNB Multiple Sclerosis Study Group, 1993). Les traitements de fond agissent sur la composante inflammatoire de la maladie. En interférant avec le système immunitaire, ils ont pour but de réduire la fréquence des poussées et de ralentir la progression du handicap. Il peut en être distingué deux familles : les immunomodulateurs et les immunosuppresseurs. Cette distinction n'est pas toujours évidente. Dans la thèse, nous évoquerons plutôt les traitements selon leur ligne d'intention thérapeutique, les regroupant en deux familles : traitements de première ligne

[1] *La Sclérose en plaques, Collège des Enseignants de Neurologie, www.cen-neurologie.fr*

et les traitements de deux et troisième lignes. Dans le groupe des deux à troisième lignes, certains traitements ont des positionnements discutés. Notons que la SEP peut également être traitée en ayant recours à la transplantation de cellules souches hématopoïétiques autologues (AHSCT). Ce type de recours apparait dans des cas particuliers où les traitements médicamenteux ne fournissent pas satisfaction au regard de l'activité de la SEP. Dans cette thèse, nous n'aborderons pas cette option et nous nous concentrerons uniquement sur les traitements médicamenteux.

Dans chacune des familles, plusieurs substances actives sont disponibles. On compte en 2019, 13 traitements disponibles. La représentation de l'historique de la disponibilité de ces traitements de fond témoigne de la diversification de l'arsenal thérapeutique au cours du temps (Figure 6). Les données de prescriptions issues de données médico-administratives (SNDS) portant sur 112 000 patients présentées en Figure 7 indiquent la place non négligeable qu'empruntent les nouvelles thérapeutiques dès leur disponibilité (Roux, 2018).



Traitements « hors AMM »: Azathioprine, cyclophosphamide, méthotrexate, mycophénolate mofétil, rituximab, cladribine

IFNβ : Interféron-β. AMM : Autorisation de Mise sur le Marché, en France
**mitoxantrone: date de disponibilité du Elsep®, la Novantrone®, spécialité de mitoxantrone pouvait être utilisée avant cette date*

Figure 6 : Disponibilité des traitements de fond de la SEP rémittente au cours du temps

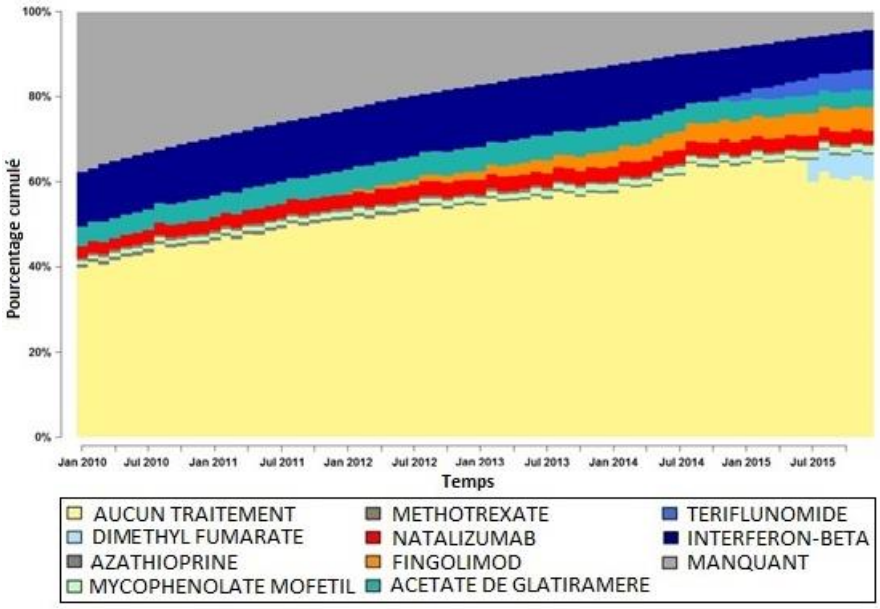


Figure 7 : Utilisation des traitements de fond de la SEP rémittente entre 2010 et 2015
Source : thèse de Jonathan Roux, 2018

2. Zoom sur les traitements de première ligne

Le premier traitement spécifique SEP est l'interféron béta-1b (Betaferon® et Extavia®)), suivi du bêta-1a (Avonex® et Rebif®). Ce n'est que depuis peu qu'une forme pégylé est disponible (Plegridy®). Les traitements interférons bêta-1a et 1b sont des immunomodulateurs. Ils agissent sur la diminution du nombre de lymphocytes et de leur activité (Vermersch and Schluep, 2017). Ils sont présentés sous forme injectable et s'administrent en sous-cutanée ou intramusculaire à l'aide de seringue ou stylo pré-remplis. Selon la spécialité, le médicament doit être pris une à trois fois par semaine ou tous les deux jours. La forme pégylé est un interféron-β modifié présentant une action prolongée. Sa fréquence d'injections est bimensuelle après une phase de mise en place avec des injections rapprochées. Les interférons bêta -1b ont démontré leur efficacité sur le plan clinique et IRM (The IFNB Multiple Sclerosis Study Group, 1993). Une réduction de 32% de la fréquence des poussées a été observée en comparaison au placebo. L'activité IRM était aussi diminuée dans le groupe traité par interféron bêta-1b. En revanche, le risque de progression du handicap n'était pas significativement modifié par le traitement. Les essais thérapeutiques conduits sur les interférons bêta-1a ont corroboré ces résultats en ajoutant une réduction du risque de handicap (Jacobs *et al.*, 1996;

PRISMS, 1998). En ce qui concerne la tolérance, les interférons-β présentent un profil plutôt satisfaisant avec des effets secondaires jugés légers ou d'intensité modérée (Vermersch and Schluep, 2017). Les effets secondaires parfois observés peuvent être des syndromes pseudo-grippaux, des ecchymoses au niveau des zones d'injections et des dépressions.

Suite à l'interféron-β et présentant le même mode d'action, l'acétate de glatiramère (Copaxone®) est arrivé sur le marché avec une première spécialité disponible en janvier 2002[1].

A l'instar de l'interféron-β, l'acétate de glatiramère est un traitement immunomodulateur qui agit sur l'orientation des cellules immunitaires (Vermersch and Schluep, 2017). Ce traitement est présenté sous forme injectable également. Il s'administre quotidiennement en sous-cutanée à l'aide d'une seringue pré-remplie. L'acétate de glatiramère a démontré son efficacité au travers d'essais thérapeutiques randomisés, notamment sur la réduction de la fréquence des poussées diminuée de 29% par rapport au groupe placebo (Johnson *et al.*, 1995; Comi *et al.*, 2001*b*). La différence sur le risque de progression du handicap n'était pas significative. Plus tard, une démonstration de son bénéfice sur l'activité IRM a aussi été faite (Comi *et al.*, 2001*b*). Ce traitement présente un bon profil de tolérance avec peu de toxicité et des effets indésirables rares, décrits comme peu sévères (Vermersch and Schluep, 2017).

Ce n'est que plusieurs années plus tard, en 2014, que l'offre de traitements de première ligne s'est complétée avec l'arrivée de deux traitements : le tériflunomide (Aubagio®) et le diméthylfumarate (Tecfidera®). Leur particularité est leur voie d'administration puisqu'ils s'administrent par voie orale. Le tériflunomide a démontré son efficacité dans les essais cliniques (O'Connor *et al.*, 2011; Confavreux *et al.*, 2014). Cette substance présente un profil de tolérance généralement bon mais est rigoureusement contre-indiquée pendant la grossesse. (Vermersch and Schluep, 2017). Le diméthylfumarate a démontré son efficacité sur la réduction des poussées, ainsi que sur le risque de progression et l'activité IRM (Fox *et al.*, 2012; Gold *et al.*, 2012). Il présente également un profil de tolérance satisfaisant marqué tout de même par des troubles digestifs et des bouffées de chaleur (Vermersch and Schluep, 2017).

[1] *Rapport public d'évaluation de la Copaxone® par l'ANSM, disponible sur www.ansm.sante.fr*

3. Zoom sur les traitements de deuxième et troisième ligne

La mitoxantrone (Elsep®), disponible depuis fin 2003, est un traitement puissant dit de haute efficacité. C'est le premier traitement de deuxième à troisième intention disponible dans la SEP. Elle est administrée par perfusion une fois par mois sur une durée limitée à quelques mois. Les essais thérapeutiques ont démontré son efficacité sur la réduction de la fréquence des poussées, l'évolution du handicap et l'activité IRM (Edan *et al.*, 1997; Hartung *et al.*, 2002). Ce traitement présente un profil de tolérance assez risqué. Il est en effet associé à des effets indésirables rares mais graves de type maladies cardiovasculaires, problèmes endocriniens, pouvant conduire à des aménorrhées, parfois des infertilités, et hématologiques, pouvant conduire dans certains cas à des leucémies (Marriott *et al.*, 2010; Le Page *et al.*, 2011). Notons qu'avant la disponibilité du Elsep®, le Novantrone®, spécialité de mitoxantrone non spécifique, était déjà utilisé dans la SEP.

Parmi les autres immunosuppresseurs, nous détaillerons les données concernant le natalizumab, disponible à partir de 2007 et le fingolimod, disponible à partir de 2011. Le natalizumab (Tysabri®) est un traitement administré par perfusion à l'hôpital tous les 28 jours. Il a montré son efficacité sur la réduction du nombre de poussées, le handicap et l'activité IRM en comparaison au placebo ainsi qu'aux traitements de première ligne (Polman *et al.*, 2006; Rudick *et al.*, 2006). Il présente des effets indésirables, notamment, un haut risque de leucoencéphalite multifocale progressive (LEMP) (Kleinschmidt-DeMasters, B.K. Tyler, 2005; Langer-Gould *et al.*, 2005). La LEMP est une pathologie traduite par des troubles neurologiques divers, graves et localisés évoluant vers un état végétatif, voire la mort en quelques mois. Le risque de LEMP a été documenté et on peut citer désormais trois facteurs de risques associés au développement d'une LEMP : (1) Un traitement par natalizumab de plus 18 à 24 mois, (2) un index anticorps contre le virus John Cunningham (JC) positif, (3) une exposition à un immunosuppresseur précédent l'exposition au natalizumab (Vermersch and Schluep, 2017). En particulier, depuis 2014, le risque de LEMP a été quantifié chez les patients présentant une séropositivité au virus JC (Annexe B). Ce test est donc réalisé avant d'exposer les patients au natalizumab. Selon le statut vis-à-vis du virus et la stratification du risque, l'exposition au natalizumab peut être contrainte.

Disponible depuis 2011, le fingolimod (Gilenya®), est le premier traitement oral en prise quotidienne de la SEP. A l'instar du natalizumab, son efficacité a été démontrée sur les poussées, le handicap et l'activité IRM (Cohen *et al.*, 2010; Kappos *et al.*, 2010). Bien que le fingolimod puisse être associé à quelques effets indésirables, le risque de LEMP est faible (Vargas and Tyor, 2017).

Plusieurs autres traitements tels que les anti-CD20 ont également récemment alimenté l'arsenal thérapeutique des traitements spécifiques SEP, comme exposé sur la Figure 6. Récents et/ou peu utilisés en France, ces derniers ne seront pas détaillés dans ce travail.

4. Stratégies thérapeutiques

Avec ces différents traitements, il se dessine diverses stratégies d'utilisation même s'il n'existe pas d'algorithme thérapeutique qui fasse consensus à ce jour (Vermersch and Schluep, 2017). On peut tout de même distinguer trois stratégies thérapeutiques proposées au regard de l'activité clinique et IRM : la stratégie d'escalade, la stratégie de traitement intensif et l'induction. La Figure 8 propose une schématisation de ces stratégies adaptée d'une précédente publication (Vermersch et al., 2016).

Le principe de l'escalade est de proposer d'abord un traitement présentant une certaine efficacité et des risques modérés ou quasi nuls (Freedman et al., 2018). L'interféron-β, l'acétate de glatiramère, le tériflunomide ou le diméthyl fumarate sont alors proposés dans ce cas. Le choix du traitement au sein des traitements de première intention dépend surtout de la voie d'administration et des risques d'effets secondaires. Ce choix est souvent fait en concertation avec le patient.

Ensuite au regard de l'évolution clinique et IRM, des changements pourront être proposés. Si le neurologue propose un autre traitement de première ligne, on parlera de changement ou « switch » horizontal. Cela se produit généralement pour des raisons de tolérance. En cas d'évolution témoignant d'une activité non contrôlée, le neurologue envisagera alors un changement vertical vers un traitement de deuxième, voire de troisième ligne.

Lorsque la maladie est marquée par une activité assez forte sur le plan clinique et/ou IRM, il pourra être proposé de commencer d'emblée par un traitement de deuxième ligne. Cette stratégie intensive est marquée par l'utilisation d'un traitement fort sur une période potentiellement longue.

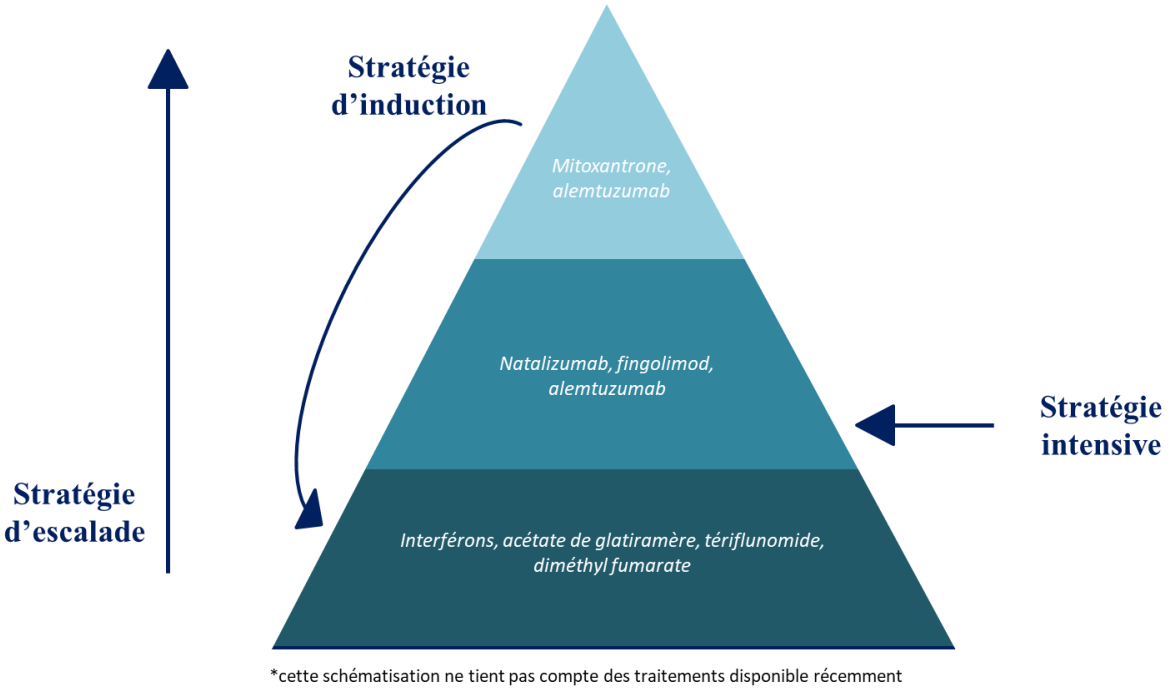


Figure 8 : Proposition d'algorithme thérapeutique dans la prise en charge médicamenteuse
Source : adaptée de Vermersch et al., 2016

Finalement, dans des cas particuliers de maladie très sévère (Edan and Le Page, 2013), il peut être discuté de mettre en place une stratégie d'induction (Freedman et al., 2018). Dès le début, un traitement de troisième ligne, comme la mitoxantrone ou l'alemtuzumab, sera administré sur un court terme dans le but de stopper l'inflammation. Suite à ce traitement qualifié d'agressif, la SEP sera mieux contrôlée et un relais par un traitement de première ligne pourra être envisagé.

Qu'on opte pour une stratégie d'escalade, intensive ou d'induction, il existe un consensus sur le moment d'intégrer les thérapeutiques dans la prise en charge d'un patient, à savoir au plus tôt. En effet, plusieurs arguments incitent à agir tôt dans le développement de la maladie. Des travaux d'épidémiologie (Leray et al., 2010), d'immunologie (McDonald et al., 1992) et les extensions d'essais thérapeutiques (Comi et al., 2013; Kappos et al., 2016) ont montré l'intérêt d'un traitement précoce. Certaines recommandations ont d'ailleurs émergé récemment (Montalban et al., 2018).

Les résultats des essais thérapeutiques montrent l'intérêt de ces différents traitements dans la prise en charge des patients atteints de SEP, en particulier en raison de leur efficacité sur la réduction de la fréquence des poussées. Cependant, les résultats obtenus en essais cliniques ne peuvent être généralisés à l'ensemble de la population. Ainsi, les études observationnelles fournissent des résultats à plus long terme et dans des conditions de prise en charge de vie réelle. Dans la section suivante, nous revenons sur les spécificités de l'évaluation de traitement en vie réelle, en faisant le parallèle avec l'évaluation en conditions contrôlées.

C. Efficacité d'un traitement en vie réelle

La pharmaco-épidémiologie peut être définie comme l'étude de l'utilisation et des effets des médicaments sur un nombre important de personnes, couramment des cohortes (Strom, 2012). A la lecture de l'intitulé de cette discipline, on fait rapidement un parallèle avec deux autres disciplines : la pharmacologie et l'épidémiologie.

La pharmacologie regroupe l'ensemble des connaissances relatives à la découverte, à l'étude et à l'utilisation des médicaments chez l'homme[1]. La pharmacologie vise à décrire différentes caractéristiques d'un médicament : sa composition, ses propriétés, ses interactions, sa toxicité et finalement ses effets recherchés ou indésirables. Elle contribue à comprendre comment fonctionne un médicament et comment il devrait être utilisé, fournissant une connaissance essentielle dans le développement de nouveaux médicaments (Barett and Zuppa, 2012).

L'épidémiologie peut être définie comme l'étude de la distribution et des déterminants des maladies des populations. L'approche peut être descriptive ou analytique. En pharmaco-épidémiologie, les méthodes issues de l'épidémiologie seront appliquées à un contexte de pharmacologie à des fins de surveillance des médicaments, d'analyse des facteurs associés à leur utilisation, de quantification de leurs effets positifs comme négatifs, et ce, après leur mise sur le marché (Strom, 2012).

[1] *https://pharmacomedicale.org/pharmacologie*

1. Processus de développement d'un médicament

Le processus du développement d'un médicament débute par un besoin médical. Pour y répondre, plusieurs molécules ciblées peuvent être testées sur les animaux dans le cadre des essais pré-cliniques. Suivant cette phase, la ou les substances sont évalué.e.s dans le cadre d'un essai clinique[1]. Un essai clinique est défini par l'Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM) comme une recherche biomédicale organisée et pratiquée sur l'homme en vue du développement des connaissances biologiques ou médicales[2]. L'objectif de l'essai clinique est d'évaluer le médicament quant à sa sécurité et son efficacité. Cette évaluation peut se faire chez des volontaires sains ou malades. L'essai clinique se déroule en quatre phases d'évaluation[3].

La phase I constitue les premiers tests chez l'homme, elle permet notamment d'évaluer la toxicité de la substance.

La phase II correspond à l'étude d'efficacité pharmacologique, aussi appelée étude pilote. Les objectifs de cette phase sont multiples : (1) évaluer la sécurité d'emploi à court terme, (2) définir la dose minimale efficace, (3) estimer les relations dose-effet et concentration-effet, (4) rechercher le schéma posologique utilisé en phase III[3]. Après l'étude pilote, il peut arriver qu'un médicament soit disponible pour certains patients. Ceci n'arrive que dans des conditions exceptionnelles et sous réserve d'autorisation par l'ANSM. Cette autorisation est nommée Autorisation Temporaire d'Utilisation et peut être délivrée à titre individuel, on parle d'ATU nominative, ou pour un ensemble de patients, ATU de cohorte[4].

La phase III constitue la phase d'évaluation de l'intérêt thérapeutique, on parle d'étude pivot. Elle est conduite sur un nombre plus important de personnes malades que la phase II. Elle est aussi souvent plus longue. Dans cette phase, on évalue l'efficacité, la sécurité d'emploi et la balance bénéfice/risque, on compare la substance au médicament de référence et on définit l'indication du traitement. Cette phase est multicentrique et nécessite une méthode standardisée d'évaluation sur laquelle nous reviendrons dans la partie suivante. Dans les suites de cette

---

[1] *www.inserm.fr/information-en-sante/dossiers-information/medicament-developpement*
[2] *www.ansm.sante.fr/Activites/Essais-cliniques/Repertoires-des-essais-cliniques-de-medicaments/Repertoires-des-essais-cliniques-de-medicaments/Glossaire-relatif-aux-essais-cliniques*
[3] *www.pharmacomedicale.org/pharmacologie/developpement-et-suivi-des-medicaments/28-essais-cliniques-chez-l-homme*
[4] *www.ansm.sante.fr/Activites/Autorisations-temporaires-d-utilisation-ATU/Qu-est-ce-qu-une-autorisation-temporaire-d-utilisation*

évaluation et au regard des résultats, cela pourra conduire à une autorisation de mise sur le marché (AMM) qui est délivrée, à l'échelle nationale, par l'agence nationale de sécurité du médicament et des produits de santé (ANSM) et à l'échelle européenne, par l'European Medical Agency[1]. La Haute Autorité de Santé (HAS) est également consultée via sa commission de transparence pour évaluer le service médical rendu et/ou l'amélioration du service médical rendu[2]. Le prix et les modalités seront alors fixés.

Finalement, la phase IV, post-AMM, concerne le suivi de l'utilisation d'un médicament, notamment dans l'évaluation des effets indésirables rares qui auraient pu ne pas être observés en conditions contrôlées. Elle permet également de mettre en évidence de potentielles complications tardives ou des mésusages[3].

Dans la suite, on s'intéresse en particulier à ces deux dernières phases qui incluent, notamment, les études d'efficacité. L'une des évaluations est réalisée en conditions contrôlées (en anglais : « efficacy »), l'autre est conduite en vie réelle (en anglais : « effectivenness »). Si dans le premier cas, le cadre est strict et défini, le deuxième cadre pose des questions méthodologiques complexes.

2. Evaluation de l'effet d'un médicament

Pour évaluer l'effet d'un médicament, on étudie l'association entre l'exposition au médicament et la survenue d'un évènement, comme un changement d'état par exemple.

Dans l'estimation d'une association entre deux éléments, il est primordial de tenir compte des facteurs de confusion. Ces facteurs sont des variables qui, si elles ne sont pas prises en compte, peuvent créer une association erronée, ou, au contraire, masquer une association existante ou encore la sur ou sous-estimer. On peut par exemple citer l'exemple de l'association entre les doigts jaunis et le cancer du poumon qui, de toutes évidences, est « confondue » par la consommation de tabac. L'association que l'on cherche à mesurer dans notre contexte est une association causale. On cherche à établir si l'exposition au traitement est la cause de l'état que l'on observe. Cette association causale n'est pas triviale et souvent difficile à démontrer. Une situation idéale serait de pouvoir observer, dans les mêmes conditions, la même personne

[1] *www.ema.europa.eu/en*
[2] *www.has-sante.fr /le-service-medical-rendu-smr-et-l-amelioration-du-service-medical-rendu-asmr*
[3] *www.inserm.fr/recherche-inserm/recherche-clinique/essais-cliniques-recherches-interventionnelles-portant-sur-produit-sante*

étant exposée et non exposée. Evidemment, cette situation est irréaliste. Pour pallier à cela, l'approche populationnelle pourra aider, reposant sur la constitution de deux groupes de personnes, que l'on construit comparables. Si la construction de ces deux groupes est réussie, leur comparaison fournira une mesure d'efficacité du médicament, c'est à dire l'effet causal que l'exposition au médicament a eu sur l'évènement d'intérêt.

Comme évoqué précédemment, on distingue deux contextes d'évaluation. L'évaluation en conditions contrôlées fournit une information sur la capacité d'un médicament à apporter l'effet souhaité (Strom *et al.*, 2012). Ce type d'évaluation fournit un haut niveau de preuve. Nous y reviendrons ci-après. En revanche, pour des raisons éthiques et économiques, cette évaluation ne peut être réalisée qu'à court terme. L'évaluation en conditions réelles d'utilisation indique si le médicament atteint dans un monde réel l'effet souhaité (Strom *et al.*, 2012). Les évaluations en vie réelle complètent les évaluations en conditions contrôlées et permettent notamment d'éclairer les décisions sur les différentes alternatives cliniques, de positionner les nouvelles technologies par rapport aux anciennes, d'augmenter l'utilisation des alternatives cliniques les plus efficaces, de diminuer l'utilisation des moins efficaces et d'identifier les sous-groupes de personnes répondant ou non aux traitements (Strom *et al.*, 2012). En particulier, l'évaluation en vie-réelle peut permettre de comparer l'effet de deux traitements présentant une indication similaire ou utilisés de manière interchangeable dont la comparaison ne nécessite pas une évaluation en condition contrôlée, plus coûteuse. Les études en vie réelle seront également utiles dans un contexte d'évaluation d'un effet à long terme.

2.1 Evaluation en conditions contrôlées

Dans cette section on s'intéresse particulièrement à l'évaluation du médicament conduite dans la phase III d'un essai clinique. A cette étape d'évaluation, le médicament va être évalué dans des conditions particulières (Rothwell, 2005). Les essais cliniques utilisent une méthodologie stricte visant à construire deux groupes comparables afin d'éliminer tout facteur de confusion. De par cette méthodologie, ces essais, appelés essais cliniques randomisés, sont qualifiés comme fournissant le plus haut niveau de preuve (Guyatt *et al.*, 1992; Concato *et al.*, 2000).

Une des premières étapes importantes consiste en la définition des critères d'inclusion ou d'exclusion. Ils sont spécifiques à chaque essai et sont, en général, basés sur l'âge, les

spécificités liées à la maladie de type : avancement de la maladie et/ou gravité et des antécédents du patient (Grossmann and Alper, 2013). Ces critères définissent l'ensemble des caractéristiques de la population à laquelle les résultats pourront s'appliquer. Souvent les critères d'inclusion ou d'exclusion des essais cliniques sont assez stricts. Reflétant ainsi l'effet du médicament dans une population très spécifique, ils ne permettent pas toujours une généralisation des résultats à l'ensemble de la population. Notamment, les patients sélectionnés présentent souvent une maladie moins grave ou avancée, sont souvent plus jeunes et ne présentent pas de comorbidités ou de caractéristique particulière telle que la grossesse, par exemple (Hannan, 2008).

A partir de cette population d'étude, les deux groupes à comparer sont constitués en utilisant des méthodes de randomisation (Lachin, 1988; Chevret, 2003). Cette étape cruciale consiste en la répartition aléatoire, par tirage au sort, des patients dans deux groupes qui feront l'objet ou non de l'intervention. Cette étape de randomisation, ne dépendant que du hasard, assure la comparabilité des deux groupes au démarrage de l'essai.

Après constitution de deux groupes identiques, l'un des groupes, le groupe expérimental, est exposé au médicament que l'on souhaite tester. L'autre groupe, le groupe contrôle, est le plus souvent exposé à un traitement placebo, un « faux » médicament n'étant composé d'aucune substance active. Dans d'autres cas, le groupe contrôle peut être exposé à un traitement de référence dont l'effet est déjà connu. Ensuite, l'essai est conduit de manière prospective.

Afin d'assurer la comparabilité des deux groupes au cours du suivi, l'essai est souvent conduit en double aveugle. Ainsi, ni le patient ni le médecin ne connait le produit auquel le patient est exposé. Ceci permet d'assurer une prise en charge similaire. Enfin, l'analyse est la plupart du temps conduite en intention de traiter. Cela signifie que tous les patients inclus au début de l'étude sont analysés comme étant dans le groupe auquel ils ont été alloués, qu'ils aient finalement reçu le médicament prévu ou non.

2.2 Evaluation en conditions réelles d'utilisation, en vie réelle

Dans un contexte de vie réelle, contrairement à l'évaluation contrôlée, les patients exposés et non-exposés ne sont pas tirés au sort, rendant la construction de deux groupes comparables plus délicates (Strom, 2007).

Des erreurs dans l'évaluation peuvent impacter l'association mesurée et conduire à masquer une association ou à démontrer une association fausse, artificielle ou indirecte. Ceci est l'une des principales difficultés pour conclure sur la question à partir d'études observationnelles. Le résultat obtenu, s'il est biaisé, empêche de fournir une conclusion claire. Ces biais doivent être, au mieux contrôlés, en tout cas être identifiés et décrits, en particulier sur leur sens et leur impact potentiel ou suspecté. Plusieurs biais sont bien connus et attirent l'attention des chercheurs. Au regard des questionnements de ce travail, on peut citer notamment : le biais de sélection, résultant d'un manque de représentativité de la population étudiée et le biais de confusion, incluant le biais d'indication observé lorsqu'un médicament est prescrit préférentiellement à des patients ayant un risque plus élevé de présenter l'évènement considéré, présentant une « nécessité » de recevoir le traitement.

L'évaluation d'un traitement en vie réelle est ainsi marquée par des questionnements méthodologiques spécifiques (Dreyer, 2010). Le biais d'indication peut être contrôlé en utilisant des schémas d'étude ou des méthodes statistiques spécifiques. Nous nous intéresserons en particulier aux méthodes basées sur les scores de propension.

Les scores de propension sont de plus en plus utilisés, en témoigne une rapide recherche dans Pubmed des mots-clés « propensity score » Figure 9.

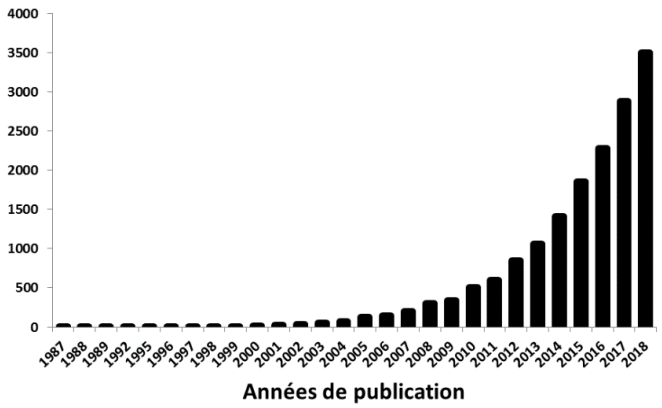


Figure 9 : Nombre de publications mentionnant « propensity score » dans Pubmed jusqu'en 2018
Source : Pubmed, recherche effectuée le 26/07/2019

Le score de propension est un score d'équilibre qui permet de se rapprocher des conditions d'un essai thérapeutique randomisé (Austin, 2011*a*). En effet, un tel score pourra permettre d'obtenir deux groupes, l'un exposé, l'autre non exposé, similaires conditionnellement au score

de propension et par conséquent sur les variables utilisées dans sa construction. Les variables à inclure dans le score sont déterminées à partir de connaissances existantes, des données observationnelles disponibles et de critères statistiques sur lesquels nous reviendrons. De plus, l'utilisation du score de propension peut se faire de plusieurs façons. La méthode utilisée dépend de la question de recherche.

3. Quelles données observationnelles disponibles en SEP ?

Deux types de données observationnelles peuvent être utilisées pour étudier la SEP : les données médico-administratives et les données des dossiers médicaux (Ogdie *et al.*, 2012).

Les données médico-administratives sont issues des systèmes administratifs de suivi des soins. Principalement utilisées à des fins économiques, elles rassemblent des informations concernant la facturation et les remboursements de soins (Ogdie *et al.*, 2012). En France, les principales données de ce type constituent le Système National des Données de Santé (SNDS). Il a déjà été utilisé pour produire des données dans le champ de la SEP, à des fins de recommandations de prise en charge (Roux, 2018). Les données médico-administratives présentent l'avantage de rassembler une information riche pour un grand nombre de personnes (98% de la population française). En revanche, elles ne contiennent que très peu d'informations individuelles, notamment cliniques ou médicales, rendant difficile l'évaluation de l'état de la personne atteinte de SEP. Les travaux pour pallier à cette limite se multiplient, proposant des algorithmes divers pour évaluer l'état clinique de la personne (Roux *et al.*, 2016).

Les données de dossiers médicaux constituent également une source de données importantes en SEP. On s'y intéressera plus particulièrement dans ce travail. Ces données, plus proches de la situation clinique du patient, peuvent également présenter certaines limites, en particulier en termes de représentativité (Ogdie *et al.*, 2012).

En France, l'origine du recueil des données observationnelles en SEP peut être attribuée au projet EDMUS pour « European Database for multiple Sclerosis », initié par le professeur Christian Confavreux dans les années 80 (Confavreux *et al.*, 1992). Il s'inscrit dans une démarche de clarification d'un langage commun et d'informatisation de la donnée, dans le but de pouvoir utiliser les données issues de dossiers médicaux à des fins de recherche et

d'amélioration du soin. L'objectif est de fournir une base de données spécifique à la SEP et un logiciel permettant une collecte standardisée et facilitée. EDMUS est utilisé dans le suivi individuel des patients et dans la réalisation d'études cliniques. En effet, l'interface EDMUS propose notamment un graphique récapitulatif et un synopsis qui permet en consultation d'avoir une vue d'ensemble de l'histoire de la maladie du patient. Un exemple est présenté en annexe C. Son adoption par l'ensemble des centres experts de SEP en France a permis de créer l'OFSEP pour Observatoire Français de la Sclérose en Plaques[1].

L'OFSEP est un projet collaboratif en place depuis 2011 et coordonné par la professeure Sandra Vukusic (Vukusic *et al.*, 2018). Il bénéficie d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche au titre du programme Investissements d'Avenir (référence : ANR-10-COHO-002 OFSEP). L'objectif est de recueillir des informations cliniques, biologiques et d'imagerie concernant les personnes atteintes de SEP, afin de favoriser les recherches et le soin pour mieux comprendre la maladie et améliorer le diagnostic, la prise en charge et le suivi.

La base de données OFSEP est une base de données observationnelles longitudinales comptant environ 68 000 patients suivis dans un des 41 centres experts participants à l'OFSEP[2]. Une fiche minimale définit les données à collecter obligatoirement. De nature démographique et clinique, les données portent sur la description des patients, de leur maladie et de leur prise en charge thérapeutique.

Les données sont collectées en consultation et à une fréquence dépendante du suivi du patient. A la première visite, l'historique est reconstitué à partir de l'interrogatoire auprès du patient et du dossier médical antérieur fourni par le médecin qui a orienté le patient en centre expert. Ensuite, les données de suivi sont collectées de manière prospective. Des suivis, en dehors du centre expert, chez un neurologue libéral par exemple, peuvent tout de même parfois manquer. D'après le descriptif de la cohorte en ligne, au 15 juin 2018, 50% des patients avaient été vus pour la dernière fois en consultation au cours des deux dernières années précédant l'export. Ces patients sont décrits comme faisant partie de la file active[10]. En revanche, 25% des patients n'avaient pas été revus depuis 8 ans ou plus.

Des échantillons biologiques et les IRM sont également collectés. Le travail d'homogénéisation des IRM a été réalisé en 2014 pour proposer des paramètres communs et

[1] *www.edmus.org/fr/proj/observatoire.html*
[2] *www.ofsep.org/fr/la-cohorte-ofsep/descriptif-de-la-cohorte, consulté le 25/07/2019*

formuler des recommandations sur les paramètres à mesurer (Cotton and Edan, 2015). Ce protocole a depuis été révisé[1].

Parmi les centres participants à l'OFSEP, nous évoquerons, en particulier, les données du centre expert rennais qui ont été utilisées dans une partie de ce travail. Le centre expert de Rennes est un centre régional de référence pour la SEP dans l'ouest de la France et a déjà fait l'objet de plusieurs publications scientifiques (Leray *et al.*, 2010; Le Page *et al.*, 2011; Kerbrat *et al.*, 2015; Yaouanq *et al.*, 2015; Ahrweiller *et al.*, 2019). Les patients viennent majoritairement de la Bretagne, ainsi que des régions limitrophes. Depuis janvier 1976, tout nouveau cas de SEP est référencé dans le système, quelle que soit la date de premiers symptômes. A partir de 1996, à la mise à disposition d'EDMUS, les données ont été informatisées.

En France, les seules données de registre disponibles dans la SEP sont les données lorraines, également collectées dans l'OFSEP. Un registre est défini comme « un recueil continu et exhaustif de données nominatives, intéressant un ou plusieurs événements de santé, dans une population géographiquement définie »[2]. Le registre lorrain, nommé ReLSEP est coordonné par le Professeur Francis Guillemin et le professeur Marc Debouverie[3]. Ce registre, labellisé INSERM et INVS-Santé Publique France depuis 2009, rassemble les données de toute la région lorraine que le patient soit suivi en centre expert ou par un neurologue libéral. Au total, 9 établissements, 19 neurologues libéraux et 10 services de médecine physique et réadaptation y participent. Des visites sur le terrain dans toute la région sont organisées pour consulter les dossiers de patients porteurs de SEP mis à disposition par les neurologues. En plus des visites sur le terrain, les caisses d'assurance maladie, les départements d'information médicale des hôpitaux et l'agence technique sur l'information de l'hospitalisation sont sollicités régulièrement. Les données de ce centre ont été utilisées dans une des études de cette thèse, parmi un ensemble de centres. Du fait du recrutement exhaustif en population et du suivi exhaustif y compris hors centre expert, le registre offre une source de données très intéressantes.

En Europe, d'autres pays ont développé des bases de données de type cohorte ou registre (Glaser *et al.*, 2019). Un tableau résumant les caractéristiques de 19 cohortes proposé par Glaser et al est disponible en annexe D. Plusieurs initiatives nationales ou plus localisées, parfois centre

---

[1] *www.ofsep.org/fr/espace-pros/l-imagerie/protocoles-d-acquisitions, consulté le 25/07/2019*
[2] *www.inserm.fr/recherche-inserm/recherche-en-sante-publique/registres-morbidite*
[3] *www.reseau-chu.org/article/relsep-unique-en-france-le-registre-nanceen-sur-la-sclerose-en-plaques/*

spécifique, existent. Les données monocentriques sont utilisées dans certains pays présentant l'avantage de faciliter la consultation des dossiers si cela est nécessaire. On peut par exemple citer, la cohorte longitudinale de Barcelone rassemblant plus de 1300 patients présentant un syndrome cliniquement isolé dont la collecte a débuté en 1995 (Tintoré, 2018). Ces données locales peuvent également être mutualisées entre plusieurs centres, voire à l'échelle du pays, fournissant des données plus larges et éventuellement plus représentatives. On peut évoquer l'exemple de l'Italie, où la collecte a débuté par plusieurs initiatives locales et est devenu une cohorte nationale (Trojano *et al.*, 2019). En particulier, dans ce travail, parmi les données européennes, on s'est intéressé au registre danois (Magyari *et al.*, 2016). Ce registre populationnel comporte tous les patients pour lesquels un diagnostic de SEP a été réalisé. Depuis 1996, cela représente environ 25 000 patients. L'inscription est obligatoire à l'échelle du pays. Les expositions aux traitements, les poussées et de l'EDSS sont récoltés selon le suivi des patients. Notons qu'au Danemark, seuls les neurologues des départements de neurologie des hôpitaux publics sont autorisés à prescrire un traitement de fond.

A l'échelle internationale, depuis 2004, il existe la base MSBase (Butzkueven *et al.*, 2006). Loin d'être exhaustive, elles rassemblent les données collectées dans 34 pays du monde. La base de données MSBase est une base de données observationnelles longitudinales comptant environ 62 000 patients pris en charge dans 129 cliniques ou centres volontaires . Les données sont issues du suivi médical des patients. Elles sont collectées au cours des consultations. En théorie, les patients devraient avoir au moins une visite annuelle avec une mise à jour des données mais ceci n'est pas un critère requis pour l'inclusion. Les pratiques et recommandations ainsi que les traitements disponibles dans les pays inclus dans les cohortes MSBase ne sont pas uniformes.

Enfin, les données médico-administratives et les données issues du dossier médical peuvent être chaînées. Ces données chaînées offrent la possibilité de combiner les avantages de l'exhaustivité de suivi des patients au travers des données de consommations de soins et de l'apport primordial des données issues de la prise en charge comme les mesures cliniques et les variables individuelles. Comme exemple, on peut citer la cohorte rétrospective canadienne chaînant les données de la base British Columbia Multiple Sclerosis (BCMS) et les données médico-administratives issues de diverses bases de données relatives aux motifs de recours aux soins, aux séjours hospitaliers, aux prescriptions médicamenteuses, mais également au statut vital, d'immigration et socio-économique (Sweeney *et al.*, 1986; Shirani *et al.*, 2012).

En France, le chaînage des données de l'OFSEP aux données du SNDS est en cours de réalisation en collaboration avec l'équipe REPERES. Ce chaînage pourrait concerner environ 16 000 patients inclus dans l'OFSEP pour lesquels les données déjà disponibles pourront être complétées avec les données de consommations de soins issues de la base médico-administrative. Ce chiffre correspond à des patients vus entre 2010 et 2018 à l'OFSEP (période de disponibilité du SNDS) et qui ont tous donné leur consentement pour un tel chaînage. La base obtenue offrira de belles opportunités d'études mais présentera toujours certaines limites déjà évoquées. En effet, les données individuelles ou celles sur l'évaluation de l'état clinique du patient ne pourront être plus complètes puisque non disponibles dans le SNDS. En revanche, les consommations de soins et traitements, éventuellement des variables proxy permettant d'établir une estimation de l'état du patient, pourront nous permettre de récolter des données de suivi plus exhaustives, en particulier pour les patients dont le suivi s'effectue en dehors du centre expert malgré des consultations ponctuelles en centre. Enfin, ce chaînage ne permettra pas de limiter le biais de recrutement de l'OFSEP dû à la collecte de données en centre expert mais permettra de le caractériser en comparant les patients OFSEP aux patients du SNDS.

D. Problématique et objectifs de la thèse

1. Problématique

Comme nous l'avons vu précédemment, le premier traitement spécifique à la SEP, disponible au milieu des années 90, a largement fait évoluer la prise en charge de la maladie. Depuis, l'arsenal thérapeutique s'est développé, ainsi que les connaissances sur l'histoire naturelle de la maladie et sur le rapport bénéfice-risque des traitements et l'expérience des neurologues dans leur utilisation. Ces évolutions ont influencé les stratégies thérapeutiques de la SEP. Dans ce contexte, il apparait important d'avoir une meilleure compréhension des effets de ces traitements, de ces stratégies en conditions réelles d'utilisation et en particulier à long terme.

Des études ont déjà été conduites sur ces questionnements, plus spécifiquement concernant l'effet des traitements de première ligne. La question de leur bénéfice à long terme a déjà été explorée dans un contexte de vie réelle mais les résultats controversés ne permettent pas de conclure. Parmi les éléments pouvant expliquer ces variations, le contrôle du biais

d'indication a souvent posé question et n'a jamais été correctement contrôlé, fournissant des résultats biaisés. De plus, ces traitements de première ligne étant des traitements largement prescrits, le groupe contrôle identifié n'est pas toujours adapté (Sormani and Bruzzi, 2014).

Les premiers traitements de deuxième ligne ont également été comparés entre eux dans la littérature. De nouveau, les résultats apportés n'ont pas permis de répondre à la supériorité de l'un ou de l'autre, en raison de résultats variables d'une étude à l'autre, empêchant ainsi de fournir un message clair aux neurologues.

Dans ce travail, nous avons donc cherché à éclairer l'évaluation des traitements en vie réelle.

Outre la comparaison de traitements, l'efficacité des différentes stratégies thérapeutiques n'a encore été que peu explorée. Si l'importance d'une initiation précoce a été démontrée, son efficacité à long terme mérite d'être mieux documentée. De plus, si la littérature converge déjà vers une recommandation d'initiation précoce de traitement, la quantification du gain de vie sans progression du handicap en fonction du délai d'initiation du traitement contribuerait à préciser ces recommandations.

Enfin, les trois principales stratégies thérapeutiques ont été peu décrites et, à notre connaissance, jamais comparées en France. Dans ce travail, nous nous sommes donc également interrogés sur le moment de l'initiation et sur l'efficacité des stratégies principales utilisées : l'escalade thérapeutique, le traitement intensif et l'induction.

2. Objectifs

2.1 Objectif principal

Ce travail de recherche, encadré par Professeur Gilles Edan et Docteure Emmanuelle Leray, avait pour objectif principal d'évaluer l'effet des stratégies thérapeutiques en conditions réelles d'utilisation sur la progression du handicap chez les patients atteints de SEP rémittente. Quatre sous-objectifs peuvent être distingués :

(1) Etudier l'effet des interférons-β et de l'acétate de glatiramère sur la progression du handicap en conditions réelles d'utilisation (centre expert de Rennes)
(2) Comparer l'efficacité du fingolimod et du natalizumab à partir de 3 cohortes observationnelles (MSBase, Registre danois, OFSEP)

(3) Mesurer l'effet d'un traitement précoce sur la progression du handicap et l'évaluer en fonction du délai d'initiation du traitement depuis le début de la SEP, à partir des données de l'OFSEP
(4) Décrire et comparer les stratégies d'induction et d'escalade : patients concernés, comparaison des effets sur la progression du handicap à partir des données de l'OFSEP

2.2 Objectif secondaire

L'objectif secondaire portait sur l'exploration des méthodes de score de propension, utilisées dans le but de contrôler le biais d'indication en vie réelle.

Chapitre 2

II. Méthodes

A. Inférence causale 38
1. Approche graphique 39
1.1 Principe 39
1.2 Définitions 40
1.3 DAG en pharmacoépidémiologie 43
2. Approche contrefactuelle 44
B. Score de propension 46
1. Principe 46
2. Méthodes utilisant le score de propension 46
3. Appariement 48
4. Pondération 49
5. Sélection des variables pour spécifier le score de propension 50
6. Les hypothèses en pratique 50
C. Analyse de survie 52
1. Délai et fonctions : définitions et notations 52
2. Notion de censure 53
3. Temps de participation 54
4. Estimation de la fonction de survie 55
4.1 Définition et notation 55
4.2 Utilisation et interprétation 56
4.3 Comparaison de deux groupes à partir des courbes de survie 57
4.4 Pondération de la courbe de survie 57
5. Modèle à risques proportionnels de Cox 57
5.1 Hypothèses du modèle de Cox 58
5.2 Les modèles dérivés du modèle de Cox original 59

Dans ce deuxième chapitre, nous proposons de présenter les méthodes utilisées de manière transversale dans les trois axes de la thèse. Nous y ferons référence dans le reste du document. Les notions associées aux scores de propension et aux méthodes relatives à l'analyse de survie sont détaillées dans ce chapitre. Au préalable, nous avons choisi de présenter quelques notions d'inférence causale.

A. Inférence causale

Comme présenté précédemment, en pharmaco-épidémiologie, on cherche notamment à établir un lien causal entre une exposition à un traitement T et la survenue d'un évènement d'intérêt E, souvent un changement d'état (décès, apparition de symptômes, maladie, etc.). Etablir un lien causal, c'est établir une relation entre un évènement, la cause, ici le traitement, et un deuxième évènement, conséquent du premier.

Les notions d'association et de causalité sont souvent juxtaposées évoquant que l'association n'est pas la causalité, comme nous l'avons évoqué dans la partie contextuelle. Pour autant, notamment à partir de la notion d'association, Sir Austin Bradford Hill a défini, en 1965, 9 critères pour discuter de la causalité : force d'association, relation dose-effet, stabilité de l'association ou reproductibilité, cohérence, spécificité, temporalité, plausibilité, preuve expérimentale et analogie (Hill, 1965). Ces critères permettent d'alimenter la discussion mais ne permettent pas de conclure sur la causalité de l'association. En 2016, ils ont été rediscutés, concluant que des neufs critères, seules la preuve expérimentale et la reproductibilité devraient être conservées (Ioannidis, 2016).

Depuis les critères de Hill, des méthodes ont en revanche été développées dans l'objectif d'établir des relations causales. Ces méthodes reposent sur deux approches fournissant un cadre conceptuel en inférence causale : l'approche graphique et l'approche contrefactuelle. L'approche contrefactuelle repose sur les notations algébriques tandis que l'approche graphique permet d'exprimer les relations de cause à effet à partir de schémas appelés « graphes orientés acycliques » ou DAG pour « Directed Acyclic Graphs ».

1. Approche graphique

En causalité, les DAG (Pearl, 2009) ont été proposés comme un outil permettant de formaliser les hypothèses de causalité. Ils donnent une vision schématique du cadre d'analyse et peuvent faciliter la compréhension et la formalisation des hypothèses, puis des analyses. Les hypothèses formalisées dans un DAG sont qualitatives, c'est-à-dire qu'elles traduisent l'existence ou non d'une relation mais ne préjugent pas de sa nature et son intensité.

La rédaction de cette partie est basée sur le cours en ligne proposé par Miguel Hernàn de l'université de Harvard intitulé « Causal Diagrams: Draw Your Assumptions Before Your Conclusion »[1].

1.1 Principe

Dans un DAG, la relation entre deux variables est représentée par des flèches. Ces flèches sont tracées entre des nœuds (ancêtres et descendants), représentant des variables dans des relations pouvant être directes (la variable 1 affecte directement la variable 2) ou indirectes (une autre variable est impliquée dans la relation entre les variables 1 et 2). Le sens de la flèche nous informe sur le sens de la relation.

La Figure 10 représente un DAG. On y trouve trois nœuds (A, T, E) formalisant trois variables, par exemple T, le traitement, E, l'évènement et A, l'activité de la SEP. La flèche entre A et T indique que A a un effet sur T. Sur le graphe, on peut également représenter que T a un effet sur E et A sur E. En ce sens, le graphe est dit orienté (« directed »). On dit également que le graphe est acyclique (« acyclic »). Cela fait référence au fait qu'une variable ne peut jamais être reliée à elle-même, même par des chemins indirects faisant intervenir d'autres variables. Par exemple, on ne pourrait pas représenter une flèche allant de T vers A. En ce sens, sur la schématisation, nous avons ajouté une flèche en pointillé gris clair formalisant le temps. Cette flèche n'appartient pas au DAG. Elle illustre simplement le fait que le passé impacte le futur mais que le futur n'impacte jamais le passé.

[1] *www.edx.org/course/causal-diagrams-draw-your-assumptions-before-your-conclusions*

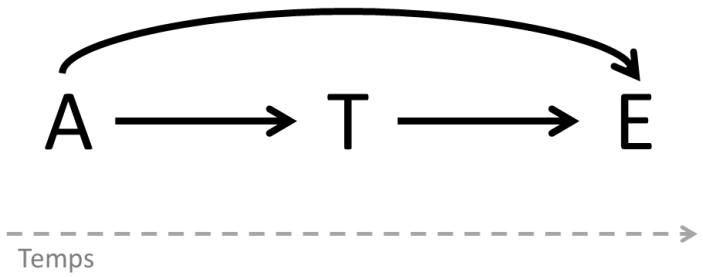


Figure 10 : Exemple de graphe orienté acyclique ou « directed acyclic graph » (DAG)

Deux variables non reliées témoignent d'une hypothèse forte, illustrant qu'il n'y a pas d'effet entre elles, excluant donc tout effet causal de l'une sur l'autre. En revanche, la présence d'une flèche ou d'un chemin entre deux variables (en suivant le sens des flèches) implique que l'une est la cause de l'autre. Dans l'exemple, la variable A, affectant T et E, est dite cause commune du traitement T et de l'évènement E. La formalisation d'un DAG implique une bonne connaissance de la question de recherche et de la littérature.

1.2 Définitions

Quelques éléments complémentaires méritent encore d'être définis pour simplifier la compréhension de cette partie.

Une variable sur laquelle pointent deux flèches est un « collider » ou en français collisionneur mais ce terme est peu utilisé. Nous privilégierons donc l'utilisation du terme de collider. Le collider est spécifique au chemin. Cela signifie qu'une variable peut être collider sur un chemin du DAG mais ne pas l'être sur un autre chemin de ce même DAG. Par exemple, sur le DAG représenté en Figure 11, la variable B est un collider sur le chemin $A \rightarrow B \leftarrow C$ mais ne l'est pas pour le chemin $A \rightarrow B \rightarrow D$.

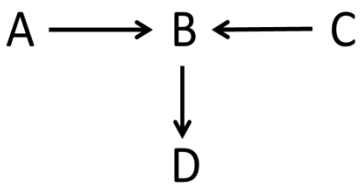


Figure 11 : Variable « collider » et « non-collider »

Tout contrôle ou conditionnement réalisé sur une variable est formalisé par une boîte, encadrant la lettre. Il peut s'agir par exemple de la restriction de l'étude chez les personnes ayant réalisé l'évènement. Dans ce cas, E sera encadré pour schématiser cette restriction. Ce peut être également un ajustement sur une variable ou une stratification dans une analyse statistique.

Enfin, les chemins peuvent être ouverts ou fermés. Cette qualification est reliée à la notion de « D-séparation », dont les règles sont résumées ci-après :

- Règle 1 : Si aucune variable n'est conditionnée, un chemin est fermé si et seulement si 2 pointes de flèches entrent en collision avec une variable du chemin (Figure 12)

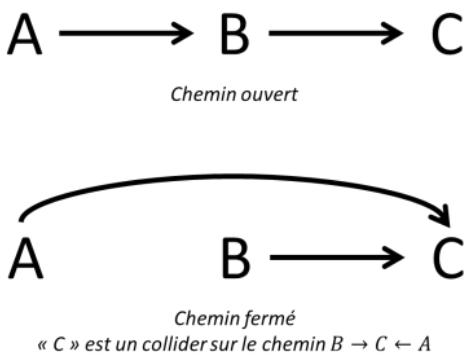


Figure 12 : règle 1 de « D-séparation »

- Règle 2 : Tout chemin contenant une variable non- collider sur laquelle un conditionnement ou contrôle est réalisé est fermé (Figure 13)

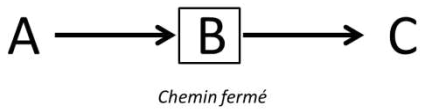


Figure 13 : règle 2 de « D-séparation »

- Règle 3 : Conditionner sur ou contrôler un collider ouvre le chemin (Figure 14)

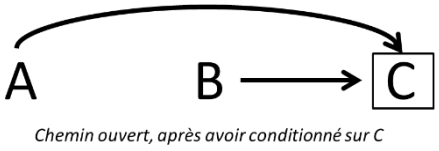


Figure 14 : règle 3 de « D-séparation »

- Règle 4 : Conditionner ou contrôler sur/pour un descendant d'un collider ne ferme pas le chemin (Figure 15)

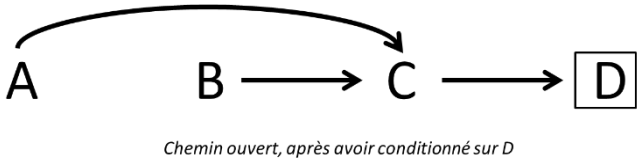


Figure 15 : règle 4 de « D-séparation »

Finalement, dans un DAG, deux variables sont dites D-séparées, ou indépendantes par un ensemble de variables Z si et seulement si tous les chemins entre les deux variables sont fermés sachant Z. Par exemple, dans le DAG représenté Figure 12, les variables A et B sont dites D-séparées ou encore indépendantes sachant C (collider, fermant donc le chemin). Dans le DAG représenté Figure 15 en revanche, A et B sont dites D-connectées ou dépendantes sachant D (variable sur laquelle un conditionnement est réalisé, ouvrant le chemin).

Ces représentations graphiques sont utiles pour formaliser les hypothèses. Elles peuvent faciliter la compréhension d'une question de recherche. Ensuite, un lien entre ces représentations graphiques et la statistique peut être fait. En effet, un chemin ouvert traduit l'existence d'une association statistique entre les variables situées aux extrémités du chemin. En revanche, un chemin fermé illustre l'absence d'association entre deux variables, bien qu'une association puisse néanmoins exister via d'autres chemins.

En conditionnant sur ou en contrôlant une variable, on pourra, en ouvrant ou fermant le chemin et selon la nature de la variable (collider ou non), engendrer ou masquer une association.

1.3 DAG en pharmacoépidémiologie

Revenons sur notre question spécifique qui concerne l'efficacité d'un traitement. En situation observationnelle, dans une population, deux groupes se distinguent, l'un traité, l'autre non, mais l'utilisation du traitement dépend d'un choix fait au regard de données observées. Ces éléments sont à prendre en compte dans l'évaluation de l'effet causal. Dans un tel contexte, des variables qui sont causes communes du traitement et de l'évènement seront alors ajoutées au DAG, comme la variable L de la Figure 16.

La Figure 16 représente un DAG que l'on pourrait utiliser pour formaliser nos hypothèses dans un contexte d'étude observationnelle. Ainsi, la flèche entre T et E représente une association entre l'exposition à un traitement et la survenue d'un évènement. Le traitement et l'évènement ont une ou des causes communes que nous formalisons par L pour simplifier.

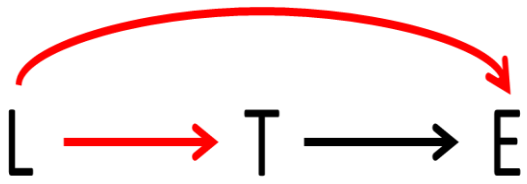


Figure 16 : DAG en situation observationnelle, notion de chemin "porte-arrière"

D'après les éléments précédemment présentés, l'association entre T et E a deux composantes :

- Un effet de T sur E
- Un effet dû à la cause commune L

Le chemin, représenté en rouge dans la Figure 16 est appelé chemin porte arrière ou « backdoor path ». Ce chemin, ouvert, permet une association de T et E par L. La fermeture du chemin, en contrôlant sur L entraîne que les variables T et E ne sont plus associées par l'intermédiaire de L mais restent associées en raison du chemin direct reliant T à E (en noir). Plus généralement, pour que l'effet que l'on évalue de T sur E soit causal, il faut fermer tous les chemins « porte arrière » qui relient T et E. Ceci, en prenant garde à ne pas contrôler à tort sur une autre variable. Par exemple, en contrôlant sur un collider ou un descendant d'un collider, on ouvre le chemin, créant une autre association non causale. Ce peut être le cas d'un biais de sélection. En effet, dans une étude, on peut sélectionner un sous-groupe de patients qui présentent des caractéristiques particulières, liées directement ou non à l'évènement.

On peut noter que, dans un essai randomisé, le traitement étant alloué aléatoirement, il n'y a par conséquent pas de cause commune au traitement et à l'évènement. Le DAG formalisant l'association causale du traitement et de l'évènement est donc simplement représenté par $T \rightarrow E$ (Figure 17).

Figure 17 : DAG d'un essai randomisé

Les DAG offrent la possibilité d'expliciter les hypothèses de causalité relatives à une question de recherche. Dans la suite, nous détaillerons simplement, sans entrer dans un détail mathématique trop complexe, l'approche contrefactuelle.

2. Approche contrefactuelle

L'approche contrefactuelle en pharmaco-épidémiologie repose sur le fait que pour un patient, deux issues sont possibles : son devenir s'il est traité et son devenir s'il ne l'est pas. Lorsque l'on évoquait dans la partie de contexte, la situation idéale mais irréaliste de pouvoir « ré-observer » le devenir d'un patient avec une autre exposition que celle qui l'a reçue, on abordait d'ailleurs cette notion de monde contrefactuel.

La situation contrefactuelle peut donc être définie comme « ce qui aurait été observé si le traitement avait été différent ». De fait, l'effet causal est défini comme étant la différence entre le devenir observé et le devenir contrefactuel. Autrement dit, dans le cas d'une exposition binaire, le résultat que l'on observerait dans le cas où le patient i aurait été traité $Y_{T=1}(i)$ et celui qu'on observerait dans le cas où il ne l'aurait pas été $Y_{T=0}(i)$. De toute évidence, une telle observation est impossible, les deux issues potentielles ne pouvant être observées simultanément. Le raisonnement s'effectue par conséquent à l'échelle populationnelle. L'idée étant de « reconstruire » une population contrefactuelle. Ce raisonnement repose sur trois hypothèses : la cohérence, l'échangeabilité et la positivité (Hernàn and Robins, 2018).

L'hypothèse de cohérence, parfois appelée consistance car on y réfère comme l'hypothèse de « consistancy » en anglais, repose sur la notion de monde contrefactuel. Elle est définie comme le fait que le devenir d'un patient s'il recevait l'exposition observée est égale au

devenir observé. Autrement dit, l'évènement d'un patient sous l'exposition qu'il a vraiment reçue est égal à l'évènement réellement observé.

L'hypothèse d'échangeabilité fait référence au fait que si les patients du groupe exposés restaient non-exposés, on observerait les mêmes devenirs moyens que dans le groupe de non exposés et vice-versa. Dans un essai randomisé, cette hypothèse est assurée par la randomisation. Dans une étude observationnelle, on ne peut pas vérifier cette hypothèse. Elle se traduit par l'absence de facteurs de confusion non mesurés.

Finalement, l'hypothèse de positivité donne un sens au devenir contrefactuel. Tous les patients inclus dans l'étude auraient pu être exposés ou non-exposés. Ainsi, tous les patients doivent avoir une probabilité non nulle d'appartenir à chaque groupe. Cela repose par exemple sur le fait qu'il n'y ait pas de contrindication à recevoir le traitement ou que les deux traitements soient disponibles pour tous les patients.

Certains auteurs proposent d'appréhender la construction de cette pseudo-population comme on chercherait à réaliser un essai randomisé : « Target trial » (Hernàn and Robins, 2018). Est-ce que chaque patient à une probabilité non nulle de recevoir l'intervention ? Connaît-on la probabilité de recevoir l'intervention ? Peut-on identifier le temps t_0 c'est-à-dire le moment de la randomisation ? On peut même poursuivre le raisonnement et définir un protocole en spécifiant les éléments clés : critère d'inclusion, intervention, critère de jugement, suivi, analyse statistique.

La formalisation des modèles causaux ainsi faite, des estimations d'effets causaux peuvent être établies à partir de données observationnelles et de méthodes spécifiques faisant appel à la notion de monde contrefactuel. Dans ce travail, ce sont sur les scores de propension que nous avons porté notre intérêt.

B. Score de propension

1. Principe

Le but de l'utilisation d'un score de propension est d'équilibrer les groupes que l'on souhaite comparer afin de se rapprocher des conditions d'un essai thérapeutique randomisé (Austin, 2011a).

Ce score est défini comme la probabilité (ou propension) qu'un patient soit traité sachant ses facteurs de confusion potentiels (Rosenbaum and Rubin, 1983). Mathématiquement, on notera p_i la probabilité qu'un patient i soit exposé sachant les facteurs de confusion Z_i. Ainsi le score de propension peut être défini par l'expression suivante : $p_i = P(T_i = 1 \,|Z_i = z_i)$ où T est la variable d'exposition ($T = 1$ pour les exposés et $T = 0$ pour les non exposés). Le score de propension est généralement estimé à partir d'une régression logistique ajustée sur les facteurs de confusion potentiels mesurés. Notons dès à présent, l'importance de ce dernier point : les méthodes utilisant les scores de propension ne pourront contrôler la confusion que sur les facteurs connus et mesurables (Strom, 2007). Ainsi les patients seront rendus comparables uniquement sur ces facteurs. Une fois le score de propension estimé, il pourra être utilisé à partir de différentes méthodes.

2. Méthodes utilisant le score de propension

On distingue quatre méthodes d'utilisation du score de propension : (1) l'ajustement sur le score, (2) la stratification sur le score, (3) l'appariement sur le score et (4) la pondération à partir du score (Austin, 2011a). L'ajustement, proche de la méthode classique de régression multiple, consiste en l'utilisation du score comme une variable résumant l'information des variables de confusion. La stratification consiste à comparer les groupes par strates définies par le score de propension. Les quintiles sont souvent utilisés. L'appariement consiste à former des paires de sujets exposés et non exposés ayant des valeurs de score de propension proches. Enfin, la pondération consiste à pondérer les patients par des poids dépendants du score de propension.

Dans la littérature, ces méthodes ont été comparées par simulation afin d'établir les performances de chacune dans l'analyse des critères de jugement (mesure d'un odds ratio, hazard ratio, estimation d'effet absolu, différence de proportions). Il en résulte que la stratification et l'ajustement présentent de mauvaises performances (Austin, 2007*a*, *b*, 2013). Leur utilisation n'est donc pas recommandée. En revanche, l'appariement et la pondération ont de bonnes performances, en particulier elles présentent une bonne capacité à bien équilibrer les groupes (Austin, 2009, 2010, 2013; Gayat *et al.*, 2012; Austin and Schuster, 2014).

Notons que deux effets peuvent être distingués : l'effet conditionnel et l'effet marginal. L'effet conditionnel est l'effet moyen de l'exposition au niveau individuel, c'est-à-dire l'effet moyen associé au passage d'un individu du statut d'exposé à celui de non-exposé (ou inversement). En présence de facteurs de confusion, cet effet est souvent estimé par des modèles de régression classique. L'effet marginal est l'effet moyen de l'exposition au niveau populationnel, c'est-à-dire l'effet moyen associé au passage d'une population entière du statut d'exposé à celui de non-exposé (ou inversement). Dans ce travail, on cherche à estimer des effets marginaux, c'est-à-dire, au niveau d'une population (Austin, 2011*a*). Parmi les méthodes ci-dessus, l'appariement et la pondération permettront d'obtenir des effets marginaux. Deux effets marginaux peuvent être distingués (Imbens, 2004; Pirracchio *et al.*, 2013; Hajage, 2017).

L'un est l'effet moyen associé au passage d'une population entière d'un statut d'exposé à celui de non-exposé ou l'inverse. Il est appelé ATE pour « average treatment effect in the entire population », c'est-à-dire effet moyen de l'exposition dans l'ensemble de la population. L'autre est l'effet moyen associé au passage de l'ensemble de la population exposée à un statut de non-exposé. Il est appelé ATT pour « average treatment effect on the treated », c'est-à-dire effet moyen de l'exposition dans la population des exposés.

Dans les essais cliniques randomisés, ces deux effets sont similaires puisque la randomisation assure que la population exposée est similaire à la population globale. En revanche dans les études observationnelles, ils peuvent différer. Choisir de mesurer un effet ATE ou ATT dépend donc de la question clinique (Austin, 2011).

Concernant les méthodes évoquées précédemment, l'appariement permet d'estimer l'ATT. La pondération, selon le poids que l'on utilise, permet d'estimer l'ATT et l'ATE. Dans les parties suivantes, ces deux méthodes vont être détaillées.

3. Appariement

Le principe de l'appariement consiste à apparier les patients exposés aux patients non exposés présentant une valeur proche de score de propension. Plusieurs algorithmes d'appariement ont été développés mais nous nous intéresserons particulièrement à l'approche la plus simple et la plus fréquemment utilisée (Austin, 2011*a*, 2014). Elle consiste en un appariement :

- 1:1, c'est-à-dire une paire de chaque statut
- appelé « greedy » en anglais, un patient exposé est apparié avec le patient non exposé ayant le score de propension le plus proche, même si ce dernier aurait pu constituer une meilleure possibilité d'appariement avec un autre patient
- sans remplacement : les patients ne sont appariés qu'une seule fois
- par méthode du plus proche voisin, en utilisant une distance maximale appelé un « caliper » : la distance maximale entre les patients appariés ne peut pas être supérieure à ce caliper. Il est défini comme une proportion de l'écart-type du score de propension. Il a été montré qu'un caliper de 0.2 fournissait de bonnes performances (Austin, 2011*b*). Notons que plus le caliper est faible, plus l'appariement est contraint.

Une fois que l'appariement a été obtenu, l'effet du traitement peut être estimé directement en comparant le(s) critère(s) de jugement entre les patients des groupes exposés et non exposés. Notons tout de même qu'après appariement, les données appariées ne sont plus indépendantes. La nature dépendante des données doit être prise en compte dans l'estimation de la variance (Austin, 2011*a*) ou dans le choix du modèle dans lequel on utilise les données appariées. Notamment, un cluster pourra être utilisé (Nayan *et al.*, 2017). Notons dès à présent que la méthode d'appariement conduit à exclure des patients par construction. En effet, les patients exposés sans correspondance dans le groupe non exposé seront exclus. De plus, certains patients non exposés pourront ne pas être appariés. Cette méthode est donc associée à une perte de puissance statistique ainsi qu'une modification potentielle de la population cible initiale.

4. Pondération

La méthode de pondération consiste à pondérer chaque patient selon un poids obtenu à partir du score de propension p_i.

D'une part, lorsque on cherche à estimer l'ATE, on utilise un poids appelé IPTW pour « inverse probability of treatment weighting ». Comme son nom l'indique, ce poids est donc inversement proportionnel à la probabilité qu'a le patient d'appartenir à son groupe d'exposition. Ainsi, on attribue aux patients avec un profil sous représenté, un poids élevé et aux patients présentant un profil surreprésenté un poids faible (Le Borgne, 2016). Mathématiquement, le poids attribué à un patient exposé est $w_i = \frac{1}{p_i}$ et le poids attribué à un patient non exposé est $w_i = \frac{1}{1-p_i}$. Cette définition peut conduire à l'estimation de poids très élevés pour les patients ayant un profil très rare. De ce fait, il a été recommandé d'utiliser des poids dits « stabilisés » c'est-à-dire des poids tenant compte de la probabilité d'appartenir au groupe auquel on appartient. Ainsi, le poids stabilisé (sIPTW) est défini par la formule suivante :

$$w_i = \begin{cases} \dfrac{P(T_i = 1)}{p_i} \; pour\ les\ patients\ exposés \\ \dfrac{P(T_i = 0)}{1 - p_i} \; pour\ les\ patients\ non\ exposés \end{cases}$$

Notons qu'il a également été proposé de limiter la valeur maximale d'un poids (Xu et al, 2010). On parle alors de troncature des poids. Ces poids sont construits de manière à utiliser l'ensemble de l'échantillon. Par construction, la distribution des covariables dans l'échantillon pondéré est la même que la distribution globale initiale (cohorte originale entière, non pondérée). C'est pourquoi il permet d'estimer l'ATE.

Une autre pondération qui permet d'estimer l'ATT est appelée pondération par les odds. Elle est calculée de manière à construire un échantillon pondéré présentant les mêmes distributions de covariables que les patients exposés (cohorte exposée originale). Ainsi, un poids de 1 est attribué aux patients exposés et un poids de $w_i = \frac{p_i}{1-p_i}$ est attribué aux non exposés.

5. Sélection des variables pour spécifier le score de propension

Les variables à inclure dans le score peuvent répondre à quatre définitions différentes : (1) toutes les variables associées à la fois à l'exposition et au critère de jugement d'intérêt, appelées vrais facteurs de confusion, (2) toutes les variables associées à l'exposition, qu'elles soient ou non associées au critère de jugement, (3) toutes les variables associées au critère de jugement, qu'elles soient ou non associées à l'exposition et (4) toutes les variables observées (Austin, 2011*a*).

Plusieurs auteurs se sont intéressés à la question de la spécification du modèle en étudiant par simulation l'estimation de la variance et les biais. Les études conduites indiquent qu'il est préférable d'inclure dans le score de propension des variables associées au critère de jugement qu'elles soient ou non associées à l'exposition (Brookhart *et al.*, 2006; Austin *et al.*, 2007; Witte and Didelez, 2018). Il a parfois également été recommandé l'inclusion de toutes les variables disponibles (Austin, 2007*b*). En revanche, les variables liées au traitement mais pas au critère de jugement, dites variables instrumentales, ont montré qu'elles entrainaient une augmentation du biais et de la variance dans la mesure de l'effet (Myers *et al.*, 2011). Leur utilisation n'est donc pas recommandée.

Dans tous les cas, il est important que ces variables soient mesurées avant le début de l'exposition. Des variables mesurées après l'initiation du traitement peuvent être influencées ou modifiées par le traitement (Austin, 2007*a*).

6. Les hypothèses en pratique

Comme évoqué précédemment, dans les méthodes d'inférence causale, trois hypothèses sont requises : la cohérence, l'échangeabilité, la positivité. L'utilisation des scores de propension conduit à en définir une quatrième : l'absence d'erreur de spécification du modèle (Cole and Hernàn, 2008; Austin and Stuart, 2015). Il convient de vérifier ces hypothèses, ou tout du moins, de les discuter.

Certaines d'entre elles sont difficiles à vérifier en pratique. L'hypothèse de cohérence repose sur la spécification correcte du modèle et celle d'échangeabilité en s'assurant d'avoir collecté l'ensemble des variables affectant l'exposition et/ou le critère de jugement (Le Borgne,

2016). Ces hypothèses ne peuvent être évaluées formellement, leur vérification repose plutôt sur une discussion.

Pour évaluer le respect de l'hypothèse de positivité, l'évaluation du support commun peut être réalisée (Vaughan *et al.*, 2015). Le support commun est défini par le chevauchement des distributions entre les groupes d'exposition. Ce chevauchement indique qu'un patient du groupe exposé pourrait appartenir au groupe non exposé et que pour chaque niveau du score de propension, on peut retrouver des patients des deux groupes. Un bon chevauchement supporte l'hypothèse d'échangeabilité et de positivité. Un manque de support commun témoigne de différences persistantes entre les groupes.

De plus, il est également important d'évaluer la qualité de l'équilibre (Vaughan *et al.*, 2015). Comme évoqué, le score de propension est un score d'équilibre, et de fait conditionnellement au score de propension, les distributions de variables sont similaires chez les patients exposés et non exposés. On évalue donc la capacité du score de propension à bien rendre les deux groupes comparables sur les variables utilisées dans le score. Pour conduire cette évaluation, il est recommandé d'utiliser les différences standardisées pour les variables continues et binaires (Austin, 2011*a*). Pour les variables continues, la différence standardisée est définie par la formule ci-après :

$$d = \frac{\hat{p}_{T=1} - \hat{p}_{T=0}}{\sqrt{\frac{\hat{p}_{T=1}(1-\hat{p}_{T=1}) + \hat{p}_{T=0}(1-\hat{p}_{T=0})}{2}}}$$

où $\hat{p}_{X=1}$ et $\hat{p}_{X=0}$ sont les proportions de patients dans l'échantillon pondéré ou du groupe apparié présentant la caractéristique d'intérêt dans le groupe exposé et non exposé respectivement.

Pour les variables binaires, la différence standardisée est définie par la formule ci-après :

$$d = \frac{\bar{z}_{T=1} - \bar{z}_{T=0}}{\sqrt{\frac{{s^2}_{T=1} + {s^2}_{T=0}}{2}}}$$

où $\bar{z}_{X=1}$ et $\bar{z}_{X=0}$ sont les moyennes pondérées ou du groupe apparié de la variable d'intérêt dans le groupe exposé et non exposé respectivement, de même, ${s^2}_{X=1}$ et ${s^2}_{X=0}$ sont les variances correspondantes.

Pour évaluer l'équilibre des variables catégorielles à plus de 2 catégories, nous avons utilisé la distance de Mahalanobis vue comme la dissimilarité entre deux vecteurs de données (Mahalanobis, 1936), dont le carré est défini par :

$$d^2 = (y_{T=1} - y_{T=0})^T \Sigma^{-1} (y_{T=1} - y_{T=0})$$

Où y représente le vecteur de données dans chaque groupe et Σ est la matrice de variance covariance.

Des différences autour de 10% sont considérées comme acceptables (Austin, 2011a). Lorsque l'équilibre n'est pas satisfait, les résultats resteront biaisés et il sera nécessaire de rechercher un meilleur score de propension. Dans la littérature, des solutions comme l'inclusion de facteurs supplémentaires ou d'interactions sont évoquées (Le Borgne, 2016).

C. Analyse de survie

L'analyse de survie consiste en l'étude du délai de survenu d'un évènement. Cet événement correspond à un changement d'état, comme le décès, la contraction d'une maladie ou l'apparition de symptômes par exemple. Ces analyses sont réalisées à partir d'un groupe d'individus, couramment des cohortes ou des essais thérapeutiques pour lesquels on a collecté des données longitudinales. En raison d'un temps de suivi trop court ou de sortie d'étude, ces données sont souvent incomplètes ce qui confère un caractère particulier à l'analyse de survie.

Cette section est essentiellement basée sur le chapitre 4 de l'ouvrage de Daniel Commenges et Hélène Jacqmin-Gadda intitulé « Modèles biostatistiques pour l'épidémiologie » (Commenges and Jacqumin-Gadda, 2015).

1. Délai et fonctions : définitions et notations

Comme évoqué précédemment, dans une telle analyse, nous cherchons à étudier le délai de survenue d'un événement d'intérêt. Il découle ainsi naturellement la nécessité de définir deux dates : la date d'origine, pour laquelle on utilise parfois le terme anglais « baseline » et la date de la survenue de l'évènement. Le choix de la date d'origine dépend à la fois de la question de recherche et de l'évènement. Souvent il existe un choix naturel, parfois, la date d'origine est plus difficile à identifier. Concernant l'évènement d'intérêt, il est nécessaire d'en avoir une définition précise.

Dans la définition du délai, d'autres dates peuvent être pertinentes à évoquer : la date de dernières nouvelles et la date de point. La date de dernières nouvelles est la date la plus récente à laquelle des données ont été collectées pour un patient. La date de point est la date au-delà de

laquelle on ne tiendra pas compte des données observées. Cette date est souvent la date de fin d'étude mais peut y être antérieure dans certains cas.

Le délai est représenté par une variable aléatoire positive T. Cette variable aléatoire est supposée continue. Plusieurs fonctions peuvent définir la loi de probabilité de T. On évoquera ici seulement d'eux d'entre elles : la fonction de survie, notée $S(t)$ et la fonction de risque, notée $h(t)$.

La fonction de survie $S(t) = P(T > t)$ représente la probabilité de subir l'évènement au-delà d'un temps t. Cette fonction décroissante de 1 à 0 signifie que au temps $t = 0$, la probabilité de ne pas avoir subi l'évènement est 1 et que lorsque t tend vers l'infini, elle tend vers 0.

La fonction de risque est définie par la formule suivante $h(t) = \lim_{\Delta t \to 0^+} \frac{P(t \leq T \leq t+\Delta t \mid T \geq t)}{\Delta t}$. Elle représente la probabilité que pour un patient l'évènement survienne entre t et $t + \Delta t$, sachant que le patient était toujours à risque au temps t. Ces deux fonctions sont liées par la relation suivante $S(t) = \exp\left(-\int_0^t h(u)du\right)$.

2. Notion de censure

Comme évoqué en introduction de cette partie, l'une des caractéristiques des données d'étude de survenue d'un évènement est l'existence d'observations incomplètes. Les observations incomplètes peuvent être induites par des temps de suivis trop courts pour observer l'évènement ou des pertes de vue en raison de sortie d'étude (déménagement, refus de continuer à participer à une cohorte, etc.). Dans ce cas, on parle de censure à droite. Un délai est dit censuré à droite si le patient n'a pas subi l'évènement à la date de dernières nouvelles ou à la date de point. On suppose généralement qu'il y a indépendance entre la cause de censure et l'évènement étudié.

Notons, qu'il existe également la censure à gauche. Celle-ci est induite par le fait que le patient ait déjà subi l'évènement avant le démarrage de l'étude. Ainsi la date d'évènement est antérieure à la date d'origine. Ce n'est généralement pas le cas dans les études épidémiologiques puisque nous définissons les critères d'inclusion de façon à ne pas inclure les patients ayant déjà subi l'évènement d'intérêt.

Enfin, on peut également évoquer la censure par intervalle. Celle-ci est induite par l'observation des évènements, non pas à une date exacte, mais dans un intervalle de temps. Cette dernière est rarement prise en compte.

3. Temps de participation

L'analyse de survie tient compte du temps de participation à l'analyse et non pas seulement du fait que le patient ait atteint ou non l'évènement d'intérêt. Ce temps de participation, défini entre la date d'origine et la date de fin d'étude, dépend bien évidemment, de la définition de la « fin de l'étude ». En particulier, les données peuvent être analysées en intention de traiter ou en per protocol. Dans une analyse en intention de traiter, les patients sont étudiés dans le groupe de traitement où ils se trouvaient à la date d'origine. Par conséquent, on ignore les éventuels arrêts ou changements de traitement. Dans une analyse en per protocol, l'analyse sera finalement restreinte aux patients qui adhèrent au traitement qui leur a été attribué. A l'arrêt du traitement initial ou au moment du changement vers un autre traitement, le suivi sera alors censuré.

Ces définitions confèrent à l'analyse en intention de traiter un caractère plus conservateur. L'analyse en per protocol, semblant plus proche de ce qu'on observerait en vie réelle, est parfois préférée par les cliniciens et les patients. Pour autant, l'analyse en per protocol peut être dangereuse puisqu'elle ne garantit pas de conserver l'équilibre. En effet, en étudiant un sous-groupe de patients qui poursuivent le traitement, on sélectionne la population pouvant conduire à l'estimation d'un effet particulier. En pratique, dans une étude de supériorité où l'on cherche à démontrer qu'un traitement A est plus efficace qu'un traitement B, l'analyse en intention de traiter est recommandée (European Agency for the Evaluation of Medicinal Products, 2000). Dans une étude de non-infériorité ou d'équivalence, où l'on cherche à démontrer, respectivement, que A n'est pas moins bénéfique que B, ou que les traitements A et B sont équivalents, on pourra conduire une analyse en per protocol. Il est tout de même fortement recommandé de la compléter par une analyse en intention de traiter (Gupta, 2011). Il convient également de bien décrire les raisons de sorties d'étude dans l'analyse en per protocol. D'ailleurs, un nombre important d'exclusion devrait interpeller. Ceci étant dit, le choix n'est pas toujours simple et il est difficile de déterminer l'analyse qui s'approche le plus de la « bonne » évaluation. Une des solutions peut être de présenter les deux et discuter les différents résultats.

4. Estimation de la fonction de survie

4.1 Définition et notation

Couramment dans les analyses de survie, on estime la probabilité de ne pas avoir subi l'évènement. Cette probabilité est appelée fonction de survie et peut être représentée par ce qui s'appelle la courbe de survie. Cette courbe de survie peut être notamment estimée à partir de l'estimateur de Kaplan-Meier (Kaplan and Meier, 1958). L'estimateur de Kaplan-Meier est l'estimateur non paramétrique le plus fréquemment utilisé. Il découle du raisonnement suivant : ne pas avoir subi l'évènement au temps t, c'est ne pas l'avoir encore subi juste avant t et ne pas le subir en t.

Basé sur ce raisonnement et avec les notations suivantes :

- $t_1 < \cdots < t_j < \cdots < t_k$: les différents temps d'évènements observés
- d_j : nombre de patients subissant l'évènement au temps $t_j, j = 1, \dots, k$
- n_j : nombre de patients à risque au temps $t_j, j = 1, \dots, k$

On peut écrire l'estimateur de Kaplan-Meier de la probabilité de ne pas avoir subi l'évènement au moins jusqu'en t ainsi :

$$\hat{S}(t) = \prod_{j:t_j \leq t} \frac{n_j - d_j}{n_j}$$

Cette fonction est en escalier décroissante, constante entre deux temps d'évènements consécutifs. Au-delà de la dernière observation, l'effectif à risque étant nul, on ne peut pas la calculer. Elle n'est donc plus définie.

4.2 Utilisation et interprétation

Outre l'estimation de la probabilité de ne pas avoir subi l'évènement à un temps donné, l'estimateur de Kaplan-Meier permet de déterminer la médiane de survie. La médiane de survie est le délai t_M tel que $\hat{S}(t_M) = 0.5$. Si le recul de l'étude n'est pas assez important par rapport à l'incidence de l'évènement, il n'est pas possible d'estimer cette médiane.

A partir de cette estimation, on peut également déterminer la survie moyenne restreinte (en anglais « restricted mean survival time »), notée RMST (Irwin, 1949; Zhao and Tsiatis, 1999). Celle-ci est évaluée à un temps donné τ, on note alors $\mathrm{RMST}(\tau)$. Cela correspond à l'espérance de vie sans subir l'évènement jusqu'à un temps τ. Elle est déterminée par l'aire sous la courbe de survie restreinte au temps τ. Mathématiquement, on peut écrire de la manière suivante $\mu_\tau = \mathrm{RMST}(\tau) = \int_0^\tau S(t)dt$ où τ est le temps d'évaluation que l'on se donne. L'interprétation d'une telle quantité est la suivante : « Quand les patients sont suivis pendant une durée de τ, la survie moyenne s'élève à μ_τ. ». S'il n'y a pas d'observation censurée, on peut utiliser le délai moyen de survie $\mu_\infty = \int_0^\infty S(t)dt$. L'exemple proposé dans le document explicatif du package du logiciel R permettant de calculer ces quantités[1], montre que la durée moyenne de survie sur 10 ans de suivi chez des patients traités était de 7,15 ans. En d'autre termes, durant les dix ans de suivi, les patients traités ont perdu en moyenne 2,85 ans.

Classiquement, la durée de suivi médiane correspond à la médiane des longueurs des durées de suivis. Elle permet donc d'évaluer la longueur du suivi des patients. Son calcul peut être discuté au regard des sujets inclus dans le calcul : seulement des sujets n'ayant pas atteint le critère de jugement ou l'ensemble des patients. Cette valeur est parfois qualifié d'instable (Altman et al., 1995).

A partir de l'estimateur de Kaplan-Meier, on pourra également déterminer la durée de suivi. En utilisant le Kaplan-Meier inversé pour estimer la durée de suivi, on tiendra compte de la longueur du suivi et de sa complétude (Shuster, 1991). La courbe de Kaplan-Meier inversée est construite sur le même principe que la courbe de survie mais en inversant les statuts des patients. Le patient ayant subi l'évènement sera alors censuré à la date de l'évènement.

[1] *https://cran.r-project.org/web/packages/survRM2/vignettes/survRM2-vignette3-2.html*

4.3 Comparaison de deux groupes à partir des courbes de survie

La comparaison de courbes de survie peut être pertinente si l'on souhaite comparer la survie de deux groupes de patients. En particulier, lorsque l'on veut tester globalement l'égalité de deux distributions de temps de vie sans subir l'évènement, le test du Log-rank peut être utilisé (Mantel, 1966).

La RMST pourra également être utile pour comparer deux groupes (Zhao et al., 2016). En effet, on peut quantifier les années de vie gagnées sans subir l'évènement d'un groupe par rapport à l'autre. Cela revient à déterminer la différence entre les RMST de chaque groupe. Cette mesure fournit une information plus intuitive et plus parlante par rapport à une mesure de risque, notamment.

4.4 Pondération de la courbe de survie

En présence de facteurs de confusion et d'utilisation d'un score de propension, l'estimateur de Kaplan-Meier peut être utilisé si un appariement a été réalisé. En revanche, il n'est plus adapté dans le cadre d'une pondération., ainsi que les éléments de comparaison qui en découlent. En se basant sur la méthode de pondération par la probabilité inverse d'être traité, des auteurs ont proposé des estimateurs tenant compte des facteurs de confusion (Cole and Hernán, 2004; Xie and Liu, 2005). Cette estimation est basée sur le nombre pondéré de patients subissant l'évènement au temps t_j et sur le nombre pondéré de patients à risque au temps t_j. Ainsi, la construction de l'estimateur pondéré est une adaptation de l'estimateur usuel. De même, une version du test du Log-rank pondéré a été proposée (Xie and Liu, 2005).

5. Modèle à risques proportionnels de Cox

Dans les études de survie, on peut également chercher à analyser l'influence de variables explicatives sur le temps sans avoir subi l'évènement. Pour ce faire, un modèle de régression pourra être utilisé tenant compte des délais de survenue de l'évènement. Dans un tel contexte, le modèle le plus utilisé est le modèle de Cox ; il est décrit ci-dessous (Cox, 1972).

Le modèle de Cox est un modèle semi-paramétrique à risques proportionnels. Il permet d'établir une relation entre la fonction de risque associée à la survenue de l'évènement et les variables explicatives, notées Z. Le modèle s'écrit comme suit :

$$h(t, Z, \beta) = h_0(t)\exp(\beta^T Z)$$

où β est le vecteur de coefficient de régression et $h_0(t)$ la fonction de risque de base

Le modèle de Cox permet d'évaluer la modification du risque de base par la présence de tel ou tel facteur de confusion. On interprète le rapport des fonctions de risques appelé « hazard ratio » (HR), défini de la manière suivante :

$$HR = \frac{h(t, Z = 1, \beta)}{h(t, Z = 0, \beta)} = \frac{h_0(t)\exp(\beta.1)}{h_0(t)\exp(\beta.0)} = \exp(\beta)$$

Ainsi, le risque instantané d'un patient exposé ($Z = 1$) sera égal au risque instantané d'un patient non exposé ($Z = 0$) multiplié par HR.

5.1 Hypothèses du modèle de Cox

Le modèle de Cox repose sur deux hypothèses : l'hypothèse de proportionnalité des risques et l'hypothèse de log-linéarité.

La proportionnalité des risques signifie que les fonctions de risque pour les différentes modalités d'une variable sont proportionnelles et que leur rapport est constant au cours du temps. Plusieurs méthodes sont disponibles pour tester cette hypothèse mais aucune d'elles n'est parfaite. Sans consensus, des méthodes aussi bien graphiques qu'analytiques peuvent être utilisées. Parmi les méthodes graphiques, il est courant de tracer les courbes $\log[-\log(\hat{S}(t))]$ en fonction du temps et de vérifier que l'écart entre les courbes est constant au cours du temps. Les méthodes graphiques sont parfois critiquées pour leur subjectivité. L'étude du respect de cette hypothèse peut donc être complétée par une approche analytique, comme l'analyse des résidus de Schoenfeld (Schoenfeld, 1982) ou celle des résidus de martingale (Lin *et al.*, 1993).

Lorsqu'on introduit une variable explicative continue dans le modèle de Cox, l'augmentation d'une unité de cette variable est associée à HR égal à $\exp(\beta)$, quelle que soit sa valeur initiale. Pour vérifier que cette hypothèse, une solution consiste à transformer la variable quantitative en variable catégorielle (3 ou 4 classes). En utilisant cette variable catégorielle dans le modèle, on obtient plusieurs HR (2 ou 3) avec leurs intervalles de confiance. On détermine la différence entre le centre de chaque classe par rapport à la référence. Ensuite on calcule le HR correspondant à la variable quantitative continue. L'hypothèse de log-linéarité

est acceptable si le HR calculé pour la variable quantitative est contenu dans les intervalles de confiance obtenus avec la variable catégorielle.

5.2 Les modèles dérivés du modèle de Cox original

Dans un modèle de Cox, les variables explicatives sont généralement des variables mesurées à l'inclusion de patients et fixes jusqu'à la fin de l'étude. Dans certains cas, il est intéressant de prendre en compte l'évolution des valeurs des variables au cours du suivi. Les variables explicatives sont alors appelées dépendantes du temps. Le traitement statistique d'un tel modèle reste le même mais l'interprétation peut être moins aisée.

A l'instar d'un modèle de Cox multiple, un modèle de Cox pondéré peut être utilisé pour tenir compte des facteurs de confusion pour estimer l'effet d'une exposition sur la survenue d'un événement. En utilisant un modèle de Cox multiple, on mesurera un effet conditionnel alors qu'un effet marginal sera obtenu en utilisant le modèle de Cox pondéré. Dans ce travail, on utilisera un modèle de Cox pondéré utilisant des poids et un estimateur robuste de la variance adaptée à l'utilisation d'une pondération IPTW (Cole and Hernán, 2004). Tenant compte de la pondération, il consiste finalement en un modèle univarié dans lequel seule la variable relative à l'exposition est incluse.

Enfin, les modèles à fragilités sont utilisés pour analyser les données de survies dans les cas de regroupement (cluster, zone géographique, etc.). Il s'agit d'une extension du modèle de Cox dans lequel on ajoute une variable dite de fragilité, correspondant à un effet aléatoire, spécifique à chaque groupe. Ces modèles ont été utilisés pour tenir compte du caractère regroupé des données issues d'un appariement.

Au-delà de ces considérations méthodologiques usuelles, des développements spécifiques ont été apportés afin d'utiliser un appariement utilisant un score de propension dépendant du temps. Cet aspect du travail ayant constitué un des axes de la thèse, il est développé en chapitre 3 où nous abordons la question de l'évaluation de l'efficacité d'un traitement en vie réelle. En chapitre 4, nous étudions une question portant sur la comparaison d'efficacité de deux traitements. Dans le chapitre 5, il s'agit de comparer diverses utilisations de traitements. Enfin, dans le chapitre 6, nous discutons les points transversaux.

Chapitre 3

III. Efficacité d'un traitement en vie réelle : l'exemple des interférons-β et de l'acétate de glatiramère

A. Contexte **64**
1. Une méta-analyse tendant à montrer le bénéfice de ces deux traitements à long terme 64
2. Quatre études à la loupe 65

B. Une réflexion méthodologique étape par étape **69**
1. Population d'étude : date d'inclusion, groupe contrôle 69
2. Analyse statistique 70

C. Un score de propension dépendant du temps **71**
1. Principe de la méthode 72
2. Adaptation du score dépendant du temps au contexte de la SEP rémittente 74

A. Contexte

Les interférons-β et l'acétate de glatiramère sont deux traitements de première ligne de la SEP rémittente présentant des indications similaires. Leur indication large, associée à leur bon profil de tolérance, a conduit à les prescrire aux patients présentant une SEP active, peu sévère, c'est-à-dire un grand nombre de patients. Leur efficacité sur la réduction de la fréquence des poussées est démontrée, en revanche, leur efficacité à long terme semble plus controversée, notamment sur la progression du handicap. L'objectif de ce travail était donc d'évaluer l'effet des traitements de première ligne injectables sur la progression du handicap à long terme.

1. Une méta-analyse tendant à montrer le bénéfice de ces deux traitements à long terme

L'efficacité à long terme, notamment sur l'aggravation du handicap, a été explorée par plusieurs équipes en utilisant des données observationnelles. Les résultats obtenus sont variables. Par exemple, l'étude du délai d'atteinte d'un score EDSS 6, marquant la nécessité d'une aide unilatérale pour marcher 100 mètres, chez les traités et les non traités, a conduit à des HR variant de 0,14 à 1,30 (Figure 18), donc à des conclusions divergentes. Ces études ont été rassemblées dans une méta-analyse conduisant, par exemple, à un HR global de 0,56 (Intervalle de confiance à 95% = [0,46 ; 0,68]) concernant l'étude du délai d'atteinte d'un score EDSS 4 (Signori *et al.*, 2016). Pour autant, les études incluses dans cette méta-analyse présentent plusieurs biais qui ont été identifiés et décrits (Sormani and Bruzzi, 2014). Parmi les éléments identifiés ayant pu altérer le résultat, le biais d'indication, le biais de temps immortel et le choix du groupe contrôle ont été évoqués. L'auteur de la méta-analyse conclut d'ailleurs son article en indiquant que bien que les méta-analyses soient considérées comme fournissant un haut niveau de preuve, rassembler des résultats biaisés conduit à un résultat global biaisé.

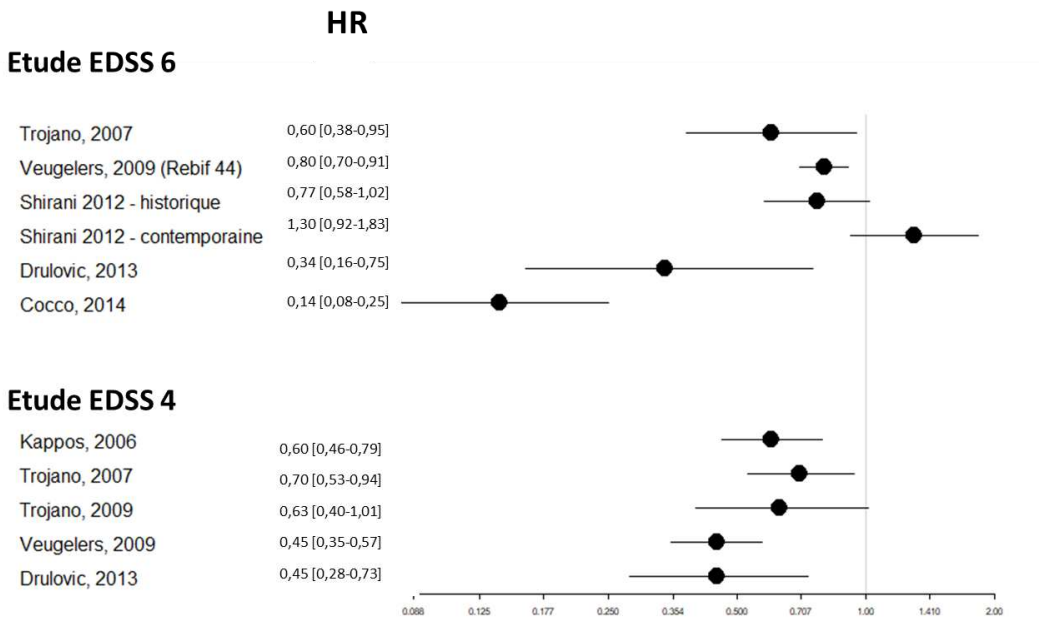


Figure 18: Etudes portant sur l'efficacité des traitements de première ligne injectables incluses dans la méta-analyse de Signori et al, 2016

2. Quatre études à la loupe

Parmi les études réalisées dans ce contexte, nous en avons sélectionné quatre qui avaient des schémas auxquels nous pouvions nous identifier. En effet, nous nous sommes concentrés sur les études observationnelles ayant pour objectif d'évaluer l'effet d'un ou des traitements de première ligne injectables à long terme (Trojano *et al.*, 2007; Shirani *et al.*, 2012; Drulovic *et al.*, 2013; Cocco *et al.*, 2015). Les autres études consistent en une comparaison avant et après initiation du traitement (Veugelers *et al.*, 2009), une extension d'essai thérapeutique (Kappos *et al.*, 2006*b*) et une étude portant uniquement sur des patients traités (Trojano *et al.*, 2009). Le Tableau 1 rassemble les éléments méthodologiques caractérisant ces quatre études (Trojano et al., 2007; Shirani et al., 2012; Drulovic et al., 2013; Cocco et al., 2015). Notons que l'un des articles présentait deux études (Shirani et al., 2012). L'une était basée sur une cohorte contrôle contemporaine, c'est-à-dire que les patients du groupe contrôle étaient sélectionnés durant la même période d'étude que la cohorte de patients du groupe traité. L'autre était basée sur une cohorte contrôle historique, c'est-à-dire que les patients du groupe contrôle étaient sélectionnés sur une période d'étude antérieure à la disponibilité des traitements. Dans ce tableau, nous avons intégré seulement la première analyse qui correspond au schéma de notre étude.

Certains choix méthodologiques réalisés dans ces études peuvent avoir biaisé les résultats. En particulier, on peut remarquer des différences dans le choix de la date d'inclusion. Dans un

contexte d'étude de comparaison de traitements (incluant traitement *versus* pas de traitement), cette date devrait correspondre à la décision thérapeutique. Pour le groupe traité, la réponse apparait triviale, la décision thérapeutique pouvant être assimilée à l'initiation du traitement. En revanche, pour le groupe contrôle, cela est moins évident. Dans les études précédemment conduites, on remarque différentes situations :

1. Une date dépendant du statut d'exposition : l'initiation du traitement pour les traités et la première visite au centre pour les non traités.
2. Une date commune : début de la SEP ou premier examen dans le centre expert.

Le premier choix peut être critiqué comme favorisant le traitement en raison du biais de temps immortel introduit en incluant les patients traités au démarrage du traitement et non à la date de première visite au centre, au même titre que les patients non traités. En effet, entre le premier examen au centre et la décision thérapeutique potentielle, les patients du groupe non traités peuvent réaliser des événements qui seront comptabilisés alors que cela est rendu impossible pour le groupe traité puisque non observés, d'où le terme de « temps immortel » (en référence à l'évènement de décès).

Le deuxième choix peut également conduire à un résultat biaisé puisque les caractéristiques des patients au début de la SEP ou à l'entrée dans le centre expert peuvent être différentes de celles qui conduisent à proposer le traitement. Ainsi, les deux groupes sont rendus comparables sur des éléments qui ne les caractériseront plus au moment de la décision thérapeutique.

Dans toutes ces études, les groupes contrôles correspondaient à des patients non traités. Dans un contexte de traitements largement prescrits et d'un critère de jugement à long terme, on peut s'interroger sur la pertinence d'un tel groupe. En effet, les patients restant non traités correspondent probablement à une population spécifique de patients présentant notamment, des facteurs de bon pronostique.

Enfin, l'équilibre des deux groupes a été obtenu en utilisant diverses méthodes, parfois un score de propension. Si cette méthode est très populaire de nos jours, il semble qu'elle questionne en pratique. On note notamment, que les variables ont pu être mesurées à des temps différents, avant comme après, la date d'inclusion. Depuis la réalisation de la première étude s'intéressant à cette question, les connaissances, notamment autour des scores de propension ont largement évolué. La littérature sur ce sujet est désormais plus large et plus accessible.

Ainsi, dans ce travail, nous avons cherché à mieux mesurer le moment de la décision thérapeutique. Nous avons également cherché à établir une population contrôle plus adaptée au contexte des maladies chroniques dans lequel l'évaluation de critères de jugement à long terme est requise.

Tableau 1 : Caractéristiques des études observationnelles portant sur l'efficacité des traitements de première ligne injectables

	Trojano, 2007	**Shirani, 2012**	**Drulovic, 2013**	**Cocco, 2015**
Critères inclusion/exclusion	SEP rémittente prise en charge dans deux centres italiens (Bari et Florence)	RRMS, plus de 18 ans, EDSS <6,5 Au moins 2 mesures d'EDSS entre entrée dans l'étude et fin de suivi	SEP rémittente prise en charge dans le centre expert de Belgrade, Serbie	SEP rémittente prise en charge dans le centre expert de Cagliari en Sardaigne
Période	Non communiqué	1995-2004	2004-2010	1985-2012
Inclusion	Groupe traité: initiation du traitement Groupe non traité: 1ère visite au centre	Eligibilité = 1ère visite au centre expert	Début de la SEP	Début de la SEP
Exposition	Interférons-β	Interférons-β	Interférons-β	Interférons-β et acétate de glatiramère
Méthode Equilibre	Pondération IPTW Exclusion des poids extrêmes	Ajustement sur le score de propension Exclusion des poids extrêmes	/	Ajustement sur le score de propension
Variables	Sexe, âge, durée de la SEP, nombre de poussées dans l'année précédent inclusion, EDSS en quintiles, interactions non détaillées	Sexe, âge, score EDSS, durée de SEP, SES, indice de comorbidités, taux annualisé de poussées, interactions de la durée de la SEP avec : EDSS ou taux annualisé de poussée ou âge	Sexe, âge, durée de SEP, EDSS, nombre de poussées dans l'année précédant l'initiation du traitement pour le groupe traité et l'admission au centre pour le groupe non traité	Sexe, âge, date de début de SEP, durée de SEP
Analyse	Modèle de Cox, avec censure à la date de dernière visite ou réalisation évènement Exclusion des patients réalisant l'évènement l'année suivant inclusion	Modèle de Cox dépendant du temps avec censure à la date de dernière visite ou réalisation évènement	Modèle de Cox ajusté *censuré après 7 ans de suivi*	Modèle de Cox dépendant du temps ajusté sur score de propension fixe
Résultat	***EDSS 4 : 0,70 [0,53 ; 0,94]*** ***EDSS 6 : 0.60 [0,38 ; 0,95]***	***EDSS 6 : 1,30 [0,92 ; 1,83]*** ***(ajusté variables du score)*** ***EDSS 6 : 1,34 [0,93 ; 1,92]***	***EDSS 4 : 0,45 [0,28 ; 0,73]*** ***EDSS 6 : 0.34 [0,16 ; 0,75]***	***EDSS 3 : 0,06 [0,04 ; 0,09]*** ***EDSS 6 : 0,14 [0,08 ; 0,25]***

B. Une réflexion méthodologique étape par étape

Au regard de la littérature et dans l'optique de mieux évaluer l'efficacité des traitements de première ligne injectables, plusieurs explorations ont été conduites. Nous revenons brièvement sur les différents choix méthodologiques envisagés avant d'exposer la méthodologie que nous avons finalement souhaité valoriser.

1. Population d'étude : date d'inclusion, groupe contrôle

Tout d'abord, nous avons cherché à définir une date d'inclusion adaptée, c'est-à-dire correspondant à la date de décision thérapeutique et qui serait la même pour les patients traités et les patients non traités. Plusieurs ensembles de critères d'inclusion ont été envisagés.

Dans un premier temps, nous avons exploré des critères définis à partir des critères d'éligibilité selon l'AMM de l'interféron-β. En ce sens, les patients étaient inclus dans l'étude s'ils étaient âgés de plus de 18 ans, avaient une SEP de forme rémittente, un score EDSS inférieur à 6.5 et étaient naïfs de tout traitement à l'inclusion. De plus, au moins deux poussées en 3 ans étaient requises pour être éligible au traitement. Concernant cette étude, la date d'inclusion dans l'étude correspondait à la date de deuxième poussée. A ce stade, si les patients étaient traités en dehors de l'AMM, ils étaient tout de même inclus. Ce dernier point présentait l'avantage de ne pas exclure de patients traités mais de déséquilibrer les critères puisqu'aucun témoin correspondant aux profils de patients traités hors « AMM » n'était inclus. De plus, à partir de ces critères, les patients y répondant avant la disponibilité des traitements étaient exclus. Ceci a donc été revu par la suite en proposant une règle pour réévaluer la date d'éligibilité, après la disponibilité des traitements. Cette règle reposait sur le nombre de poussées dans les trois ans précédant la date de disponibilité. Ce choix induisait en revanche un biais de sélection. En effet, parmi les patients dont la date d'éligibilité pouvait être réévaluer pouvaient présenter un meilleur état de santé que ceux qui avaient été exclus en raison d'une exposition à un autre traitement.

Dans tous les cas, cette sélection de patients, basée sur des critères d'AMM, présentait la limite majeure de ne pas refléter les pratiques réelles. Ces pratiques ont notamment été influencées par les connaissances acquises au cours du temps sur l'utilisation des traitements.

Les critères d'inclusion ont ensuite été élargis afin de se rapprocher des pratiques : avoir une SEP de forme rémittente et présenter une activité clinique. Les patients étaient inclus dès

la première poussée si les traitements étaient disponibles ou à la date de disponibilité des interférons-β. Comme précédemment, être naïf de tout traitement à l'inclusion était requis pour être inclus.

Qu'il s'agisse des critères AMM ou des critères « pratiques réelles », le schéma proposé conduisait à comparer des patients traités à des patients non traités. Comme évoqué précédemment, cela avait tendance à favoriser le traitement puisque les patients restant sans traitement ont probablement un profil particulier. Afin de mieux documenter cela, nous avons consulté les dossiers médicaux de 424 patients vus au centre expert de Rennes qui n'avaient jamais été traités au cours de leur suivi afin de mieux comprendre pourquoi ces patients n'étaient pas traités. Parmi les dossiers que nous avons pu consulter, 55% n'avait eu aucune proposition de traitement. La plupart du temps, la raison de non proposition était une maladie bénigne et la présentation de facteurs de bon pronostic. Pour 32% des patients, le traitement était proposé mais refusé par les patients en raison de réticence face aux traitements, de peur des injections ou des effets indésirables ou encore de préférence pour des médecines alternatives. Les autres patientes étaient des femmes ayant un désir de grossesse, donc non éligibles au traitement à ce moment-là. Enfin, des patients ont été perdus de vue après la proposition de traitement. Ce travail, nous a permis d'étudier l'impact de l'exclusion de la population pour laquelle aucun traitement n'avait été proposé.

Ces éléments montrent la complexité de la définition des critères d'inclusion dans l'évaluation en vie réelle, d'un traitement largement prescrit, en particulier pour la définition du groupe contrôle. En outre, les choix relatifs aux méthodes statistiques ont également fait l'objet d'une discussion complexe.

2. Analyse statistique

Dès le début de la thèse, l'utilisation d'un score de propension semblait évidente. En revanche, la manière de l'utiliser et l'impact de son utilisation sur le résultat ont constitué un choix moins trivial. Les méthodes d'appariement et de pondération ont par conséquent été explorées. Une première issue de ce travail aurait d'ailleurs pu porter sur la discussion de l'utilisation de telle ou telle méthode et de son impact sur le résultat. Ces questionnements ont finalement été abordés dans le chapitre 4 dans une étude comparative de deux traitements.

Nous nous sommes également interrogés sur la manière de tenir compte du caractère temporel de l'exposition. A la difficulté de déterminer la date d'inclusion, s'ajoutait le potentiel biais de temps immortel qui pouvait être introduit en incluant les patients à des dates différentes par exemple. Au regard des études précédemment conduites, nous nous sommes donc également questionnés sur l'utilisation d'un modèle tenant compte de la dépendance dans le temps du statut d'exposition. A l'instar de deux des trois études présentées précédemment, un modèle de Cox dépendant du temps a été utilisé pour étudier la progression du handicap. Dans ce contexte, les patients étaient inclus à une date d'éligibilité qui n'était pas nécessairement le démarrage du traitement. Ils étaient rendus comparables à cette date d'éligibilité en sachant leur futur groupe d'exposition. Cette exposition était considérée comme dépendante du temps. Cette différenciation entre la date d'inclusion à laquelle on réalisait l'équilibre et le vrai démarrage du traitement, nous a conduit à abandonner cette méthode d'analyse.

Ces éléments ainsi résumés ont fait l'objet d'analyses conduites dans les premières années de thèse qui ne sont jamais apparues satisfaisantes au regard des critiques précédemment évoquées et du choix du groupe contrôle (patients jamais traités) qui ne semble pas adéquat. Nous faisons le choix de ne pas montrer les résultats pour ne pas troubler la lecture. Cependant, notons que ces études ont fourni des résultats très fluctuants, modifiant même le sens des conclusions d'une étude à l'autre. Ces fluctuations ont d'ailleurs conforté l'idée que les variations observées dans la littérature pouvaient s'expliquer par les différences de critères d'inclusion et de méthodes. La complexité de cette question réside dans le caractère dépendant du temps de l'exposition. D'une part, il rend difficile la définition du moment de la décision thérapeutique dans le groupe n'ayant pas reçu le traitement. D'autre part, ce groupe contrôle, groupe de patient n'ayant pas reçu de traitement présentant une indication large et ce sur une longue période d'évaluation semble être un groupe inadapté aux caractéristiques particulières.

C. Un score de propension dépendant du temps

Dans ce contexte, au regard des réflexions menées à ce stade de l'étude, nous avons collaboré avec l'équipe SPHERE de Nantes : MethodS in Patients-centered outcomes and HEalth ResEarch, en particulier avec Yohann Foucher, maitre de conférence en biostatistique

et Rémi Lenain qui avait rejoint l'équipe SPHERE dans le cadre de son stage de Master 2 de Modélisation en pharmacologie clinique et épidémiologie. Son stage portait sur l'étude de la survie des patients transplantés rénaux en comparaison aux patients restant en dialyse. De nos jours, le recours à la greffe en cas d'insuffisance rénale est la plupart du temps privilégié. Ce stage s'inscrivait ainsi dans des problématiques similaires aux nôtres (traitements largement prescrits et critère de jugement à long terme). Ainsi, nos questionnements méthodologiques ont convergé. Nous avons alors pu bénéficier de leurs connaissances sur le plan méthodologique pour développer ce travail. Leurs recherches préalables les avaient orienté vers une méthodologie, originalement proposée par Lu, consistant en un appariement sur un score de propension dépendant du temps (Lu, 2005).

Bien que la méthode soit présentée dans l'article que nous proposons en fin de chapitre, il nous semble important de la décrire dans cette première partie et d'expliciter les choix que nous avons réalisés pour obtenir la version du travail que nous sommes actuellement prêts à publier.

1. Principe de la méthode

Sur le même principe qu'un appariement sur un score de propension conventionnel, le but est de faire correspondre un patient traité avec un autre patient qui présente une chance similaire de recevoir le traitement mais pour autant non traité au moment de l'initiation du traitement du patient traité. La principale différence réside donc dans cette notion de temps, et est la valeur ajoutée de la méthode proposée par Lu (Lu, 2005). La procédure est divisée en deux étapes.

La première étape consiste à estimer la probabilité de recevoir le traitement au cours du temps. Cette première étape fournit alors le score de propension. Cette valeur peut être appréhendée comme un score de propension classique. En revanche, au lieu d'être estimé au démarrage du traitement uniquement, il est évalué à partir d'une date donnée, comme le début de la maladie par exemple, jusqu'au démarrage du traitement. Le score est « révisé » en fonction des caractéristiques du patient à chaque temps d'évaluation entre ces deux dates d'où l'appellation dépendant du temps.

Le score de propension ainsi établi, la seconde étape consiste en un appariement sur cette valeur. On cherche alors à apparier un patient traité, au moment du démarrage de son traitement

avec un patient qui à ce moment-là n'est pas traité mais présente un score de propension similaire.

Dans la Figure 19, nous proposons une schématisation de l'étape d'appariement. Sur la figure, les valeurs de score de propension de trois patients ont été représentées du temps 0 (entrée dans l'étude) au dernier temps d'évaluation disponible. A chaque temps d'évaluation où les données permettant de déterminer le score de propension sont mises à jour, le score de propension évolue. Le patient représenté en bleu initie le traitement au temps 2800 (correspondant au nombre de jours depuis le début de la SEP), marqué par un point bleu. A ce temps, un patient pas encore traité mais présentant une valeur de score de propension équivalente, selon la tolérance maximale que nous nous sommes accordés, est recherché. Dans l'exemple, le patient représenté en orange constitue un bon appariement, contrairement au patient représenté en jaune dont la valeur du score de propension est trop basse.

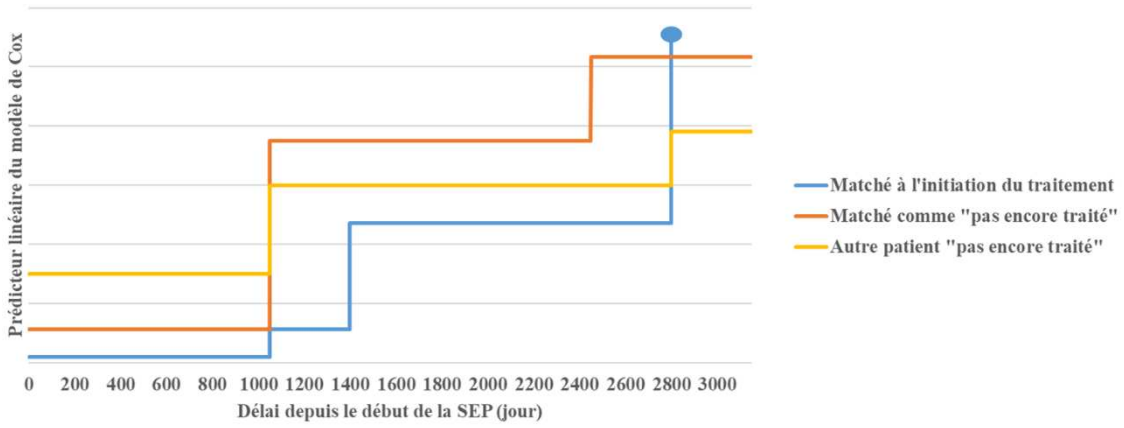


Figure 19 : Schématisation de l'appariement sur le score de propension dépendant du temps

Dans la suite, et pour simplifier le discours, on considèrera que la date d'entrée est le début de maladie. Dans notre travail, nous avons utilisé le début de la SEP, une date commune et généralement bien définie. Pour d'autres pathologies ou dans d'autres contextes, on pourrait utiliser une autre date. Par exemple, dans son travail sur la transplantation, Rémi Lenain avait utilisé la date d'inscription sur la liste d'attente de greffe.

Ceci étant défini, la première étape consiste alors à modéliser le temps entre le début de la maladie et la première initiation du traitement, en utilisant un modèle de Cox dépendant du temps. Les variables dépendantes du temps sont mises à jour au cours du temps. Les mises à

jour dépendent des temps d'évaluation des données. Mathématiquement, le modèle de Cox dépendant du temps peut se définir par la formule suivante : $h(t\,|L(t), Z) = h_0(t)\exp(\beta_Z Z + \beta_L L(t))$ où Z sont les variables dont les valeurs sont fixées dans le temps (par exemple, le sexe) et $L(t)$ sont les variables dépendantes du temps (par exemple l'évaluation du handicap à chaque visite). Comme la fonction de risque de base $h_0(t)$ est constante à chaque temps, le score de propension correspond simplement au prédicteur linéaire $\beta_Z Z + \beta_L L(t)$.

A chaque temps t auquel un patient devient traité, on recherche à associer au patient nouvellement traité, un patient présentant une valeur de prédicteur linéaire similaire mais qui au temps t n'a pas encore été traité. Un tel appariement conduit finalement à construire deux groupes : un groupe de patients traités et un groupe de patients « pas encore traités ». En effet, au moment de l'appariement, le patient « contrôle » n'est pas traité mais il se peut qu'il le devienne au cours du suivi. Dans cette méthode, si un patient traité est apparié comme « pas encore traité » à un moment antérieur à son propre traitement, il / elle est exclu.e de l'ensemble des patients traités et ne changera pas de statut.

Le principe de la méthode ainsi explicité, nous revenons sur les choix que nous avons réalisés dans l'application de cette méthode à notre problématique.

2. Adaptation du score dépendant du temps au contexte de la SEP rémittente

Comme évoqué précédemment, le score de propension dépendant du temps a été estimé en utilisant un modèle de Cox dépendant du temps, modélisant le délai entre les premiers symptômes de SEP et l'initiation du premier traitement par interférons-β ou acétate de glatiramère. Les variables dépendantes du temps incluses dans ce modèle étaient mises à jour à chaque fois qu'une mesure EDSS était réalisée. Dans ce travail, nous avons suivi la recommandation consistant à intégrer les variables qui étaient associées au critère de jugement, indépendamment de leur association avec l'exposition. Ainsi, chaque modèle de score de propension était spécifique au critère de jugement d'intérêt. Les variables potentielles étaient les suivantes : le sexe, l'âge, l'année de début de SEP, le nombre de poussées dans l'année précédant la visite, le score EDSS, la variation du score depuis la dernière visite et le délai depuis la dernière visite au centre expert. Les variables relatives au nombre de poussées, à

l'EDSS, à la variation d'EDSS et le délai depuis la dernière visite étaient considérées comme variables dépendantes du temps.

Une fois le score de propension dépendant du temps estimé, nous avons utilisé un appariement tel que recommandé dans la littérature (Austin, 2014). Il s'agissait d'un appariement 1:1, aléatoire, sans remplacement avec un caliper de 0.2 de l'écart-type du score de propension. Comme dans la méthode originale, lorsqu'un patient était apparié comme « pas encore traité », il était alors exclu du groupe des patients pouvant être appariés comme traités même s'il le devenait plus tard.

Ce n'est que tardivement que nous avons eu à l'esprit d'un élément incorrect dans la modélisation. Le début de la SEP ayant pu survenir avant la période de disponibilité des traitements, plusieurs patients étaient appariés comme « pas encore traités », avant la disponibilité des interférons-β donc sans avoir la chance de pouvoir recevoir le traitement. Ceci est contraire à l'hypothèse de positivité que nous avons évoquée dans le chapitre 2. Pour respecter celle-ci, il est nécessaire que les patients aient tous une chance d'être traités. Ainsi, nous avons ajouté une contrainte dans la méthode d'appariement afin que seuls les temps après le début clinique de la SEP et la disponibilité du traitement en France (à savoir le 30 novembre 1995) soient pris en compte. Ainsi, par exemple, sur la Figure 20, le patient 1 ne pouvait être apparié qu'à partir du 1er décembre 1995, à la suite de la disponibilité de l'interféron-β alors que l'on pouvait d'emblée apparier le patient 2.

Patient 1

Identifiant	Date début	Date fin	Inclusion
Patient_1	15/06/1992	01/03/1995	0
Patient_1	02/03/1994	30/11/1995	0
Patient_1	01/12/1995	24/05/1996	1
Patient_1	25/05/1996	06/06/1997	1

Début de SEP : 15/06/1992,
Trois dates de visites ont été enregistrées :
01/03/1995, 24/05/1996, 06/06/1997

Patient 2

Identifiant	Date début	Date fin	Inclusion
Patient_2	15/06/2002	01/03/2003	1
Patient_2	02/03/2003	15/10/2004	1
Patient_2	16/10/2004	24/05/2006	1

Début de SEP : 15/06/2002,
Trois dates de visites ont été enregistrées :
01/03/2003, 15/10/2004, 24/05/2006

Figure 20 : Contrainte d'appariement sur le score de propension dépendant du temps

Il nous semble pertinent de montrer le résultat obtenu avant cette contrainte. En effet, cette première analyse a conduit à montrer que, sur 15 ans, le temps moyen avant d'atteindre un score EDSS de 3 était significativement retardé de 1,6 ans (Intervalle de confiance à 95% = [0.40; 2.83]) chez les patients traités par l'interféron-β ou l'acétate de glatiramère. En se référant à

l'article proposé ci-après, on constate que cette contrainte, pourtant primordiale et oubliée à tort, a largement modifié le résultat final.

Ce travail a été soumis au journal « Pharmacoepidemiology and Drug Safety » dans la version proposée ci-après. Nous sommes actuellement en révision depuis le 6 septembre 2019.

Time-dependent propensity score matching: a novel method that warrants consideration in comparative effectiveness studies

Authors:

Mathilde Lefort[(1)(2)], Yohann Foucher[(3)(4)], Remi Lenain[(3)(5)], Sandra Vukusic[(6-9)], Gilles Edan[(2)(10)], and Emmanuelle Leray[(1)(2)(11)]

Affiliations:

[(1)] *Rennes University, EHESP, REPERES – EA 7449, F-35000 Rennes, France*

[(2)] *Univ Rennes, CHU Rennes, Inserm, CIC 1414 [(Centre d'Investigation Clinique de Rennes)], F-35000 Rennes, France*

[(3)] *INSERM UMR 1246 – SPHERE, Nantes University, Tours University, Nantes, France*

[(4)] *Nantes University Hospital, Nantes, France*

[(5)] *Lille University Hospital, Lille, France*

[(6)] *Hospices Civils de Lyon, Service de Neurologie, sclérose en plaques, pathologies de la myéline et neuro-inflammation, F-69677 Bron, France*

[(7)] *Observatoire Français de la Sclérose en Plaques, Centre de Recherche en Neurosciences de Lyon, INSERM 1028 et CNRS UMR 5292, F-69003 Lyon, France*

[(8)] *Université de Lyon, Université Claude Bernard Lyon 1, F-69000 Lyon, France*

[(9)] *Eugène Devic EDMUS Foundation against multiple sclerosis, state-approved foundation, F-69677 Bron, France*

[(10)] *Department of Neurology, CHU Pontchaillou, Rennes, France*

[(11)] *METIS Department, EHESP French school of public health, Rennes, France*

Abstract

Purpose: Studies of treatment effectiveness need to take into account aspects relating to the control of bias. The long-term effect of beta-interferon and glatiramer acetate on multiple sclerosis (MS) disability progression has resulted in controversial results, probably due to a lack of appropriate control of such biases. In particular, the time of the therapeutic decision is difficult to define when the controls are not treated.

Methods: This retrospective observational study was based on a series of patients from the MS expert center in Rennes, France. We used a time-dependent propensity score defined as the linear predictor of a Cox model estimating the hazard of being treated at each time from MS onset. The matching procedure resulted in two groups: patients matched as treated and as not yet treated. The restricted mean times (RMST) to reach a moderate level of disability or worsening of the disability were compared between the two groups in an intention-to-treat analysis.

Results: Of the 2,383 patients included in the study, 556 were matched as treated. The matching procedure provided a good balance of both the time-fixed and the time-dependent covariates. A slight difference was observed for the time to reach a moderate level of disability, in favor of the "not yet treated" group (difference in the RMST: -0.62 [-0.91; -0.33]) while no difference was found in terms of worsening of the disability (-0.03 [-0.24; 0.33]).

Conclusion: This unexpected result is probably due to unmeasured confounders. However, this time-dependent PS warrants consideration in long-term effectiveness studies.

Keywords: Propensity score, indication bias, effectiveness, multiple sclerosis

Key points

- Despite the increased use of propensity scores, there is often still an important methodological issue related to indication bias.
- The time of the therapeutic decision, i.e. the baseline in comparative effectiveness study, is hard to identify when the controls are untreated.
- Time-fixed propensity scores are often used while confounders are time-dependent.
- The time-dependent propensity score is a useful tool that warrants consideration in treatment effectiveness studies in a chronic disease context.

Abbreviation:

MS: multiple sclerosis

RRMS: relapsing-remitting multiple sclerosis

β-IFN: beta-interferon

EDMUS: European database for multiple sclerosis

EDSS: expanded disability status scale

GA: glatiramer acetate

ITT: intention to treat

PS: propensity score

RMST: restricted mean survival time

SD: standard deviation

SMD: standardized mean differences

Introduction

Comparative effectiveness research allows treatments in real-life conditions to be studied when randomized clinical trials are lacking or not feasible. It needs to take into account specific issues related to the control of bias. Treatment options for multiple sclerosis (MS) have become more diverse and more complex over the past twenty years. This context necessitates up-to-date and rigorous assessment of the effectiveness of such treatments.

MS is a chronic disease that affects the central nervous system. This lifelong disease often starts in early adulthood, and it affects more women than men [1]. Relapsing-remitting MS (RRMS), characterized by the occurrence of relapses [2], is the most common phenotype. In the long run, the symptoms accumulate and lead to irreversible disability. Thus, most patients ultimately convert to secondary progressive MS.

The first available MS disease-modifying agents, beta-interferon (β-IFN) and glatiramer acetate (GA), were approved in France [3] in November 1995 and January 2002, respectively [4]. They are safe and they are still often prescribed. Their efficacy compared to placebo in the short- and medium-term, especially in terms of reducing relapses, has been shown in randomized clinical trials as well as by observational studies [5,6]. Nevertheless, their long-term effects remain less clear.

The studies comparing treated versus treatment-naive patients have mostly demonstrated a beneficial effect of β-IFN and GA, as concluded in a recent meta-analysis [7]. Very different results, however, have been obtained with hazard ratios for the time to having a need for a walking aid ranging from 0.14 to 1.30. In observational studies, results can be skewed by several biases [8], in particular, indication bias, resulting in non-comparable characteristics of the treated versus the control patients. The methods to control for bias have evolved over the past several decades. In particular, propensity scores (PS) are a commonly used solution to control for indication bias. PS aim to mimic a randomized clinical trial [9] by balancing the observed confounding factors at the time of the therapeutic decision. Nevertheless, this time is difficult to define when the controls are not treated. Indeed, it corresponds to the theoretical and unobserved time at which an untreated patient would have been treated. This time is clearly identified in clinical trials because it corresponds to the randomization.

The use of a time-dependent PS may solve this issue, as it enables control patients to be identified who are comparable to the treated

patients when the latter start treatment. In the present study, we propose to adapt a matching method originally devised by Lu [10] to assess the effectiveness of β-IFN and GA in MS.

Materials & Methods

Study population

This retrospective study was based on a series of patients from the MS expert center in Rennes, France [11]. Patients who are monitored by private neurologists can schedule consultations at expert centers, when needed. In Rennes, data have been systematically compiled since 1976 in the European Database for Multiple Sclerosis (EDMUS) software [12]. This information is updated at each consultation [13], and in May 2017 it comprised data for a total of 5,150 patients. Information on treatments and relapses are collected at each consultation. MS-related disability is measured using the Expanded Disability Status Scale (EDSS), which ranges from 0 (normal) to 10 (death due to MS). [16] All of the enrolled patients provided informed consent for participation and for their data to be stored and used for research. Data confidentiality and safety were ensured according to the recommendations of the French "Commission Nationale Informatique et Libertés" (CNIL).

All of the patients with MS according to the Poser [14] or the McDonald [15] criteria; a relapsing MS onset before December 31st 2005; and who were treatment-naive before 30 November 1995 (the date when β-IFN was approved in France), did not reach an irreversible EDSS score of 3, and did not convert to secondary progressive MS before November 30th 1995, were included in the present study.

Exposure

Over the study period, four preparations of β-IFN were available: interferon beta-1b (Betaseron®, 250 µg), interferon beta-1a (Avonex®, 30 µg and Rebif®, 22 µg and 44 µg) and one preparation of GA (Copaxone®, 20 mg). As all these preparations have similar indications, all of the patients exposed at least one day were considered in a single group.

Outcomes

Three outcomes were defined as follows: (1) the time from matching to reaching an irreversible EDSS score of 3, which corresponds to moderate disability [16], (2) the time from matching to the first worsening of the EDSS, and (3) the number of relapses within 5 and 10 years following matching.

An EDSS score was considered to be irreversible when it persisted for at least 6 months, and until the last consultation. A

relapse was defined as the occurrence of new symptoms or the worsening of symptoms that had occurred previously, for at least 24 hours. Worsening of the EDSS was defined as an increase by 1½ points when the initial EDSS score was 0, by 1 point when the initial score was comprised between 1 and 5½, or ½ a point if the initial score was more than 5½.

Time-dependent propensity

We used the method initially proposed by Lu [10]. This method consisted of matching treated patients with untreated patients who had the same chance of being treated at the same time. This procedure is divided into two steps. The first step consisted of modeling the time from the disease onset to the first prescription of β-IFN or GA using a cause-specific updated Cox regression. The times were right-censored at the earliest of the four following events: initiation of a treatment other than β-IFN or GA, conversion to secondary progressive MS, assignment of an irreversible EDSS score of 3, or the most recent clinical information. The time-dependent PS was defined as the linear predictor of this Cox model with time-dependent covariates.

As recommended [17], this model was adjusted for variables associated with the outcome. Thus, a PS model was specified for each of the three outcomes (Table S1).

The following variables were considered: sex, age, the year of MS clinical onset, the number of relapses in the past 12 months, the EDSS score, the EDSS variation since the last consultation, and the delay since the first consultation at the MS expert center. The number of relapses in the past 12 months, the EDSS score, the EDSS variation, and the delay since the first consultation were time-dependent covariates. The number of relapses was divided into 5 modalities: zero, one, two, three, and more than three. The EDSS score was categorized as: ≤ 1, 2-3, and ≥ 4. The EDSS variation was categorized as: (1) improvement if there was a decrease by 1 point when the initial EDSS was < 5.5, or 0.5 a point when the initial score was ≥ 5.5; (2) worsening when there was an increase by 1 point if the initial EDSS was < 5.5, or 0.5 a point when the initial score was ≥ 5.5, and (3) stable in all other cases. The delay since the first consultation at the MS expert center was categorized into 3 modalities: (1) none, if no consultation, (2) recent, if the delay from the first consultation was less than 3 years, and (3) old, if it was more than 3 years.

Matching procedure

The second step of the procedure consisted of matching treated patients at the time of treatment start with untreated patients who had a comparable PS value at that time.

According to the positivity assumption, all of the patients should have a not-null probability of being treated in order to be matched. Thus, only time points after both MS clinical onset and treatment availability in France (i.e., 30 November 1995) were considered. As patients from the untreated group can initiate treatment later in the disease course, this group is referred to as the "not yet treated" group. Nearest neighbor matching at random, with a caliper of 0.2 of the standard deviation without replacement, was used [18]. If a treated patient was matched as "not yet treated" at a time before his own treatment initiation, he was excluded from the set of treated patients and could not be matched as treated later. To check the balance in the matched cohort, standardized mean differences (SMD) were computed, with 10% being considered as an acceptable difference [19].

Baseline was the time of matching, i.e. the decision of treatment start for patients matched as treated and the corresponding time for patients matched as "not yet treated".

Main analyses from the matched cohort

The median follow-up duration from matching was estimated using the reverse Kaplan-Meier method [20]. The distributions of time-to-events were estimated using the Kaplan-Meier estimator. The area under this survival curve up to 10 years represented the restricted mean survival time (RMST) [21]. This interval of 10 years was determined according to the follow-up duration available in the dataset. Moreover, the numbers of relapses within 5- and 10-year following matching were analyzed using a negative binomial regression, with the log-transformed observation time as offset.

An intention to treat (ITT) analysis was performed (all the subjects were considered regardless the post-matching events such as treatment modification). Moreover, bootstrapping (1,000 samples) was used to handle the randomness of the matching and to obtain confidence intervals. The analyses were performed using R software (R 3.4.0).

Results

A total of 2,383 patients were included in the present study (Figure 1), of whom 933 were treated with β-IFN or GA. Overall, the matching procedure led to the exclusion of 1,271 patients, although the matched population (including 556 of the 933 treated patients) approached the distribution of the initial cohort in terms of sex, age, and the period of MS clinical onset (Table 1).

Characteristics of the matched dataset

As shown by the standardized mean differences in Table 2, the matching procedure resulted in a good balance of both

the time-fixed and the time-dependent covariates. The mean age at matching was 34 ± 9.9 years, and 77% were women. The median follow-up duration was 9.5 years (95% CI from 9.1 to 10.0). It was 9.3 years (95% CI from 8.6 to 10.0) and 9.8 years (95% CI from 9.2 to 10.5) for the β-IFN/GA-treated patients and the "not yet treated" patients, respectively.

Comparisons of the outcomes

The survival curves for the probabilities of not reaching an irreversible EDSS 3 and of not having a worsening of the EDSS are presented in Figures 2A and 2B. As summarized in Table 3, no difference was found for worsening of the disability, and a slight difference was observed for the time to reach EDSS 3, albeit in favor of the "not yet treated" group.

Analysis of the number of relapses indicated that the incident rate ratio was 1.23 [1.07; 1.40] over 5 years and 1.16 [1.01; 1.32] over 10 years, in favor of the "not yet treated" group.

Therapeutic sequences in the two groups

In the treated group, β-IFN/GA was initiated after a median time of 3.4 years (interquartile range: 1.2 - 7.9) after MS clinical onset, and its median duration was 2.0 years (interquartile range: 1.0 – 4.6). The reason for stopping treatment (known for 364 of the patients) was inefficacy for 36% and intolerance for 37%. For others PS models, the balance was also good (Table S2 to S4). As shown in Figure 3A, after the initial exposure to β-IFN or GA, 22% of the patients in the treated group switched to another first-line treatment and 12% switched to a second or third-line treatment. Of the 556 patients from the "not yet treated" group, 259 were treated with β-IFN/GA later in the disease course. The β-IFN/GA-treatment-free survival is presented in Figure S1. The cumulative probability of being treated was 18% (95% CI from 15% to 22%) at one year and 47% (95% CI from 42% to 51%) at 5 years. The RMST over 10 years showed that treatment was delayed by 5.7 years (95% CI from 5.3 to 6.1) for the "not yet treated" patients. Moreover, patients from the "not yet treated" group could have received other treatments, as shown in Figure 3B.

Discussion

In this study, we investigated how to assess long-term treatment effectiveness in real-life settings. As discussed previously [8], this issue entails considerable challenges, such as the choice of the appropriate control group. The novelty of the present study was the use of a time-dependent PS to compare patients who had received treatment to patients who had not yet initiated treatment

but who had the same probability of being treated at that time.

When applied to MS, this method yielded results that were in accordance with previous studies [22,23] but that were nonetheless unexpected as we were not able to show any benefit of treatment. In our opinion, this is due to the fact that our study still suffered from unmeasured confounding. Indeed, we were not able to balance groups based on MRI data due to a high rate of missing values, although this is nowadays known to be a central prognostic parameter [24].

Moreover, it should be kept in mind that the inclusion period was 1996-2006 and hence corresponded to outdated therapeutic practices of French neurologists. Indeed, the patients were treated after a mean MS duration of 5 years, while the current recommendations are to treat as soon as possible after MS onset[25]. This delay could also be related to the fact that the first symptoms of MS (MS onset) do not always lead to a consultation with a neurologist, and a delay in the diagnosis typically results in a delay in access to MS-specific care.

In addition, the number of therapeutic options has increased in recent years. Moreover, as indicated by randomized clinical trials in particular, most of the new drugs have been shown to exhibit comparable or higher efficacies than β-IFN/GA.

That being said, in our opinion, the methodology is of interest. Therefore, rather than focusing on the results, we believe that there is more merit in discussion of the method.

The use of a time-fixed PS (i.e., classical PS) to compare treated patients to never treated ones would have been an intuitive and well-known methodology, but the results can be biased. Indeed, on the one hand, patients who were never treated are probably patients who did not need treatment due to a benign course. On the other hand, patients who were never treated could be patients who had no chance of being treated, either because treatment was not available (such as patients from a historical control cohort) or because they were not eligible for treatment (such as secondary progressive MS patients). Use of the time-dependent PS implies that patients can be matched as untreated although they could be treated later in the disease course (here, 259 patients), leading to the control group being referred to as “not yet treated”. Thus, this methodology offers a design where the controls are patients with a delayed treatment. This appears to be suitable in many research contexts and corresponds better to the care management of chronic diseases in the real-life,

especially when assessing treatments that are largely prescribed over time.

The previously mentioned meta-analysis [7] pooled data from 10 observational studies and 4 long-term extensions of randomized clinical trials. The results showed a beneficial effect of β-IFN or GA, with a hazard ratio of 0.56 (95% CI from 0.46 to 0.68) for the time to reach irreversible EDSS 4. However, it was based on different designs and methods that are subject to varying degrees to potential biases [8], especially indication bias.

For instance, in the first publication published in 2007, the patients were included at different time points according to their treatment status, since the treated patients were included at treatment initiation while the control patients were included at the first consultation at the MS expert center [26]. This difference could have biased the results and favored the treatment since events that occurred from the first consultation to treatment initiation were only excluded from the analysis for the treated group, leading to an immortal time bias [8,22]. When an attempt at correction was carried out [22], the conclusion changed, with a new HR estimate of 2.02 instead of 0.73.

Moreover, the choice of the baseline date is often questionable. Indeed, the time-points for patient inclusion differed between studies, and they did not correspond to the therapeutic decision: for instance, inclusion could be at MS onset [27,28] or at the first consultation at the MS expert center [23]. In our opinion, the patient characteristics could have changed between these baseline dates and the time of therapeutic decision. The response to treatment could have been impacted by events such as relapses or the disability becoming worse before treatment initiation. In addition, the timing when potential confounders are assessed is important. Several situations have been observed in the literature: assessment at different times [26], not necessary at baseline, and sometimes after the treatment start. Such practices have been shown to lead to misspecification in the PS model [29]. Moreover, randomized clinical trials do not include parameters measured after baseline as covariates.

The study period may have also impacted the results. For instance, a Canadian study [23] used two different control groups: a contemporary control cohort (July 1995 – December 2004) and a historical control cohort (April 1985 – June 1995). Use of a historical control cohort implies that the comparison was performed between two different periods, as the treated patients were compared to the patients followed-up before treatment availability and who were hence, by definition, not treated. This type of design could lead to biased results [30,31] due to changes over time, for instance in diagnostic criteria, disease progression,

standard of care practices, or the quality of the data collection [32].

To conclude, the most important limitation of the present work is residual confounding. Using it with more comprehensive data, such as MRI in MS, would limit the issues due to unmeasured confounders. Moreover, due to its monocentric design, our study probably also suffers from a lack of power and representativeness. However, we do think that this innovative methodology based on time-dependent PS warrants consideration in long-term effectiveness studies. It would be interesting to see it applied to additional datasets from different contexts to better appreciate its strengths and potential limitations.

Acknowledgments

The authors wish to thank Damien Le Port (Rennes Hospital, France) for the data collection and the quality control, Prof Emmanuel Oger for his comments on the manuscript, and the “Observatoire Français de la Sclérose en Plaques” (OFSEP), which is supported by a grant provided by the French government and administered by the "Agence Nationale de la Recherche," within the framework of the "Investments for the Future" program, as reference ANR-10-COHO-002, by the Eugène Devic EDMUS Foundation against multiple sclerosis and by the ARSEP Foundation.

Funding

This work was part of Mathilde Lefort’s Ph.D., which is funded through an unconditional donation from Roche SAS, without any link to the scientific contents of the work.

Conflict of Interest Statement

M. Lefort reports travel grants from Roche SAS. Y. Foucher reports speaking honoraria from Biogen and Sanofi
R. Lenain has nothing to disclose. S. Vukusic reports consulting and lecturing fees, travel grants, and unconditional research support from Biogen, Celgène, Geneuro, Genzyme, MedDay, Merck Serono, Novartis, Roche, Sanofi Aventis, and Teva Pharma. G. Edan reports personal honoraria for lectures or consulting from Bayer, Biogen, LFB, Merck, Novartis, Roche, Sanofi. Research support from Bayer, Biogen, Genzyme, Merck, Novartis, Roche and Teva Pharma. E. Leray reports consulting and lecture fees or travel grants from Biogen, Genzyme, MedDay Pharmaceuticals, Merck Serono Novartis, and Roche

References

1. Filippi M, Or AB-, Piehl F, *et al.* Multiple sclerosis. *Nat Rev* 2018; **4**: 1–27.

2. Lublin FD, Reingold SC, Cohen JA, *et al.* Defining the clinical course of multiple sclerosis The 2013 revisions. 2014.

3. HAS. BETAFERON 250 - Transparency committee. 2010.

4. HAS. Copaxone - Avis de la commission. 2002.

5. Jacobs L, Beck R, Simon J, *et al.* Intramuscular interferon Beta-1atherapy initiated during a first demyelinating event in multiple sclerosis. *J Med* 2000.

6. Comi G, Filippi M, Barkhof F, *et al.* Effect of early interferon treatment on conversion to definite multiple sclerosis: A randomised study. *Lancet* 2001; **357**: 1576–1582. doi:10.1016/S0140-6736(00)04725-5.

7. Signori A, Gallo F, Bovis F, Di Tullio N, Maietta I, Sormani MP. Long-term impact of interferon or Glatiramer acetate in multiple sclerosis: A systematic review and meta-analysis. *Mult Scler Relat Disord* 2016; **6**: 57–63. doi:10.1016/j.msard.2016.01.007.

8. Sormani MP, Bruzzi P. Can we measure long-term treatment effects in multiple sclerosis? *Nat Rev Neurol* 2014; **11**: 176–182. doi:10.1038/nrneurol.2014.237.

9. Austin PC. A tutorial and case study in propensity score analysis: An application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate Behav Res* 2011; **46**: 119–151. doi:10.1080/00273171.2011.540480 .

10. Lu B. Propensity score matching with time-dependent covariates. *Biometrics* 2005; **61**: 721–728. doi:10.1111/j.1541-0420.2005.00356.x.

11. Leray E, Yaouanq J, Le Page E, *et al.* Evidence for a two-stage disability progression in multiple sclerosis. *Brain* 2010; **133**: 1900–1913. doi:10.1093/brain/awq076.

12. Confavreux C, Compston DAS, Hommes R, McDonald WI, Thompson AJ. EDMUS, a European database for multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1992; **55**: 671–676.

13. Vukusic S, Casey R, Rollot F, *et al.* Observatoire Français de la Sclérose en Plaques (OFSEP): A unique multimodal nationwide MS registry in France. *Mult Scler J* 2018: 1–5. doi:10.1177/1352458518815602.

14. Poser CM, Paty DW, Scheinberg L, *et al.* New Diagnostic Criteria for Multiple Sclerosis : Guidelines for Research Protocols. *Ann Neurol* 1983; **13**: 227–231.

15. Mcdonald WI, Compston A, Edan G, *et al.* Recommended Diagnostic Criteria for Multiple Sclerosis : Guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. 2001: 121–127.

16. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; **33**: 1444–1452. doi:10.1212/WNL.33.11.1444.

17. Brookhart M alan, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J*

Epidemiol 2006; **163**: 1149–1156. doi:10.1093/aje/kwj149.

18. Austin PC. A comparison of 12 algorithms for matching on the propensity score. *Stat Med* 2014; **33**: 1057–1069. doi:10.1002/sim.6004.

19. Austin PC, Stuart E a. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015; **34**: 3661–3679. doi:10.1002/sim.6607.

20. Shuster JJ. Median follow-up in clinical trials. *J Clin Oncol* 1991; **9**: 191–2. doi:10.1200/JCO.1991.9.1.191.

21. Royston P, Parmar MKB. Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* 2013; **13**. doi:10.1186/1471-2288-13-152.

22. Renoux C, Suissa S. Immortal time bias in the study of effectiveness of interferon-β in multiple sclerosis. *Ann Neurol* 2008; **64**: 109–110. doi:10.1002/ana.21352.

23. Shirani A, Zhao Y, Karim ME, *et al.* Association between Use of Interferon Beta and Progression of Disability in Patients. *Jama* 2012; **308**: 247–256.

24. Tintore M, Rovira À, Río J, *et al.* Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain* 2015; **138**: 1863–1874. doi:10.1093/brain/awv105.

25. Tintoré M. Rationale for early intervention with immunomodulatory treatments. *J Neurol* 2008; **255**: 37–43. doi:10.1007/s00415-008-1006-4.

26. Trojano M, Pellegrini F, Fuiani A, *et al.* New natural history of interferon-??-treated relapsing multiple sclerosis. *Ann Neurol* 2007; **61**: 300–306. doi:10.1002/ana.21102.

27. Cocco E, Sardu C, Spinicci G, *et al.* Influence of treatments in multiple sclerosis disability: A cohort study. *Mult Scler J* 2015; **21**: 433–441. doi:10.1177/1352458514546788.

28. Drulovic J, Kostic J, Mesaros S, *et al.* Interferon-beta and disability progression in relapsing-remitting multiple sclerosis. *Clin Neurol Neurosurg* 2013; **115**: S65–S69. doi:10.1016/j.clineuro.2013.09.024.

29. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; **46**: 399–424. doi:10.1080/00273171.2011.568786.

30. Papageorgiou SN, Koretsi V, Jäger A. Bias from historical control groups used in orthodontic research: A metaepidemiological study. *Eur J Orthod* 2017; **39**: 98–105. doi:10.1093/ejo/cjw035.

31. Spencer E, Brassey J, Heneghan C. Non-contemporaneous control bias. *Cat Bias* 2017. Available at: https://catalogofbias.org/biases/non-contemporaneous-control-bias/.

32. Walton MK. Historical Controls for Clinical Trials Contemplation on Use in Drug Development. *Present FDA* 2012: 1–30. Available at: https://events-support.com/Documents/Walton_Marc.pdf.

Tables

Table 1: The characteristics of the study population (N=2,383) and comparison of the matched and the excluded patients at disease clinical onset, according to the main outcome (EDSS 3)

		All	Matched	Excluded	p[†]
		N=2,383	**N=1,112**	**N=1,271**	
Sex	Women	75%	77%	73%	0.014
Age at MS‡ clinical onset	Mean (SD[§])	30 (9.3)	29 (8.8)	31 (9.6)	$<10^{-4}$
Year of MS‡ onset					
	Before 1998	56%	44%	67%	$<10^{-4}$
	[1998, 2001]	23%	27%	20%	
	[2002, 2005]	21%	29%	13%	

[†] p-value of Fisher's test or Student's t-test comparing the matched and the excluded patients

‡Multiple sclerosis

[§] Standard deviation

Table 2: The baseline (=matching) characteristics of the matched cohort (N=1,112) for studying the time to reach EDSS 3, overall and according to the exposure group

		All	Treated	Not yet treated	SMD[†]
		N=1,112	N=556	N=556	
Age at matching	Mean (SD[‡])	34 (9.9)	34 (9.7)	35 (10.1)	4.3%
Relapses in the past 12 months					7.6%
	0	27%	27%	27%	
	1	43%	42%	44%	
	2	22%	21%	22%	
	3	6%	6%	6%	
	More than 3	2%	3%	2%	
EDSS score					6.3%
	[0,1]	42%	43%	41%	
	[2,3]	42%	42%	43%	
	4 and more	15%	15%	16%	
EDSS variation since the last consultation					2.3%
	Improvement	7%	8%	7%	
	Stable	81%	80%	82%	
	Worsening	11%	12%	11%	
Delay since first consultation at the expert center					2.5%
	None	46%	46%	45%	
	Recent	33%	33%	33%	
	Old	21%	21%	22%	

[†] SMD: standardized mean differences between the treated and the "not yet treated" patients; [‡] SD: Standard deviation

Table 3: The mean time before reaching outcomes in the two exposure groups, over a 10-year period

	Mean	95% CI
EDSS† 3		
RMST‡ for treated	8.82	[8.59; 9.05]
RMST‡ for "not yet treated"	9.44	[9.26; 9.59]
Difference	**-0.62**	**[-0.91; -0.33]**
Disability progression		
RMST‡ for treated	9.12	[8.90 ; 9.32]
RMST‡ for "not yet treated"	9.09	[8.88; 9.30]
Difference	**0.03**	**[-0.24; 0.33]**

† EDSS: Expanded Disability Status Scale

‡RMST: restricted mean survival time

Figure legends

Figure 1: Flowchart of the study population (n=2,383)

Figure 2: Kaplan-Meier estimates of the time to reach the outcomes A, B

Legend:

A: Time before reaching an irreversible EDSS score of 3;

B: Time before the first worsening of the EDSS;

For all of the subfigures, the black line represents the survival curve for the treated patients; the grey dotted line represents the survival curve for the "not yet treated" patients

Figure 3: The characteristics of exposure after matching for the treated patients and the "not yet treated" patients

Legend:

A: The characteristics of exposure for the treated patients after matching;

B: The characteristics of exposure for the "not yet treated" patients after matching

The first-line treatment group included beta-interferon, glatiramer acetate, dimethyl fumarate, teriflunomide, and peg-interferon; The second- and third-line treatment groups included mitoxantrone, natalizumab, and fingolimod; Off-label treatments included azathioprine, cladribine, cyclophosphamide, and methotrexate.

Figure 1: Flowchart of the study population (n=2,383)

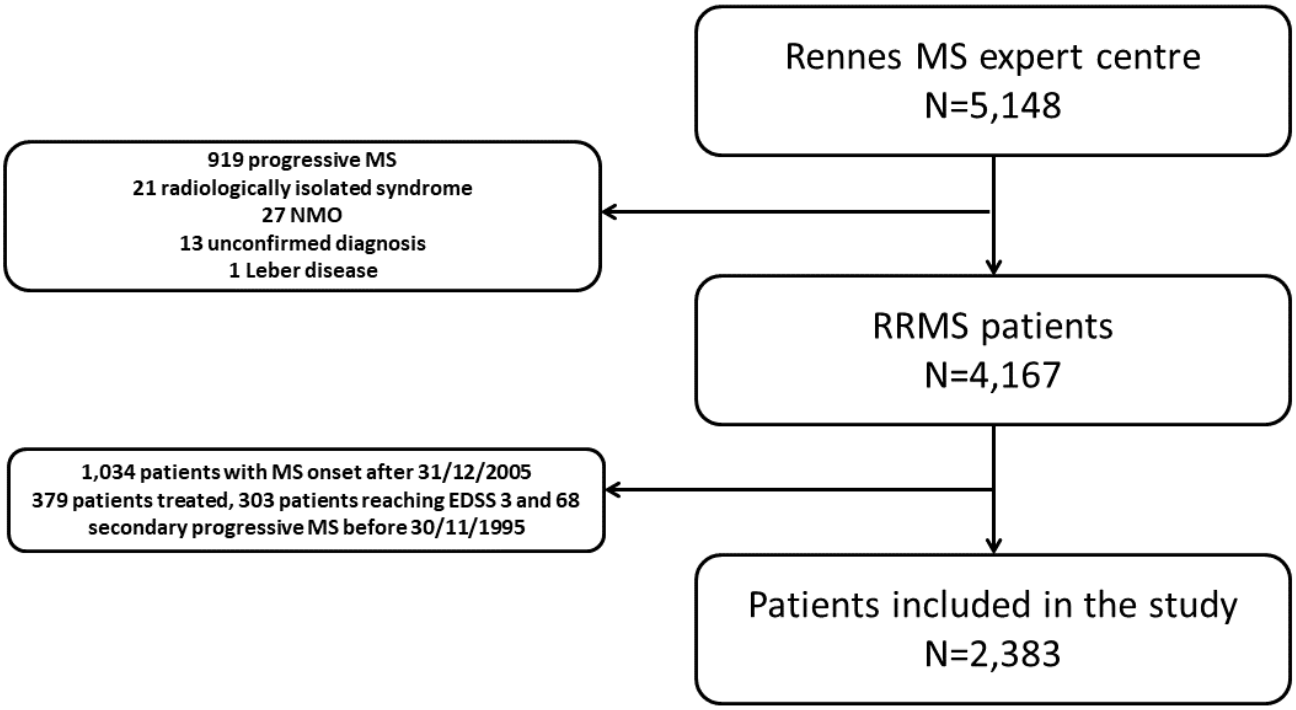
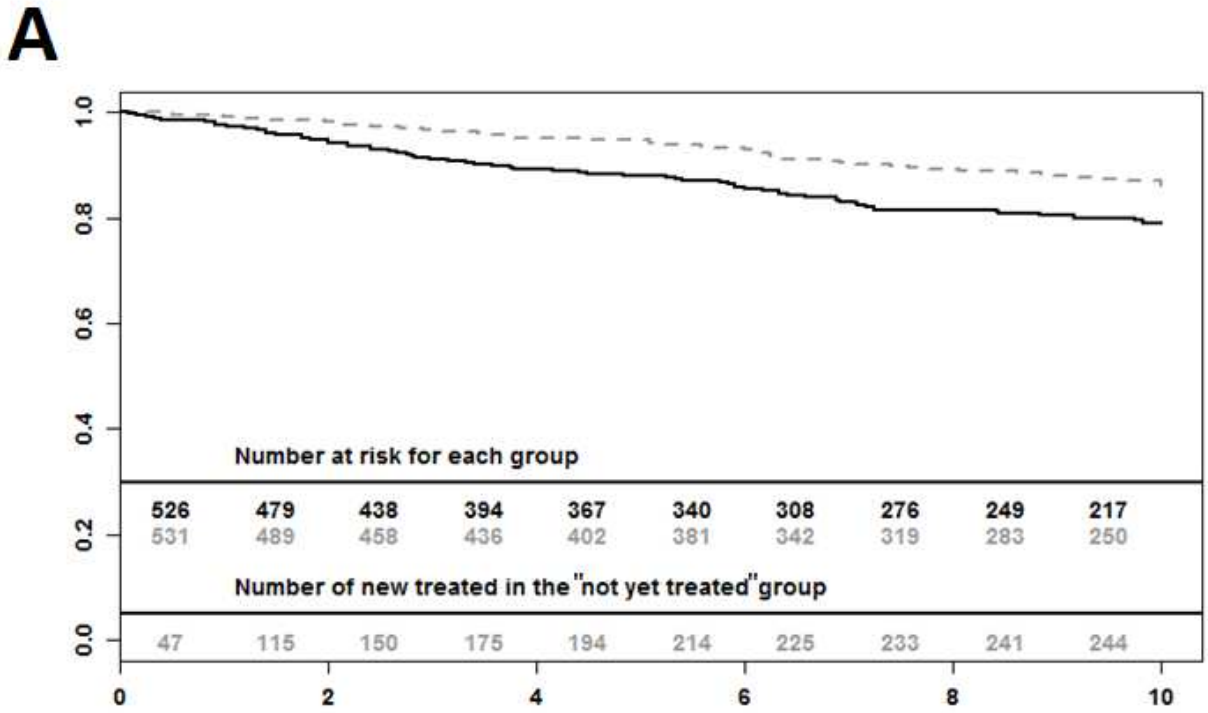


Figure 2: Kaplan-Meier estimates of the time to reach the outcomes A, B



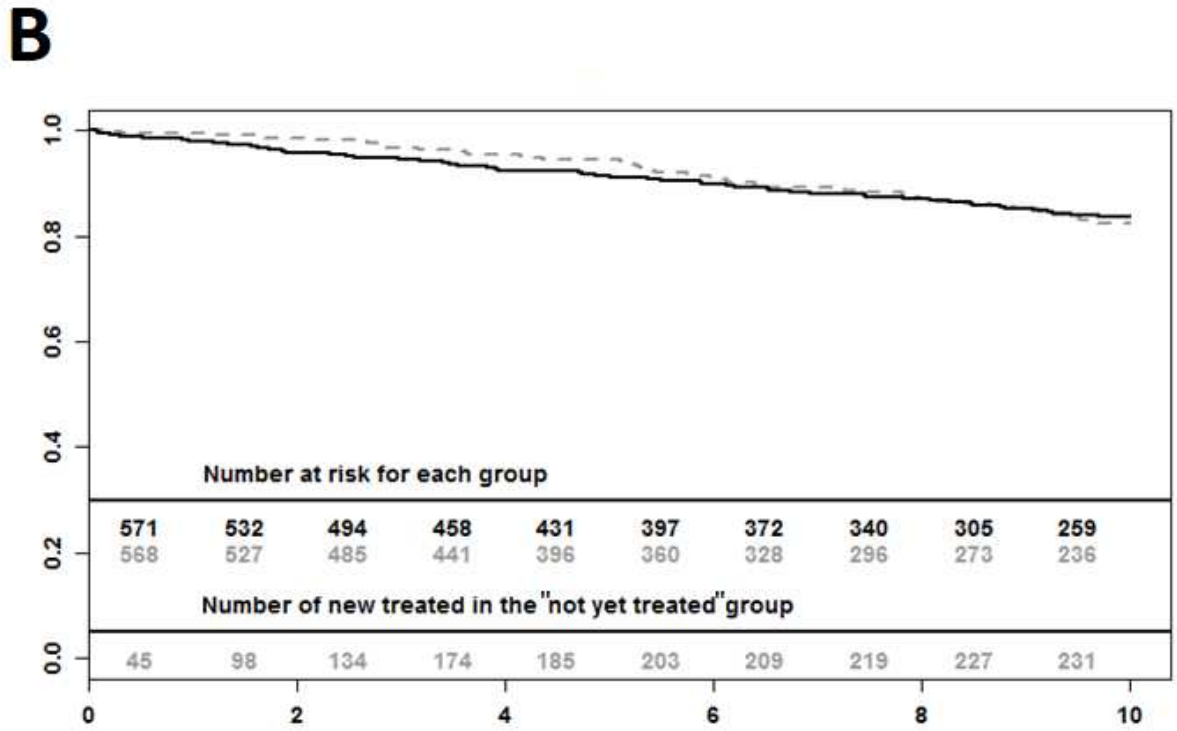
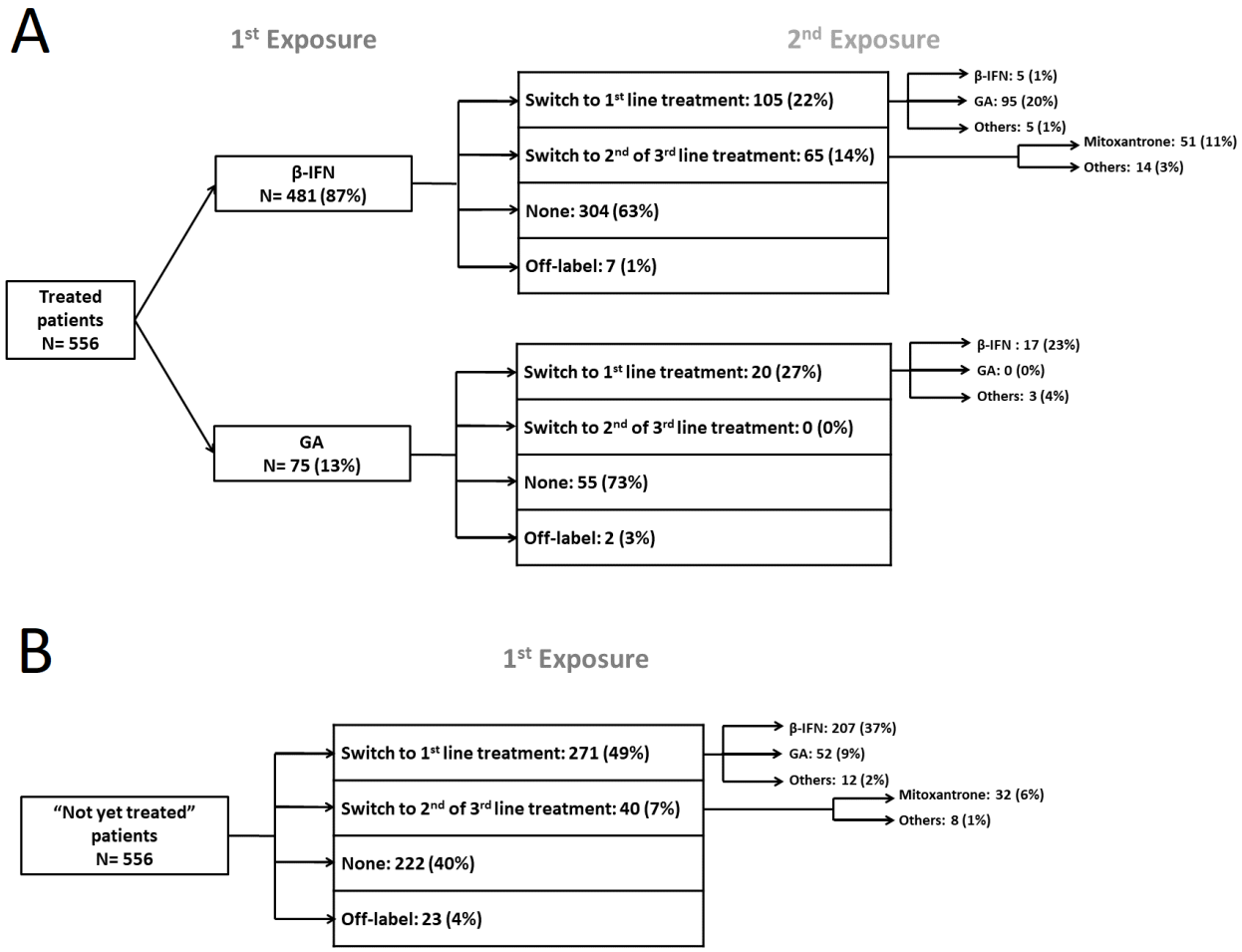


Figure 3: The characteristics of exposure after matching for the treated patients and the "not yet treated" patients



Supplementary Material

Figure S1: Kaplan-Meier estimator of the time before being treated by β-IFN or GA

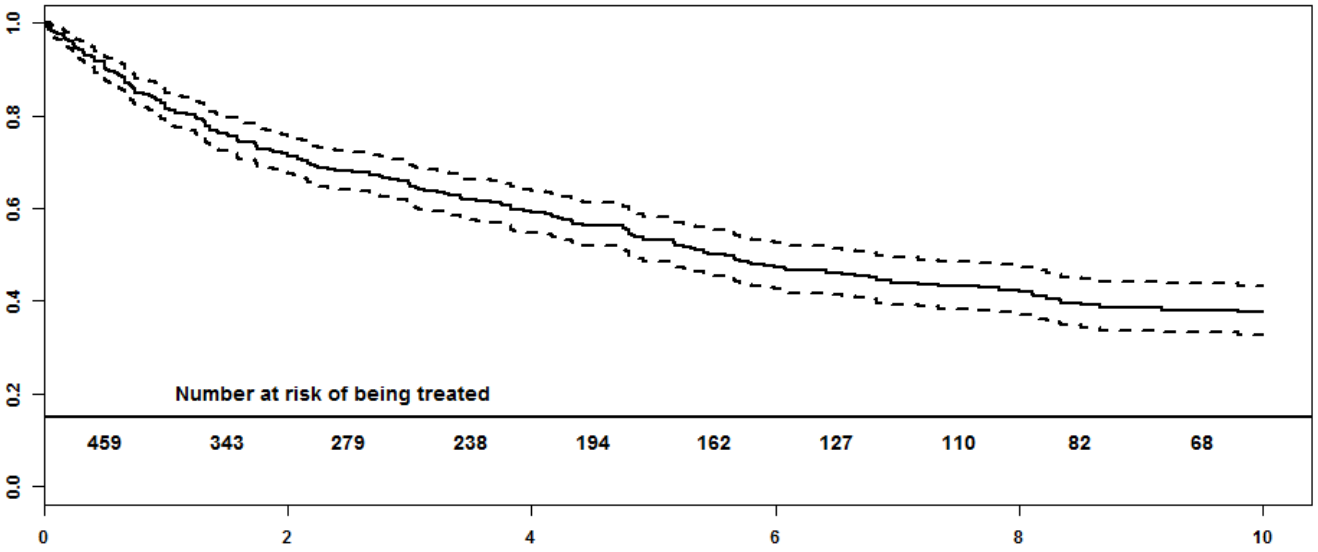


Table S1: Propensity score model definition: association between variables and outcomes

	Time to reach EDSS 3	Time to the first worsening of the EDSS	Number of relapses over a 5-year period	Number of relapses over a 10-year period
Sex			X	X
Age	X		X	X
Relapses in the past 12 months				
1 (vs. 0)	X	X	X	X
2 (vs. 0)	X	X	X	X
3 (vs. 0)	X	X	X	X
More than 3 (vs. 0)	X	X	X	X
EDSS score				
[2,3] (vs. [0,1])	X			X
4 and more (vs. [0,1])	X	X	X	X
EDSS variation since the last consultation				
Stable (vs. Improvement)		NA	NA	NA
Worsening (vs. Improvement)	X	NA	NA	NA
Delay since the first consultation at the expert center				
Recent (vs. None)	X		X	X
Old (vs. None)	X			
MS† onset year				
Before 1998 (vs. [2002; 2005])			X	X
[1998; 2001] (vs. [2002; 2005])				

† MS: Multiple Sclerosis ; Commentary: The variables associated with the outcome of interest were used to build the propensity score leading to three different matched datasets: one for studying EDSS 3, one for studying the first worsening, and one for studying the number of relapses over a 5-year or a 10-year period.

Table S2: The characteristics of the study population (N=2,383) and comparison of the matched and the excluded patients at disease clinical onset, according to the secondary outcomes

		All	Matched	Excluded	p[†]
Disability becoming worse		**N=2,383**	**N=1,190**	**N=1,193**	
Sex	Women	75%	77%	72%	0.014
Age at MS‡ clinical onset	Mean (SD[§])	30 (9.3)	29 (8.9)	30 (9.6)	$<10^{-4}$
Year of MS‡ onset					
	Before 1998	56%	44%	68%	$<10^{-4}$
	[1998, 2001]	23%	26%	20%	
	[2002, 2005]	21%	29%	12%	
Number of relapses over a 5-year or a 10-year period		**N=2,383**	**N=1,124**	**N=1,259**	
Sex	Women	75%	78%	72%	0.001
Age at MS‡ clinical onset	Mean (SD[§])	30 (9.3)	29 (8.8)	31 (9.6)	$<10^{-4}$
Year of MS‡ onset					
	Before 1998	56%	44%	67%	$<10^{-4}$
	[1998, 2001]	23%	26%	20%	
	[2002, 2005]	21%	29%	13%	

[†] p-value of Fisher's test or Student's t-test comparing the matched and the excluded patients

‡Multiple sclerosis

[§] Standard deviation

Table S3: The baseline (=matching) characteristics of the matched cohort (N=1,190) for studying the time to first worsening of the EDSS, overall and according to the exposure group

		All	Treated	Not yet treated	SMD[†]
		N=1,190	N=595	N=595	
Relapses in the past 12 months					8.3%
	0	27%	28%	27%	
	1	44%	43%	46%	
	2	20%	21%	19%	
	3	6%	6%	6%	
	More than 3	2%	2%	1%	
EDSS score					6.6%
	[0,1]	42%	43%	40%	
	[2,3]	43%	42%	44%	
	4 and more	15%	15%	15%	

[†] SMD: standardized mean differences between the treated and the “not yet treated” patients

Table S4: The baseline (=matching) characteristics of the matched cohort (N=1,124) for studying the number of relapses over a 5-year and a 10-year period, overall and according to the exposure group

		All	Treated	Not yet treated	SMD†
		N=1,124	N=562	N=562	
Sex	Women	78%	76%	80%	10.0%
Age at matching	Mean (SD‡)	35 (9.8)	35 (9.6)	34 (9.9)	6.5%
Relapses in the past 12 months					7.9%
	0	27%	27%	26%	
	1	44%	43%	45%	
	2	21%	21%	21%	
	3	5%	5%	4%	
	More than 3	2%	3%	2%	
EDSS score					7.7%
	[0,1]	42%	42%	42%	
	[2,3]	42%	41%	44%	
	4 and more	16%	17%	14%	
Delay since first consultation at the expert center					6.6%
	None	47%	46%	48%	
	Recent	32%	32%	33%	
	Old	21%	22%	19%	

† SMD: standardized mean differences between the treated and the “not yet treated” patients ; ‡ SD: Standard deviation

Chapitre 4

IV. Comparaison d'efficacité de traitements en vie réelle : l'exemple du natalizumab et du fingolimod

A. Contexte général de l'étude 106
1. Contexte scientifique 106
2. Contexte logistique 107
B. Les études originales 108
1. Des différences méthodologiques 108
2. Des résultats discordants 111
C. Une comparaison par étape 111
1. Les critères d'inclusion 111
2. L'hétérogénéité des pratiques 112
3. Les méthodes d'équilibre et d'analyse des données 113
4. Les critères de jugement 114
5. Déroulement de l'étude 115
D. Résultats issus de la mobilité 115

A. Contexte général de l'étude

1. Contexte scientifique

D'indication similaire, le natalizumab et le fingolimod sont prescrits en cas de SEP très actives, la plupart du temps en deuxième intention et quand cela est jugé nécessaire, en première intention, à des patients naïfs de tout traitement (HAS, 2017, 2019). En deuxième intention, ils sont proposés dans le cas où le patient ne répond pas ou plus à un traitement de première ligne bien conduit, i.e. s'il présente une activité de SEP importante : nouvelle(s) poussée(s), nouvelle(s) lésion(s) malgré le traitement. En première intention, ils sont proposés dans des cas de SEP sévères et d'évolution rapide (plusieurs poussées invalidantes en un an, une charge lésionnelle importante). Cependant, ces deux traitements présentent une différence notable en ce qui concerne les effets indésirables. En effet, comme évoqué en introduction, le natalizumab est associé à un haut risque de LEMP (Kleinschmidt-DeMasters, B.K. Tyler, 2005; Langer-Gould *et al.*, 2005) alors que le fingolimod présente un risque plus faible (Vargas and Tyor, 2017). Ce risque constitue un effet indésirable grave à considérer dans la décision de traiter par natalizumab.

Dans ce contexte, la comparaison de l'efficacité de ces deux traitements est utile pour guider le neurologue dans sa décision thérapeutique, le choix du natalizumab pouvant s'avérer plus risqué. Pour des raisons financières et pratiques, la comparaison du natalizumab et du fingolimod n'a pas été conduite par essai randomisé. En revanche, des études observationnelles ont été menées, mais ont conduit à des résultats discordants. En effet, certaines n'ont pas montré de différence d'efficacité entre le fingolimod et le natalizumab (Braune *et al.*, 2013; Koch-Henriksen *et al.*, 2017) et d'autres ont montré une supériorité du natalizumab par rapport au fingolimod (Gajofatto *et al.*, 2014; Kalincik *et al.*, 2015; Barbin *et al.*, 2016). Ces discordances interpellent et empêchent de fournir un message clair aux neurologues. Ces études ont été conduites à partir de différentes sources de données en tenant compte de critères d'inclusion et de méthodes différents. L'exploration de ces études pourrait permettre de formuler des explications sur les différences de résultats. Cela pourrait contribuer à fournir des clés de compréhension sur les variations observées sur les résultats afin de conclure sur la comparaison de ces deux traitements, en confirmant la supériorité du natalizumab par rapport au fingolimod ou non.

Dans ce contexte, une opportunité de collaboration de trois partenaires ayant proposé des études comparatives de l'efficacité du natalizumab et du fingolimod en vie réelle a émergé. La finalité de cette mobilité était d'explorer l'impact des choix méthodologiques sur les résultats de ces trois études comparatives de l'efficacité du natalizumab et du fingolimod en vie réelle sur la progression du handicap chez les patients atteints de SEP rémittente. Dans la prochaine partie, le contexte logistique de cette mobilité est présenté puis nous aborderons plus en détails les trois études originales.

2. Contexte logistique

Cette étude s'inscrit dans un projet collaboratif mené avec l'équipe CORe (Clincal Outcomes Research Unit) de l'université de Melbourne en Australie et l'équipe du registre danois : The Danish Multiple Sclerosis Group, de l'université de Copenhague au Danemark.

L'équipe CORe est dirigée par Tomas Kalincik. Il a été l'encadrant de proximité de ce travail. Pour le registre, Melinda Magyari était identifiée comme référente projet, ainsi que Nils Koch-Henriksen étant le premier auteur de l'étude originale danoise. Côté français, l'OFSEP nous a associé à ce projet. En particulier, Sandra Vukusic, coordinatrice de l'OFSEP, a eu l'idée originale de ce projet et m'a proposé cette mobilité. David Laplaud, initiateur de l'étude originale française était également associé. Enfin, des points réguliers sur le sujet étaient réalisés avec Emmanuelle Leray, ma directrice de thèse.

Cette collaboration a donné lieu à une mobilité de deux mois à l'université de Melbourne en juin et juillet 2018. Dans chaque équipe, un membre junior a été associé à la collaboration. Sifat Sharmin, issue d'une formation biostatistique, commençait depuis peu de temps un post doctorat dans l'équipe CORe et, de fait, découvrait la SEP. Johanna Andersen, kinésithérapeute en reconversion, a débuté un doctorat après la mobilité, en octobre 2018, à la suite d'un master en santé publique au cours duquel elle avait réalisé un stage sur la santé mentale des enfants nés de mères atteintes de SEP. Pour ma part, je travaillais déjà depuis un an et demi sur des questions d'évaluation de l'efficacité en vie réelle dans la SEP.

En arrivant à Melbourne, les objectifs de chacun divergeaient quelque peu, probablement en raison de difficultés de compréhension en amont et de différences de niveaux de connaissances du sujet et de compétences statistiques.

Johanna était venue dans une démarche de formation sur les méthodes de scores de propension. Sifat, forte de son bagage statistique, découvrait la SEP et ses traitements. Pour ma part, l'avancement de mon doctorat m'avait déjà donné plusieurs éléments utiles pour aborder les questionnements de cette mobilité, que ce soit sur le plan méthodologique ou sur la connaissance de la SEP.

Mon déplacement et mon logement à Melbourne ont été financés grâce à une bourse de mobilité internationale obtenue auprès de la Fondation ARSEP (fondation pour l'aide à la recherche sur la SEP) et à une participation de mon équipe de rattachement EA 7449 REPERES Univ Rennes/EHESP (respectivement 6000€ et 1500€).

B. Les études originales

Le projet collaboratif a été consacré à la comparaison des études originalement conduites dans les équipes associées : celle du groupe MSBase (Kalincik *et al.*, 2015), celle portant sur le registre danois (Koch-Henriksen *et al.*, 2017) et enfin celle portant sur l'OFSEP (Barbin *et al.*, 2016). Dans cette partie, nous allons présenter ces trois études en détails.

1. Des différences méthodologiques

La première différence concerne tout d'abord les données (MSBase, le registre danois et l'OFSEP). Celles-ci ont été présentées brièvement dans le chapitre 1. En particulier, elles se différencient par leur caractère exhaustif ou non et leur statut national ou non. Bien que la nature des données soient proches permettant leur analyse commune, elles reflètent des pratiques différentes, menées dans des pays aux situations économiques et culturelles pouvant impacter la prise en charge.

Le Tableau 2 résume les éléments relatifs à la méthodologie utilisée dans chaque étude. Certaines différences sont à noter, notamment les critères d'inclusion, le type d'analyse (per protocol ou intention de traiter) et la méthode de score de propension utilisée pour équilibrer les groupes à comparer.

Dans l'étude MSBase, les patients inclus étaient étudiés dans un contexte de changement de traitement, en raison de l'observation d'une activité sous traitement par interféron-β ou

acétate de glatiramère. Dans l'étude OFSEP ou celle du registre danois, en revanche, tout type d'initiation de natalizumab ou fingolimod était concerné par l'étude. Ces deux études incluaient donc les patients pour lesquels les traitements étaient utilisés dans une stratégie d'escalade (deuxième ligne en deuxième intention), qu'une activité de maladie justifiant le changement ait été observé ou non, ainsi que les patients traités en stratégie intensive (deuxième ligne d'emblée, en première intention). De ce fait, on pourrait supposer que dans l'étude OFSEP et l'étude du registre danois, un groupe de patients plus actifs ait été inclus. Notons tout de même que dans l'étude MSBase, les patients étaient tous en échec de traitement de première ligne donc présentaient une maladie active également.

Dans l'étude OFSEP, une analyse en intention de traiter a été conduite alors que des analyses en per protocol ont été menées dans l'étude de MSBase et dans celle du registre danois.

Enfin, dans l'étude de MSBase, un appariement avec censure par paire appelé « pairwise censoring » a été utilisé pour équilibrer les deux groupes. A notre connaissance, la pratique de « pairwise censoring » est utilisée seulement par l'équipe de MSBase. Elle consiste à censurer les données par paire résultant de l'appariement selon le suivi le plus court de la paire. Dans l'étude du registre, il s'agissait d'un simple appariement. Dans l'étude OFSEP, une pondération par la probabilité inverse d'être traité a été utilisée.

Notons également que dans l'étude OFSEP, les données d'imagerie avaient pu être utilisées. Dans l'étude MSBase, celles-ci avaient été utilisées après imputation.

Tableau 2 : Comparaison des méthodologies utilisées dans chacune des trois études comparant le natalizumab et le fingolimod considérées dans l'étude collaborative

	Etude MSBase	Etude OFSEP	Etude Registre Danois
Période d'étude	/ *~ 1ère exposition au natalizumab ou au fingolimod*	Juillet 2011 à Janvier 2013	Juillet 2011 à Mars 2015
Date d'entrée dans l'étude t_0	Initiation du natalizumab ou du fingolimod	Initiation du natalizumab ou du fingolimod	Initiation du natalizumab ou du fingolimod
Inclusion	SEP rémittente, escalade verticale en raison d'activité de la maladie EDSS[1] disponible à dans l'intervalle $[t_0 - 6\ mois, t_0 + 1\ mois]$, au moins 2 visites supplémentaires durant le suivi	SEP rémittente, initiant natalizumab ou fingolimod 18-65 ans, EDSS[1] <6, EDSS[1] disponible + au moins une visite et une IRM disponible dans les 12 mois précédents	SEP rémittente, initiant natalizumab ou fingolimod, EDSS[1] disponible à t_0
Exclusion	Alemtuzumab, cladribine, infliximab, mitoxantrone, rituximab, dimethyl fumarate, teriflunomide	cyclophosphamide, mitoxantrone, alemtuzumab, rituximab	/
Censure	Per protocol	Intention de traiter	Per protocol
Analyse statistique	Appariement sur le score de propension avec pairwise censoring	Pondération sur le score de propension (IPTW[2])	Appariement sur le score
Variables	Sexe, Age, durée de la SEP, EDSS[1], nombre de traitement précédents, Pays	Sexe, âge, durée de la SEP, EDSS[1], Nombre de poussées, Prise de gadolinium, centre expert (aléatoire)	Sexe, âge, durée de la SEP, taux annualisé de poussée pendant traitement précédent, type de traitement précédent, MSSS[3]
Outcome	Taux annualisé de poussée , aggravation et amélioration du handicap	% au moins 1 poussée, % au moins 1 prise de Gadolinium T1, % au moins 1 nouvelle lésion T2, Aggravation du handicap	Taux annualisé de poussée, % sans poussée, délai avant la 1ère poussée, aggravation et amélioration du handicap

[1] EDSS : Expanded Disability status scale, [2] IPTW: Inverse probability of treatment weighting, [3] MSSS : Multiple Sclerosis Severity Score (mesure de la sévérité de la maladie en utilisant le handicap et la durée de la SEP)

2. Des résultats discordants

Comme évoqué précédemment, les trois publications ont conduit à des résultats discordants. L'étude de MSBase (Kalincik *et al.*, 2015) et celle de l'OFSEP (Kalincik *et al.*, 2015) ont démontré une supériorité du natalizumab par rapport au fingolimod sur la réduction du nombre de poussées. Dans l'étude MSBase, il a été montré que l'amélioration du handicap était également meilleure avec natalizumab mais il n'a pu être montré de différence sur la réduction du handicap. Dans l'étude OFSEP, une supériorité du natalizumab était également observée sur la réduction d'activité IRM. L'étude du registre danois (Koch-Henriksen *et al.*, 2017) n'avait pu montrer aucune différence entre les deux traitements.

C. Une comparaison par étape

L'exploration des études originales a permis d'établir trois différences majeures pouvant conduire à expliquer les discordances de résultats : les critères d'inclusion, les données reflétant différentes pratiques de traitement et la méthodologie statistique.

1. Les critères d'inclusion

Dans les études originales, les critères d'inclusion étaient différents, sélectionnant des populations de patients aux profils particuliers. Dans ce contexte, quatre ensembles de critères d'inclusion ont été établis permettant d'explorer quatre populations différentes : (1) patients en situation de changement d'une première ligne vers natalizumab ou fingolimod en raison d'une activité durant l'exposition au traitement de première ligne, (2) patients en situation de changement d'une première ligne vers natalizumab ou fingolimod sans mention particulière concernant l'activité, (3) patients naïfs de tout traitement exposés au natalizumab ou au fingolimod, (4) patients initiant du natalizumab ou du fingolimod, sans condition particulière sur les antécédents d'exposition ou d'activité de SEP. Les critères d'inclusion sont présentés dans le Tableau 3.

Tableau 3 : Critères d'inclusion pour chaque population cible étudiée dans le projet collaboratif portant sur l'efficacité des traitements de deuxième ligne

Design 1 **Changement d'une 1ère ligne vers natalizumab ou fingolimod sous SEP active**	**Design 2** **Changement d'une 1ère ligne vers natalizumab ou fingolimod**	**Design 3** **Natalizumab ou fingolimod en première intention**	**Design 4** **Toute initiation de natalizumab ou fingolimod**
Atteint de SEP rémittente	Atteint de SEP rémittente	Atteint de SEP rémittente	Atteint de SEP rémittente
Initiant du NAT ou du FTY à partir du 01/01/2011	Initiant du natalizumab ou du fingolimod à partir du 01/01/2011	Initiant du natalizumab ou du fingolimod à partir du 01/01/2011	Initiant du natalizumab ou du fingolimod à partir du 01/01/2011
Non exposé aux traitements exclus[1]	Non exposé aux traitements exclus[1]	Non exposé aux traitements exclus[1]	Non exposé aux traitements exclus[1]
Ayant au moins une visite à l'initiation [t_0-6mois; t_0+1 mois]	Ayant au moins une visite à l'initiation [t_0-6mois; t_0+1 mois]	Ayant au moins une visite à l'initiation [t_0-6mois; t_0+1 mois]	Ayant au moins une visite à l'initiation [t_0-6mois; t_0+1 mois]
Ayant été traité par interféron-β ou acétate de glatiramère avant l'initiation	**Ayant été traité par interféron-β ou acétate de glatiramère avant l'initiation**	**N'ayant pas été exposé avant l'initiation**	
Ayant une activité de SEP (poussée ou dégradation de l'EDSS) sous interféron-β ou acétate de glatiramère dans l'année précédant l'initiation			
Ayant au moins une visite sous le natalizumab ou le fingolimod et deux autres visites à au moins 6 mois d'intervalle	Ayant au moins une visite sous le natalizumab ou le fingolimod et deux autres visites à au moins 6 mois d'intervalle	Ayant au moins une visite sous le natalizumab ou le fingolimod et deux autres visites à au moins 6 mois d'intervalle	Ayant au moins une visite sous le natalizumab ou le fingolimod et deux autres visites à au moins 6 mois d'intervalle

[1] Liste des traitements : alemtuzumab, cellules souches hématopoïétiques autologues, mitoxantrone, ocrelizumab, rituximab, cladribine, daclizumab, cyclophosphamide ou intégré dans un essai randomisé

2. L'hétérogénéité des pratiques

Un des éléments majeurs qui pourrait expliquer les résultats discordants, mais pour lequel il est difficile d'agir, concerne les données. De nature différente et collectées dans divers pays, elles représentent des pratiques hétérogènes dans des systèmes de soins différents, incluant un accès aux soins et/ou aux traitements variable.

Dans ce contexte, plusieurs sous-groupes de patients ont été analysés afin d'étudier si l'efficacité des traitements pouvait être différenciée selon telle ou telle caractéristique démographique ou clinique.

Les femmes et les hommes ont été étudiés séparément dans une analyse stratifiée sur le sexe. Concernant l'âge, deux analyses ont été menées. Un premier sous-groupe de patients âgés de 18 à 65 ans a été étudié puisque l'étude originale de l'OFSEP n'avait inclus que ce sous-groupe. Une analyse stratifiée selon l'âge médian à l'initiation du traitement (38 ans) a également été conduite, ainsi qu'une analyse stratifiée selon la durée médiane de la SEP à l'initiation du traitement (7 ans). Puisque la première étude de l'OFSEP avait restreint le groupe des patients inclus aux patients n'ayant pas atteint un EDSS 6 irréversible, ce sous-groupe a été analysé. Ensuite, une analyse stratifiée selon trois groupes déterminés selon la valeur de l'EDSS à l'initiation a été conduite : EDSS inférieur à 4, EDSS compris entre 4 et 5.5 et EDSS supérieur ou égal à 6. Une stratification par rapport à l'activité récente de la maladie a été réalisée, conduisant à l'analyse séparée de trois groupes : patients ayant fait des poussées dans l'année précédant l'initiation, patients ayant connu une progression du handicap sans poussée et patients sans aucune activité de maladie relative au handicap ou aux poussées. Finalement, une analyse tenant compte des données d'imagerie a été menée. Sur le sous-groupe de patients ayant une IRM disponible à l'initiation du traitement, nous avons conduit l'analyse globale, puis stratifiée selon l'activité d'imagerie : un groupe de patients n'ayant ni de nouvelle lésion en T2, ni de prise de Gadolinium et un groupe présentant l'un au moins de ces éléments.

3. Les méthodes d'équilibre et d'analyse des données

Le dernier élément sur lequel nous nous sommes interrogés concerne la méthodologie statistique. Dans les études précédemment conduites, plusieurs méthodes utilisant les scores de propension ont été utilisées. De plus, les analyses ont été conduites soit en intention de traiter, soit en per protocol.

Les différentes méthodes ont été appliquées aux différentes populations d'étude et critères de jugement. Le tableau ci-dessous résume les différents cadres analytiques appliqués à chaque design d'étude.

Tableau 4 : Analyses statistiques conduites dans le cadre du projet collaboratif portant sur l'efficacité des traitements de deuxième ligne

	Analyse des données longitudinales		
Equilibre des deux groupes	**Intention de traiter**	**Per protocol**	**Pairwise censoring**
Appariement, caliper = 0,1	x	x	x
Appariement, caliper = 0,2	x	x	x
Appariement, caliper = 0,02	x	x	x
Pondération par les odds	x	x	
Pondération sIPTW	x	x	

4. Les critères de jugement

Plusieurs critères de jugement ayant été étudiés dans les études originales, nous avons intégré les principaux dans ce travail collaboratif. Au total, six critères de jugement ont ainsi été analysés :

#1 Le nombre de poussées sous traitement

#2 Le délai avant la première poussée après l'initiation du traitement

#3 Le délai avant la première aggravation du handicap

#4 Le délai avant la première amélioration du handicap

#5 Evaluation multiple d'aggravation du handicap

#6 Evaluation multiple d'amélioration du handicap

L'aggravation de l'EDSS (Kurtzke, 1983) était définie par une augmentation de la valeur de l'EDSS selon la règle suivante : augmentation confirmé à 6 mois et maintenu jusqu'aux dernières nouvelles d'1,5 point si l'EDSS était compris entre 0 et 1,5, d'1 point si l'EDSS à l'initiation était compris entre 2 et 5,5 et enfin 0,5 si l'EDSS était supérieur ou égal à 6. L'amélioration de l'EDSS était définie selon une règle similaire, tenant compte d'une diminution de la valeur du score EDSS confirmé à 6 mois et maintenu.

5. Déroulement de l'étude

Les premières semaines de la mobilité ont été consacrées à la réplication des études originales pour lesquelles les données avaient pu être mises à jour, à savoir les données de MSBase et les données du registre. Aucune de nous trois n'avions participé à la réalisation des premières études. Durant ces premières semaines, j'ai accompagné Johanna pour mieux comprendre le cadre méthodologique, que ce soit celui de l'étude originale du registre danois ou celle que nous allions conduire.

Les données de l'OFSEP n'ayant pas été mises à jour, je disposais de plus de temps pour réfléchir à la présente étude. A partir des études originales, j'ai proposé les critères d'inclusion. Nous les avons ensuite rediscutés avec nos responsables. Pour chaque population, les données ont été étudiées ensemble (« poolées ») et par cohorte source, en prenant soin d'exclure les patients danois de la cohorte MSBase. Ainsi, un travail de nettoyage de données a été réalisé pour rassembler les trois bases de données. J'ai préparé les données danoises et françaises pour le regroupement des bases. Sifat, à l'aise avec la programmation, avait pu réaliser la préparation de données de MSBase.

Enfin, les semaines suivantes ont été consacrées à l'exploration des questionnements spécifiques précédemment évoqués. Les analyses ont été réalisées par Sifat et moi-même. Johanna ne maitrisant pas le logiciel, elle a commencé la rédaction d'un des articles en projet. Sifat a réalisé les analyses basées sur l'appariement et j'ai conduit celles basées sur la pondération.

D. Résultats issus de la mobilité

Des trois champs d'explorations, sont proposés trois articles. Chacune d'entre nous a en charge la rédaction d'un article et nous partageons toutes les trois systématiquement la place de premier auteur du fait de ce travail collaboratif.

Le premier article, intitulé « Understanding heterogeneity in comparative studies of treatment effectiveness in multiple sclerosis - Part 1: Validation of the comparisons of natalizumab and fingolimod », avait pour objectif de comparer l'efficacité du natalizumab et du fingolimod à partir des trois sources de données séparément et des données « poolées » en

uniformisant les critères d'inclusion et la méthode statistique. Cet article a été principalement rédigé par Johanna Andersen.

Dans cet article, en utilisant un schéma d'étude, des critères d'inclusion et une méthodologie unifiés, nous proposons une nouvelle analyse des données cliniques originales, séparément puis regroupées.

Dans ce travail, les patients étudiés répondaient aux critères du design 2, i.e. des patients en situation de changement de traitement après une première exposition à un traitement de première ligne. Dans le but de contrôler le biais d'indication, une pondération sIPTW a été utilisée. Les analyses de tous les critères de jugement (#1 à #6) ont été rapportées.

Les cohortes étudiées séparément ont fourni des résultats similaires aux études originales et donc, toujours discordants, malgré l'uniformisation de la méthode. Les analyses de la cohorte regroupée ont montré que le natalizumab présentait une efficacité supérieure au fingolimod dans la réduction du risque de poussée et l'amélioration du handicap. Dans cette étude, aucune différence dans le risque d'aggravation du handicap n'a pu être montrée entre les deux traitements. Ces résultats montrent que les différences d'efficacité entre le natalizumab et le fingolimod semblent dépendre des caractéristiques des cohortes étudiées.

Ces caractéristiques méritent d'être approfondies car elles apporteraient des éléments pour guider une approche personnalisée du traitement de la SEP. Il s'agit du thème du deuxième article qui avait pour objectif de comparer l'efficacité du natalizumab et du fingolimod par sous-groupe de patients déterminé à partir des caractéristiques démographiques et cliniques des patients (base poolée). Cet article est en cours de rédaction par Sifat Sharmin. Elle n'a pas été en mesure de le finaliser cette année en raison d'un congé maternité. L'article sera intitulé « Understanding heterogeneity in comparative studies of treatment effectiveness in multiple sclerosis - Part 3: Effect of cohort characteristics ». Par ailleurs, ce travail a été présenté par communication affichée au congrès de European Commitee for Treatment and Research in Multiple Sclerosis (ECTRIMS), en 2019. Le poster est disponible en annexe E. Ce travail étant moins directement lié à mon travail de thèse, seul un résumé est proposé ci-après.

Dans cette deuxième partie, nous avons cherché à savoir si la supériorité du natalizumab sur la réduction du risque de poussée ou l'amélioration du handicap est universelle pour tous les groupes de patients. Cette étude a donc évalué les avantages du natalizumab et du fingolimod pour le traitement dans les sous-groupes de patients atteints de SEP rémittente en stratifiant en fonction des caractéristiques démographiques et cliniques.

Les patients étudiés répondaient aux critères du design 4 c'est à dire tout type d'initiation de natalizumab et fingolimod. Les analyses ont été réalisées sur la base globale regroupant les trois cohortes. A l'instar de l'article précédent, une pondération sIPTW a été utilisée. De même, les critères de jugement portant sur le nombre de poussées sous traitement (#1), la progression (#5) et l'amélioration (#6) du handicap ont été rapportés. Les résultats de l'analyse per protocol sont présentés ci-après.

Le natalizumab a été associé à un nombre réduit de poussée par rapport au fingolimod chez les femmes (rapport de taux d'incidence (IRR) = 0,76 [0,65 ; 0,88]), groupes d'âge de 18 à 65 ans (IRR= 0,78 [0,68 ; 0,89]) et ≤38 ans (IRR = 0,64 [0,54 ; 0,76]), score EDSS<4 (IRR = 0,75 [0,64 ; 0,87]), <6 (IRR = 0,80 [0,70 ; 0,91]) et ≥6 (IRR = 0,52 [0,31 ; 0,86]), la durée de la maladie ≤ 7 ans (IRR = 0,63 [0,53 ; 0,76]), les patients présentant des rechutes avant l'inclusion (IRR = 0,74 [0,64 ; 0,86]) et de nouvelles lésions IRM (IRR = 0,51 [0,40 ; 0,66]) au cours de l'année précédant le début du traitement. Le natalizumab était associé à une meilleure amélioration du handicap, par rapport au fingolimod : chez les femmes (HR= 1,36 [1,10 ; 1,66]), les patients âgés de 18 à 65 ans (HR = 1,22 [1,02 ; 1,45]) et > 37 ans (HR = 1,34 [1,03 ;1,73]), score EDSS <6 (HR = 1,21 [1,00 ; 1,45]), les patients présentant des poussées avant l'initiation du traitement (HR = 1,23 [1,00 ; 1,50]) et chez les patients ne présentant aucune nouvelle lésion IRM observée 12 mois avant le début du traitement (HR = 4,35 [1,45 ; 13,04]). Pour les sous-groupes non rapportés, les résultats n'étaient pas significatifs, aucune différence ne pouvait être montrée.

Ces résultats conduisent à penser que l'efficacité supérieure du natalizumab par rapport au fingolimod, dans la réduction du risque de poussée et l'amélioration du handicap, est différenciée selon les caractéristiques des patients.

Enfin, le dernier article, que j'ai rédigé, est intitulé « Understanding heterogeneity in comparative studies of treatment effectiveness in multiple sclerosis - Part 2: Effect of analytical methodology ». Cet article a pour objectif d'étudier l'impact des méthodes statistiques sur les résultats d'une étude de comparaison d'efficacité : natalizumab versus fingolimod (base poolée). Il présente les résultats obtenus avec les différentes méthodes et reprend les étapes du processus méthodologique pour guider les choix et expliquer, si possible, l'impact qu'ils peuvent avoir sur les résultats. Cet article s'inscrit dans un travail amené à fournir des éléments de compréhension et des clés de lecture vis-à-vis des méthodes fréquemment utilisées pour comparer deux traitements en vie réelle.

Dans cet article, la population incluant tous les patients initiant du natalizumab ou du fingolimod a été étudiée (design 4). L'analyse a porté sur les critères de jugement #1 à #4 (nombre de poussées, délai avant la première poussée, délai avant la première aggravation du handicap et délai avant la première amélioration du handicap). Deux méthodes d'équilibres ont été utilisées : l'appariement et la pondération dont le descriptif a été proposé dans le chapitre 2, partie intitulée « Score de propension » (page 46). Les analyses ont été conduites en intention de traiter et en per protocol. Le « pairwise censoring » a également été utilisé.

Toutes les analyses ont conduit à des résultats proches avec des intervalles de confiance se chevauchant. Dans cet article, nous discutons les étapes et pratiques dans l'utilisation des scores de propension. Une discussion sur l'analyse en per protocol et en intention de traiter est également conduite.

Les deux articles présentés ci-dessus et disponibles à ce jour sont proposés dans la suite de ce document dans leurs versions actuelles. Le premier article proposé a été revu par les co-auteurs et est présenté dans un bon état d'avancement. En revanche, le second article proposé n'a pas encore été relu et est susceptible d'évoluer avant la soumission.

La soumission de ces articles est prévue pour le premier trimestre 2020. Une discussion est actuellement en cours avec l'éditeur du journal Jama Neurology pour considérer la soumission de ces trois papiers simultanément.

Understanding heterogeneity in comparative studies of treatment effectiveness in multiple sclerosis - Part 1: Validation of the comparisons of natalizumab and fingolimod

Authors:

Johanna Andersen*[(1)]; Sifat Sharmin*[(2)]; Mathilde Lefort*[(3)(4)]; Sandra Vukusic[(5-8)]; Helmut Butzkueven[(9-11),] , Emmanuelle Leray[(3)(4)(12)]; Nils Koch-Henriksen[(1)]; Tomas Kalincik**[(2)]; David-Axel Laplaud**[(13-15);] Melinda Magyari**[(1)]; on behalf of MSBase, Danish MS Registry, and OFSEP.

** Equal contribution as first author **Equal contribution as last author*

Affiliations:

[(1)]Department of Neurology, Danish Multiple Sclerosis Center and the Danish Multiple Sclerosis Registry, Rigshospitalet and University of Copenhagen, Copenhagen, Denmark

[(2)]CORe, Department of Medicine, University of Melbourne, Melbourne, Australia; Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia

[(3)]Rennes University, EHESP, REPERES – EA 7449, Rennes, France

[(4)]Univ Rennes, CHU Rennes, Inserm, CIC 1414, Rennes, France

[(5)]Hospices Civils de Lyon, Service de Neurologie, sclérose en plaques, pathologies de la myéline et neuro-inflammation, F-69677 Bron, France

[(6)]Observatoire Français de la Sclérose en Plaques, Centre de Recherche en Neurosciences de Lyon, INSERM 1028 et CNRS UMR 5292, F-69003 Lyon, France

[(7)]Université de Lyon, Université Claude Bernard Lyon 1, F-69000 Lyon, France

[(8)] Eugène Devic EDMUS Foundation against multiple sclerosis, state-approved foundation, F-69677 Bron, France

[(9)]Central Clinical School, Monash University, Melbourne, Australia;

[(10)]Department of Neurology, The Alfred Hospital, Melbourne, Australia;

[(11)]Department of Neurology, Box Hill Hospital, Monash University, Melbourne, Australia;

[(12)]METIS Department, EHESP French school of public health, Rennes, France

[(13)]Department of Neurology, CHU de Nantes, Nantes, France

[(14)]Univ Nantes, CHU Nantes, Inserm CIC 015, Nantes, France

[(15)]Centre de Recherche en Transplantation et Immunologie (CRTI), Inserm, UMR 1064, nantes, France

Introduction

Evidence-based 2018 guidelines (1) for the use of disease modifying drugs (DMDs) in multiple sclerosis (MS) suggest that in patients with relapsing-remitting MS (RRMS) the choice of DMD should be based upon patient characteristics and comorbidities, disease severity, drug safety profile and accessibility of the drug. In RRMS patients with inadequate treatment response it is recommended to switch to a drug with higher efficacy including natalizumab (Tysabri; Biogen, Cambridge, MA) or fingolimod (Gilenya; Novartis, Pharmaceuticals, Basel, Switzerland)(1).

No randomized clinical trial assessed the comparative efficacy of natalizumab and fingolimod in RRMS patients. Observational studies showed inconsistent results as to difference in clinical effectiveness in real life settings(2–9). Studies varied in sources of data, sample size, inclusion and exclusion criteria, study design, outcomes and statistical analyses. The applied study design may impact the sensitivity of the outcomes studied, as well as the robustness and internal validity of the results. Differences in the studied samples may influence external validity of the results and may reflect variability in reported response to treatment in different patient subgroups (10).

Kalincik et al. (7) used data from the MSBase (11) and reported a higher effectiveness of natalizumab compared to fingolimod in reducing the annualized relapse rate (ARR) and sustained disability improvement in RRMS patients switching from interferon β or glatiramer acetate.

Barbin et al. (8), using the French Multiple Sclerosis Registry (Observatoire Français de la Sclérose en Plaques, OFSEP), supported the finding of higher effectiveness of natalizumab compared to fingolimod on reducing the proportion of at least one relapse within the 1st and 2nd year of treatment, as well as the development of new gadolinium enhancing lesions and new T2-lesions at both 1 and 2 years post treatment baseline.

Conversely, Koch-Henriksen et al. (2) analysed data from the nationwide Danish Multiple Sclerosis Treatment Registry (DMSTR) and found no differences when comparing the effectiveness of natalizumab and fingolimod neither in terms of on-treatment relapse rate, time to first relapse, probability of remaining free of relapse nor changes in expanded disability status scale (EDSS). The three studies also differed to some extent in inclusion/exclusion criteria and methods mitigating indication bias.

The purposes of this study were 1) to re-evaluate the results of each of the three

published head-to-head observational studies by recalculating the results using the original study-specific methods and data (renewed if possible - the replication study); 2) to subject the same data from each database, both separately and pooled, to new analyses using uniform inclusion/exclusion criteria and statistical methods (the validation study). Table 1 shows the differences in the inclusion/exclusion criteria, methods and endpoints used in the three original studies and in the pooled study.

Material and Methods

Study design

This study is a historical cohort study of prospectively collected data, recorded in the three observational MS registries, OFSEP, DMSR and MSBase (2,7,8).

The study consists of three parts: 1) the *replication study* in which the updated cohorts from the three individual studies (2,7,8) were reanalysed using the same methods and criteria as in the original studies; 2) the *validation study* in which the same data were subjected to new and uniform selection criteria, definition of endpoints, and statistical analyses, and 3) the *pooled study* in which data from the three cohorts were pooled and subjected to the same methods and analyses as used for the validation study (see below).

Since publication of the original articles, data in MSBase and DMSR have been updated with more patients and longer follow-up (supplementary table 1a, b), whereas the presently available OFSEP data were unchanged since they were first published (8) (supplementary table 1c).

Data sources

The MSBase Registry is a large international collaboration with patient records from 129 participating MS centres located in 34 different countries (11). The MSBase protocol stipulates update on the minimum data set at least annually, although this was not a required inclusion criterion. The median inter-visit interval is 5 months. The data entry portal was either iMed MS patient record system or the MSBase online data system. An operationalised data quality procedure was applied.(12). The data provided from this register has been updated since the original study (7).

The Danish Multiple Sclerosis Registry (DMSR) (13) was founded in 1956. It comprises data on all patients diagnosed with - or suspected of having - MS by a neurologist.

The diagnostic criteria used before 2005 were the Poser criteria (14) and thereafter the current version of the McDonald criteria (15). Since 1996, acquisition of relapses and EDSS, and of the clinical characteristics has been performed in all DMD-treated patients at baseline, 3 months and thereafter every 6 months during the clinical follow-up with mandatory notification of the DMSR due to reimbursement. Only departments of neurology in the public Danish hospitals are authorized to treat MS patients with DMDs, and treating neurologists are joined in a network enabling use of uniform guidelines. Data collection is done through an online data collection platform, which enables continuous completion of data improving its completeness and validity.

The ObservatoireFrançais de la Sclérose en Plaques (OFSEP) (16) collects information from self-selected patients at 40 centres throughout France, representing more than one half of the French MS population. Minimum standardized datasets are recorded through the EDMUS database and synchronised with the OFSEP database at 6-month intervals. The data provided from this register for this work is the same as used in the original study (8).

Inclusion criteria for the validation study and the pooled study

For the validation and the pooled study the inclusion criteria were: 1) patients have commenced either natalizumab or fingolimod for the first time after 1st of January 2011 (to ensure accessibility of both drugs in Europe and Australia); 2) RRMS at commencing study treatment; 3) continuous treatment with either natalizumab or fingolimod for >3 months; 4) no prior exposure to immunotherapies with extended effect (mitoxantrone, alemtuzumab, ocrelizumab, daclizumab, rituximab, cyclophosphamide or cladribine); 5) no prior participation in any interventional randomised controlled trials; 6) prior DMD treatment of >90 consecutive days within the 12 months immediately prior to commencing natalizumab or fingolimod; 7) sufficient EDSS follow-up (consisting of EDSS recorded -6 months to +1 months of baseline; ≥ 1 EDSS assessment recorded on study therapy and ≥ 1 EDSS assessment recorded ≥ 6 months later (irrespective of the treatment status at that time)). EDSS scores recorded < 30 days after a prior relapse were ignored. Baseline was defined as the day of initiation of natalizumab or fingolimod. Patients' follow-up was censored at discontinuation of the study therapy or the last recorded follow-up. The numbers of eligible patients are presented in Table 2. All three registries used equivalent definitions of the EDSS score as derived from functional score systems described by Kurtzke

(17). Relapse was defined as occurrence of new or worsening neurological symptoms persisting for at least 24 hours in the absence of fever and infection (18) and onset year as the year of first experienced symptom of MS. MSBase and OFSEP had the date and year of onset registered, whereas only year of onset was recorded in the Danish register (date was set to 15/6 in the recorded year of onset).

Study endpoints of the validation study and the pooled study

Annualized relapse rate was calculated at the individual level as the number of relapses divided by annualized observed person-time from baseline to treatment discontinuation or censor date in years.

Time to first relapse was calculated as the time from baseline to the date of start of a first relapse.

Worsening of EDSS was defined as an increase by ≥ 1.5 step sustained for 6 months if EDSS at baseline was 0 or 1; or ≥ 1 step if EDSS at baseline was ≥ 2 and ≤ 5.5; and ≥ 0.5 step if EDSS at baseline was ≥ 6. Improvement of EDSS was defined as a decrease by ≥ 1 EDSS step if EDSS at baseline was ≤ 6 and ≥ 1.5; it was ≥ 0.5 step if EDSS at baseline was > 6; and 1.5 step if EDSS at baseline was 1.5, of which all should be confirmed by EDSS scores recorded over ≥ 6 months.

Statistical analysis: The replication study

Analytical methods as described in the original articles were used with the updated (DMSR and MSBase) and original (OFSEP) cohort.

Statistical analyses: The validation and the pooled studies

Estimation of propensity scores

To control treatment indication bias, we used propensity score. Propensity score is a balanced score representing the probability of being treated with natalizumab (relative to fingolimod), conditional on the patients' baseline clinical and demographic characteristics. It was computed using a multivariable logistic regression based on sex, age, MS duration, EDSS at baseline, number of relapses in the 12 months prior to baseline, disease activity 12 months prior to baseline (classified as relapse, EDSS progression, or both), and the number of previously DMDs commenced prior to baseline. For the MSBase cohort and the pooled cohort, the model was adjusted for country as a random effect.

Using the propensity scores, we calculated the stabilized inverse probability of treatment weighting (sIPTW) (19). Each patient who fulfilled the inclusion criteria was assigned a weight. The weight is proportional to the inverse probability of receiving the treatment that the subject actually received (20) given the individual patient's baseline characteristics,

e.g. a patient treated with natalizumab with a low probability of being treated with natalizumab was assigned a high weight.

Comparison of treatment effectiveness

Demographic and clinical characteristics of patients from either treatment group within each registry as well as the pooled cohort at baseline were reported, including their standardized differences. A difference of ≤ 10% was considered acceptable (19). The propensity score distributions in the two groups were assessed for the degree of overlap, also named the common support.

Annualized relapse rates (ARR) for natalizumab and fingolimod were reported. The counts of relapses between natalizumab and fingolimod in the treated periods were compared using generalized linear models with weighted negative binomial distribution model and with logarithmic transformed length of treatment period as offset. The exponentiated regression coefficients were thereby equivalent to the ratio of relapse rates. Weighted Cox proportional hazards models were used to evaluate the cumulative hazard of 1st relapse as well as 1st EDSS improvement and 1st EDSS worsening. Weighted Andersen-Gill proportional hazards model was used to evaluate the cumulative hazards of multiple events of EDSS worsening and improvement. Robust estimation of variance was used.

Analyses were performed per protocol. The analyses were performed using the R-software (R 3.4.0).

Ethical considerations

The MSBase registry was approved by the Melbourne Health Human Research Ethics Committee and the local ethics committees at participating centres. Enrolled patients provided written informed consent as required. OFSEP was conducted in accordance with the French law relative to clinical noninterventional research according to the French law on Bioethics. The Danish study was conducted according to the Danish laws and correct approvals were obtained with the Center for Data Review applications.

Results

Replication analyses

The replication analysis was a repeat of the original published analyses from MSBase, DMSR, and OFSEP of which MSBase and DMSR had been updated since the original publications. The replication analysis from all three registries confirmed the results of the original articles and are shown in the Appendix.

The validation analyses

Table 2 shows the number of cases included in the three original studies, the number of new study candidates and the eventually included cases from all three databases in the validation study and the pooled study.

Table 3 shows the baseline characteristics for the cases from the three databases, before and after inverse probability weighting (sIPTW). It appears that the weighting reduces imbalance between the natalizumab and the fingolimod arms which is demonstrated by the diminishing standardized mean differences.

The MSBase cohort

This part of the study included 1202 patients, 410 treated with natalizumab and 792 treated with fingolimod (Table 2). The detailed demographic and clinical baseline characteristics are shown in table 3 before and after sIPTW. The results of the present validation analysis from the MSBase using unified methodology showed annualized relapse rates (ARR) of 0.091 for natalizumab and 0.144 for fingolimod. With fingolimod as reference, the ratio of the ARRs was 0.619 (95% CI 0.454 – 0.844; p = 0.0013). The Cox-regression showed a hazard ratio (HR) for a first relapse of 0.61 (p = 0.0032). HR for a first sustained EDSS-worsening was close to unity: 1.08 (p = 0.767), but the Cox analysis of a first sustained EDSS-improvement indicated that natalizumab was associated with a greater chance of decrease in EDSS than fingolimod: HR = 1.89 (p = 0.003). For the full results and confidence intervals see Table 4.

The DMSR cohort

The DMSR cohort included 980 patients, 399 treated with natalizumab and 581 with fingolimod (Table 2). See table 3 for detailed demographic and clinical baseline characteristics before and after sIPTW. In the validation analyses the ARR was 0.178 for natalizumab and 0.151 for fingolimod. With fingolimod as reference, the ratio of the ARRs was 1.115 (p = 0.397). The risks of a first relapse [HR 1.12 (p = 0.359)], first sustained EDSS-worsening [HR 0.97 (p = 0.91)] and first sustained EDSS-improvement [HR 1.11 (p = 0.539)] were comparable between the treatment groups. For confidence intervals and additional analyses: see Table 4.

The OFSEP cohort

This part of the study included 255 patients, 159 treated with natalizumab and 106 with fingolimod (Table 2). See table 3 for detailed demographic and clinical baseline characteristics before and after sIPTW. The validation analysis of the French OFSEP data showed that the ARR was 0.183 with natalizumab and 0.387 with fingolimod. The

ratio of the ARRs was 0.66 (p = 0.002). The HR for a first relapse [0.66 (p=0.111)], first sustained EDSS worsening [0.77 (p=0.645)] and first sustained EDSS-improvement [1.57 (p=0.342)] did not find evidence of statistically significant difference between the two groups. For full results of the analysis and confidence intervals see Table 4.

In summary, with some differences, the present validation analyses for each database showed the same tendency as those published in the three individual original studies (2,7,8), using the new uniform design and statistical analyses, more patients and longer observation times for two of the registry cohorts.

The pooled analysis

The pooled cohort from the three databases consisted of 2447 cases, 968 treated with natalizumab and 1479 treated with fingolimod. In natalizumab-treated patients ARR was 0.138, compared with the ARR 0.165 in fingolimod-treated patients. The ratio of the ARRs was 0.771 (p = 0.004). The risk of a first relapse was lower [HR 0.82 (p=0.0298)] among patient treated with natalizumabs. The cumulative probabilities of a first sustained EDSS improvement [HR 1.40 (p=0.009)]. and of multiple EDSS-improvement events [1.37 (p = 0.011)] were higher on natalizumab. We did not find an evidence for a difference in sustained EDSS-worsening between the treatment groups. For detailed results and confidence intervals see Table 4.

Discussion

Using unified design and methodology, this study reanalysed original and extended clinical data from three different published studies that compared effectiveness of natalizumab and fingolimod in RRMS. The analyses of the pooled cohort confirmed a modest advantage of natalizumab over fingolimod in reducing the risk of relapses and facilitating early recovery from neurological disability. Similar to the original studies, the pooled study did not find a difference in the risk of EDSS worsening between the two therapies.

The original studies from OFSEP and MSBase (7,8) suggested that natalizumab is associated with lower risk of relapses than fingolimod. The study in the MSBase cohort also suggested that natalizumab is associated with a higher probability of recovery from disability. On the other hand, the study from DMSR (2) showed no clinical differences between the effects of the two drugs. These seemingly divergent results can potentially be explained by the differences in the clinical and demographic characteristics of the studied patient. The

OFSEP and the MSBase cohorts were enriched for younger patients with higher prior relapse activity and greater exposure to DMDs prior to their treatment with natalizumab or fingolimod than the DMSR. This suggests that a 'floor effect' exists when one compares effectiveness among highly potent DMDs, and the differences between fingolimod and natalizumab can be seen mainly among patients with more active disease.

Confounding by variables that influence the choice of treatment as well as of short-term disease outcomes is a major concern when comparing treatment arms in non-randomized open-label studies. The three original studies had dealt with this issue using differing statistical methods. The present study used a unified analytical methodology based on a consensus among the investigators - Inverse Probability Treatment Weight - to successfully reduce treatment indication bias. This is reflected by the very close balance of baseline variables between the two treatment arms.

Reassuringly, the validation analyses, applying the consensus analytical methodology to each of the registry cohorts individually, confirmed the results of the original studies.

One can therefore infer that the natural clinical and demographic composition of the studied cohorts is more critical than differences in design and methods.

The reported findings were mainly driven by the MSBase and the DMSR cohorts, which constitute 49% and 40% of the data in the pooled cohort, respectively. The size of the treatment groups in the individual cohorts (with the exception of the fingolimod group in MSBase) decreased as a result of more rigorous inclusion criteria in the unified analyses., This would have a negative impact on the power of the validation analyses before the cohorts were pooled, and the results from the validation analyses with unified analytical design should be interpreted with caution. Reassuringly, these results confirmed the results of the original studies. This study applied unified statistical design to a composite cohort in order to increase internal validity of its results. The pooled cohort consisted of data from a large number of centres in multiple countries, which improves external validity of our conclusions. On the other hand, to mitigate the risk of heterogeneity of the data, we have applied rigorous unified inclusion/exclusion criteria and have included the country of data origin in the estimated sIPTW.

Limitations

Another limitation to this study is the lack of MRI data, either as a baseline or as an endpoint parameter. A recently published guideline (1)

emphasises the advantage of using MRI activity as short- and long-term predictors of disability worsening in RRMS patients. However, two of the original analysis had used (OFSEP) or imputed (MSBase) this information in their analyses. Finally, this study did not compare incidence of adverse events, as this information was not available from all combined registries.

This study, conducted in a large combined cohort from three MS registries, reconciles the results of several previous analyses, and shows that natalizumab is associated with a superior control of relapse activity and improved chance of early recovery from disability among patients with active relapsing-remitting MS. These results are robust to different methods of controlling indication bias. The differences in effectiveness between natalizumab and fingolimod may depend on the characteristics of the studied cohorts. These characteristics warrant further research as they hold the promise of guiding personalised approach to MS therapy.

References

1. Montalban X, Gold R, Thompson AJ, Otero-Romero S, Amato MP, Chandraratna D, et al. ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. Multiple Sclerosis Journal [Internet]. 2018 Feb 20 [cited 2018 Jul 19];96–120. Available from: http://journals.sagepub.com/doi/10.1177/1352458517751049

2. Koch-Henriksen N, Magyari M, Sellebjerg F, Soelberg Sørensen P. A comparison of multiple sclerosis clinical disease activity between patients treated with natalizumab and fingolimod. Mult Scler J [Internet]. 2017;23(2):234–41. Available from: http://journals.sagepub.com/doi/10.1177/1352458516643393

3. Gajofatto A, Bianchi MR, Deotto L, Benedetti MD. Are natalizumab and fingolimod analogous second-line options for the treatment of relapsing-remitting multiple sclerosis? A clinical practice observational study. Eur Neurol. 2014;72:173–80.

4. Carruthers RL, Rotstein DL, Healy BC, Chitnis T, Weiner HL, Buckle GJ. An observational comparison of natalizumab vs. fingolimod using JCV serology to determine therapy. Mult Scler. 2014;20(10):1381–90.

5. Guger M, Enzinger C, Leutmezer F, Kraus J, Kalcher S, Kvas E, et al. Real-life clinical use of natalizumab and fingolimod in Austria. Acta Neurol Scand. 2018;137(2):181–7.

6. Braune S, Lang M, Bergmann A. Second line use of Fingolimod is as effective as Natalizumab in a German out-patient RRMS-cohort. J Neurol [Internet]. 2013 Dec 6 [cited 2018 Jul 9];260(12):2981–5. Available from: http://link.springer.com/10.1007/s00415-013-7082-0

7. Kalincik T, Horakova D, Spelman T, Jokubaitis V, Trojano M, Lugaresi A, et al. Switch to natalizumab versus fingolimod in active relapsing-remitting multiple sclerosis. Ann Neurol [Internet]. 2015 Mar [cited 2018 Jul 6];77(3):425–35. Available from: http://doi.wiley.com/10.1002/ana.24339

8. Barbin L, Rousseau C, Jousset N, Casey R, Debouverie M, Vukusic S, et al. Comparative efficacy of fingolimod vs natalizumab. Neurology [Internet]. 2016 Feb 23 [cited 2018 Jul 6];86(8):771–8. Available from: http://www.neurology.org/lookup/doi/10.1212/WNL.0000000000002395

9. Lorscheider J, Benkert P, Lienert C, Hänni P, Derfuss T, Kuhle J, et al. Comparative analysis of natalizumab versus fingolimod as second-line treatment in relapsing–remitting multiple sclerosis. Mult Scler J. 2018;24(6):777–85.

10. Kalincik T, Butzkueven H. Observational data: Understanding the real MS world. Mult Scler. 2016;22(13):1642–8.

11. Butzkueven H. Building a global disease cohort for real-work data exploration: MSBase. J Neurol Sci [Internet]. 2017 Oct 15 [cited 2018 Jul 6];381:52. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022510X17307049

12. Kalincik T, Kuhle J, Pucci E, Rojas JI, Tsolaki M, Sirbu CA, et al. Data quality evaluation for observational multiple sclerosis registries. Mult Scler J. 2017;23(5):647–55.

13. Koch-Henriksen N, Magyari M, Laursen B. Registers of multiple sclerosis in Denmark [Internet]. Vol. 132, Acta Neurologica Scandinavica. Wiley/Blackwell (10.1111); 2015 [cited 2018 Aug 2]. p. 4–10. Available from: http://doi.wiley.com/10.1111/ane.12424

14. 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;

15. Polman CH, Reingold SC, Edan G, Filippi M, Hartung H-P, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald Criteria.” Ann Neurol. 2005;

16. Vukusic S. Observatoire de la sclérose en plaques (OFSEP). [Internet]. [cited 2018 Jul 6]. Available from: http://www.ofsep.org/fr/la-cohorte-ofsep/descriptif-de-la-cohorte

17. KURTZKE JF. Disability Rating Scales in Multiple Sclerosis. Ann N Y Acad Sci [Internet]. 1984 Dec [cited 2018 Jul 9];436(1 Multiple Scle):347–60. Available from: http://doi.wiley.com/10.1111/j.1749-6632.1984.tb14805.x

18. Schumacher GA, Beebe G, Kibler RF, Kurland LT, Kurtzke JF, McDowell F, et al. Problems of Experimental Trials of Therapy in Multiple Sclerosis: Report By the Panel on the Evaluation of Experimental Trials of Therapy in Multiple Sclerosis. Ann N Y Acad Sci. 1965;122(1):552–68.

19. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med [Internet]. 2015 Dec 10 [cited 2018 Jul 11];34(28):3661–79. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26238958

20. Austin P. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399–424.

Table 1. Differences in methods for the original studies and the present validation study and pooled study

	MSBase 2015 (7)	**OFSEP** (8)	**DMSTR 2016** (2)	**Present validation study and pooled study**
Number of centers	66	27	14	174
Design	Cohort, longitudinal data.	Cohort, longitudinal data.	Cohort, longitudinal data.	Cohort, longitudinal data.
Inclusion/exclusion	Relapse or disability worsening within 6 months before start; No previous participation in randomized trials	RRMS. Age 18 to 65. EDSS ≤ 5.5	RRMS ≥1 relapse within 12 months before start or, if treatment naïve, else ≥2 relapses with residual symptoms	RRMS; > 90 days of DMD first-line treatment prior to study medication; > 3 months of study treatment; no previous participation in randomized trials;
Propensity score: Matching or weighting	Matched by ropensity score based on age; sex; number of relapses in 6 or 12 months EDSS; Disease activity under previous treatment (relapses, EDSS-worsening or both). MRI data available from a proportion of patients, multiple imputation used.	Weighted by inverse probability of treatment (IPTW) based on sex; number of relapses in previous year; EDSS; hospital; Gd-enhancing lesions on MRI.	Matched by propensity score based on sex; age; being treatment naïve; ARR during previous treatment; MSSS (derived from EDSS) with ignoring unmatchable cases. No MRI data available for matching.	Weighted by stabilized inverse probability of treatment (IPTW) based on sex; age; MS duration; EDSS; #relapses in 12 months; disease activity in 12 months (relapses, EDSS-worsening or both).
Statistical analyses	Adjusted paired proportional hazards models and weighted negative binomial model	*t* test; Wilcoxon test; chi-square.	Generalized linear models assuming negative binominal distribution; Kaplan-Meyer analysis; Mann Whitney U test; Pearson chi-square.	Negative binominal model; Cox proportional Hazards; Anderson-Gill model
Follow-up	December 2013	July 2014	October 2015	-
Clinical study endpoints	Freedom from clinical relapses. ARR. Disability worsening. Disability improvement.	Proportion of patients with at least one on-study relapse in the first year and at two years.	ARR; proportion of patients remaining free of relapses; time to 1st relapse; proportion with worsening or improving EDSS.	ARR; time to 1st relapse; increase in EDSS sustained for 6 months; improvement of EDSS sustained for 6 months

Table 2. Number of cases provided from the registers in the original studies, updated study candidates, and finally included number of patients

	MSBase		DMSR		OFSEP	
	Natalizumab	Fingolimod	Natalizumab	Fingolimod	Natalizumab	Fingolimod
Original studies (2,7,8)	560	232	520	789	303	326
Updated numbers of study candidates	690	610	793	1057	303	326
Eventually included*	410	792	399	581	159	106

*according to the uniform inclusion/exclusion criteria in the validation study and the pooled study.

Table 3. Baseline characteristics of the pooled cohort and the three individual cohorts contributing to it before and after stabilized inverse probability of treatment weighting (sIPTW)

	Before IPTW			After IPTW		
	Natalizumab	**Fingolimod**		**Natalizumab**	**Fingolimod**	
MSBase (international)	N = 410	N = 792	SMD*			SMD*
Female %	73.17	72.85	0.0071	72.02	72.8	0.017
Mean age at baseline,years (sd)	36.2 (10.3)	38.1 (9.6)	0.183	37.6 (11.1)	37.6 (9.5)	0.003
Mean MS duration, years (sd)	8.1 (6.58)	9.2 (7.18)	0.151	9.1 (7.95)	8.9 (7.00)	0.025
Mean EDSS at baseline, (sd)	2.9 (1.52)	2.28 (1.51)	0.411	2.66 (1.42)	2.49 (1.58)	0.113
Mean #relapses 12 months prior to baseline	1.35 (0.96)	0.94 (0.84)	0.447	1.14 (0.88)	1.07 (0.90)	0.789
Mean nr of previous DMDs, (sd)	1.59(0.80)	1.67 (0.88)	0.106	1.67 (0.88)	1.67 (0.87)	0.082
Disease activity 12 months prior to baseline%						
None	13.66	27.02	0.337	18.92	22.86	0.097
EDSSworsening	3.90	5.56	0.078	4.36	5.25	0.042
Relapse	51.95	45.2	0.135	49.06	46.16	0.058
Relapse and EDSS worsening	30.49	22.22	0.188	27.66	25.73	0.044
DMSR (Denmark)	N = 399	N = 581				
Female %	70.93	65.06		67.32	67.40	
Mean age at baseline (sd)	39.2 (9.5)	40.4 (9.2)	0.131	39.9 (9.5)	39.9 (9.3)	0.001
Mean MS duration, years (sd)	8.8 (7.43)	8.9 (6.72)	0.025	8.8 (7.56)	8.8 (6.63)	0.003
Mean EDSS at baseline	2.90 (1.59)	2.63 (1.46)	0.171	2.73 (1.56)	2.74 (1.50)	0.005
Mean #relapses 12 months prior to baseline	0.76 (0.84)	0.71 (0.75)	0.072	0.73 (0.80)	0.73 (0.78)	0.001
Mean #previous DMDs	1.61 (0.95)	1.51 (0.76)	0.117	1.56 (0.87)	1.55 (0.79)	0.005
Disease activity 12 months prior to baseline%						
None	25.81	30.98	0.115	28.99	28.97	0.001
Worsening	18.30	13.77	0.124	15.36	15.41	0.002
Relapse	28.32	30.98	0.058	30.10	30.02	0.002
Relapse and worsening	27.57	24.27	0.075	25.55	25.60	0.001
OFSEP (France)	N = 159	N = 106				
Female %	76.73	73.58	0.073	74.89	76.7	0.042
Mean age at baseline (sd)	37.1(10.2)	39.1 (9.2)	0.198	37.9(10.4)	37.8 (9.5)	0.023
Mean MS duration (years)	8.0 (5.42)	9.8 (6.85)	0.297	8.7 (5.76)	8.6 (6.32)	0.015
Mean EDSS at baseline	2.82 (1.58)	2.61 (1.67)	0.131	2.77 (1.54)	2.85 (1.66)	0.049
Mean #relapses 12 months prior to baseline	1.62 (1.07)	0.99 (0.93)	0.623	1.38 (1.06)	1.41 (1.1)	0.029
Mean #previous DMDs	1.69 (0.89)	1.58 (0.87)	0.11.4	1.66 (0.89)	1.68 (0.9)	0.023
Disease activity 12 months prior to baseline%						
None	6.92	25.47	0.520	14.09	14.03	0.002
Worsening	3.77	7.55	0.164	5.12	5.10	0.001
Relapse	45.91	44.34	0.032	45.29	44.31	0.020
Relapse and worsening	43.4	22.64	0.453	35.49	36.56	0.022
Pooled cohort (MSBase+DMSR+OFSEP)	N = 968	N = 1479				
Female %	72.8	69.8	0.066	70.3	71.0	0.015
Mean age at baseline (sd)	37.6 (10.0)	39.1 (9.5)	0.150	38.8 (10.5)	38.6 (9.5)	0.022
Mean MS duration, years (range)	8.4 (6.77)	9.1 (6.98)	0.110	9.0 (7.75)	8.9 (6.81)	0.026
Mean EDSS at baseline	2.89(1.56)	2.44 (1.51)	0.289	2.71 (1.50)	2.65 (1.57)	0.036
Mean #relapses 12 months prior to baseline	1.15 (0.99)	0.85 (0.82)	0.327	0.99 (0.91)	0.98 (0.91)	0.017
Mean #previous DMDs	1.61 (0.88)	1.6 (0.84)	0.011	1.61 (0.84)	1.62 (0.84)	0.004
Disease activity 12 months prior to baseline%						
None	17.56	28.47	0.261	23.12	24.29	0.027
Worsening	9.81	8.92	0.031	8.97	9.41	0.015
Relapse	41.22	39.55	0.034	40.64	39.51	0.023
Relapse and worsening	31.40	23.06	0.188	27.27	26.79	0.011

*Standardized mean difference (difference of the pooled standard deviation)

Table 4. Results of validation analyses based on weighted (sIPTW) data in the pooled cohort and the three individual cohorts contributing to it

		Pooled cohort N=2447	**MSBase N=1202**	**DMSTR N=980**	**OFSEP N=265**
Annualized relapse rate	Natalizumab	0.138	0.091	0.178	0.183
	[95% CI]	[0.116; 0.160]	[0.064; 0.117]	[0.138; 0.217]	[0.135; 0.234]
	Fingolimod	0.165	0.144	0.151	0.387
	[95% CI]	[0.144; 0.186]	[0.122; 0.166]	[0.124; 0.177]	[0.208; 0.566]
Difference of means (FTY minus NAT)		0.026	0.053	-0.027	0.204
	[95% CI]	[-0.004; 0.057]	[0.019; 0.087]	[0.0.74; -0.021]	[0.019, 0.388]
Ratio of counts of relapses given observation time§		0.771	0.619	1.115	0.466
	[95% CI]	[0.642; 0.926]	[0.454-0.844]	[0.866; 1.436]	[0.282; 0.761]
	p-value	*0.004*	*0.0013*	*0.397*	*0.002*
Hazard Ratio* for a first relapse		0.82	0.61	1.12	0.66
	95% CI	[0.68; 0.98]	[0.44; 0.85]	[0.88; 1.43]	[0.40; 1.10]
	p value	*0.0298*	0.0032	*0.359*	*0.111*
Hazard Ratio* for a first sustained EDSS-worsening		1.13	1.08	0.97	0.77
	95% CI	[0.83; 1.53]	[0.63; 1.85]	[0.63; 1.51]	[0.26; 2.30]
	p value	*0.438*	*0.767*	*0.910*	*0.645*
Hazard Ratio* for a first sustained EDSS-improvement		1.40	1.89	1.11	1.57
	95% CI	[1.08; 1.80]	[1.24; 2.88]	[0.79; 1.57]	[0.62; 3.96]
	p value	0.009	0.003	0.539	0.342
Ratio* of cumulative hazards of multiple events of EDSS-worsening		1.10	1.06	0.94	0.79
	95% CI	[0.82; 1.46]	[0.64; 1.75]	[0.62; 1.41]	[0.26; 2.34]
	p value	0.528	0.814	0.745	0.669
Ratio* of cumulative hazards of multiple events of EDSS-improvement		1.37	1.89	1.09	1.69
	95% CI	[1.08; 1.76]	[1.25; 2.86]	[0.78; 1.51]	[0.68; 4.20]
	p value	0.011	0.002	0.624	0.259

*with Fingolimod as reference

§exponentiated regression coefficient of count of relapses with logarithmic transformed observation time as offset.

Understanding heterogeneity in comparative studies of treatment effectiveness in multiple sclerosis - Part 2: Effect of analytical methodology

Authors:

Mathilde Lefort*[(1)(2)]; Sifat Sharmin*[(3)]; Johanna Andersen*[(4)]; Helmut Butzkueven[(5)(6)(7),] David-Axel Laplaud[(8)(9)(10)], Nils Koch-Henriksen[(4)], Sandra Vukusic[(11-14)]; Melinda Magyari**[(4)]; Tomas Kalincik**[(3)]; Emmanuelle Leray**[(1)(2)(15)]; on behalf of MSBase, Danish MS Registry, and OFSEP.

** Equal contribution as first author; **Equal contribution as last author*

Affiliations:

[(1)]Rennes University, EHESP, REPERES – EA 7449, Rennes, France

[(2)]Univ Rennes, CHU Rennes, Inserm, CIC 1414, Rennes, France

[(3)]CORe, Department of Medicine, University of Melbourne, Melbourne, Australia; Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia

[(4)]Department of Neurology, Danish Multiple Sclerosis Center and the Danish Multiple Sclerosis Registry, Rigshospitalet and University of Copenhagen, Copenhagen, Denmark

[(5)]Central Clinical School, Monash University, Melbourne, Australia;

[(6)]Department of Neurology, The Alfred Hospital, Melbourne, Australia;

[(7)]Department of Neurology, Box Hill Hospital, Monash University, Melbourne, Australia;

[(8)]Department of Neurology, CHU de Nantes, Nantes, France

[(9)]Univ Nantes, CHU Nantes, Inserm CIC 015, Nantes, France

[(10)]Centre de Recherche en Transplantation et Immunologie (CRTI), Inserm, UMR 1064, nantes, France

[(11)]Hospices Civils de Lyon, Service de Neurologie, sclérose en plaques, pathologies de la myéline et neuro-inflammation, F-69677 Bron, France

[(12)]Observatoire Français de la Sclérose en Plaques, Centre de Recherche en Neurosciences de Lyon, INSERM 1028 et CNRS UMR 5292, F-69003 Lyon, France

[(13)]Université de Lyon, Université Claude Bernard Lyon 1, F-69000 Lyon, France

[(14 Eugène Devic EDMUS Foundation against multiple sclerosis, state-approved foundation, F-69677 Bron, France

[(15)]METIS Department, EHESP French school of public health, Rennes, France

Abstract

Introduction

Natalizumab and fingolimod are two second-line treatments used in patients with relapsing-remitting multiple sclerosis (MS). They present similar indication but important difference in terms of safety. Several studies comparing their effectiveness showed inconsistent results. Previous studies were based on different inclusion criteria and methods, especially in control for indication bias and management of censoring in time-to-event analysis. The objective of the present study was to evaluate the impact of statistical methods on the results.

Patients and methods

This study was based on three observational MS registries (international MS Base, Danish MS register and French OFSEP registry). Four outcomes were studied: count of relapses, time to first relapse, time to first disability worsening and improvement. Two propensity scores methods were used: matching and weighting allowing for estimating average treatment effect for treated and for the entire population. Analyses were conducted in intention-to-treat and per-protocol frameworks.

Results

Overall 5,148 patients were included. Whatever the methods used, conclusions were similar as confidence intervals overlapped, and in line with the pooled analysis. Censoring management led to slight differences. Weighting and matching gave consistent results although they provided two different types of average treatment effect. In matching method, changes in the caliper value did not have much impact on the results.

Conclusion

This study provided several estimates of treatment effects based on different statistical strategies. Moreover, we discuss the practices using propensity score and censoring. To conclude, we recommend describing precisely the methodology as it affects the conclusion.

Keywords: Propensity score, indication bias, effectiveness, multiple sclerosis, censoring

Introduction

Natalizumab [1,2] and fingolimod [3,4] are two second-line treatments used in patients with Relapsing Remitting Multiple Sclerosis (RRMS). Their efficacy on reduction of relapses, disability progression and MRI activity was demonstrated through randomized clinical trials (RCT). The most serious adverse effect with natalizumab is the high risk of Progressive Multifocal Leukoencephalopathy while this risk is much lower with fingolimod[5]. This important difference in terms of safety justified the comparison of their effectiveness.

Observational studies comparing their effectiveness showed inconsistent results. Some of them showed that fingolimod was as effective as natalizumab [6] and others showed superiority of natalizumab in comparison with fingolimod [7–12]. We have shown that at least some of this variability is attributed to differences in the studied populations [13]. In the present work, we focused on three studies which have provided different results, and which used data from three multiple sclerosis (MS) registries, with differences in inclusion/exclusion criteria and methods [6,10,14]. In this work, we will focus on the impact of methodological choices on the results - in particular, the methods to control indication bias and to manage censoring in time-to-event analysis.

Indication bias is one of the most important biases in studies comparing therapies in observational data that it occurs when the risk is related to the indication for treatment use but not the use of the treatment itself [15]. In the three original studies comparing fingolimod and natalizumab, propensity score (PS) were used to control indication bias. PS is a summary score that represents known determinants of treatment allocation, which aims to balance known potential confounders of treatment effect between the two groups [16] to compare. PS can be utilized in several ways [17]. Two of our three original studies used matching [6,10] while the other study used weighted models of treatment effect [14].

Attrition bias and informative censoring result from systematic differences in the duration of follow-up between cohorts. Two approaches to the analysis of follow-up information exist. On one hand, per-protocol approach uses censoring at the treatment cessation or last recorded clinical information, whichever occurs first, and was used in two of our three studies[6,10]. On the other hand, intention-to-treat approach uses the date of last recorded clinical information as the censoring event,

irrespective of treatment status. This was used by one of our three studies [14]. Moreover, one of the three studies used pairwise censoring that consists of censoring data within each matched pair to the shorter of the recorded follow-up times within the pair [18].

We hypothesise that methodological choices may affect the results of PS-based comparisons of therapies. The objective of this empirical study is to evaluate the impact of statistical methodology on the results of comparison of therapies, using the example of natalizumab and fingolimod in multiple sclerosis and data from our three previous studies.

Patients and methods

Data source and approvals

This study used data from three observational MS registries: MSBase, the international cohort from 129 MS expert centres in 34 countries [extraction date: May 1st, 2018] [19]; Danish multiple sclerosis register (DMSR) [extraction date: December 31st, 2017] [20]; and OFSEP (« Observatoire Français de la Sclérose en plaque »), the French national cohort from 40 French expert MS centers [extraction date: July 11th, 2014] [21].

Ethical considerations

The MSBase registry was approved by the Melbourne Health Human Research Ethics Committee and the local ethics committees at participating centres. The Danish study was conducted according to the Danish laws and correct approvals were obtained with the Center for Data Review applications. Data confidentiality and safety are ensured according to the recommendations of the French Commission Nationale Informatique et Libertés (CNIL). OFSEP has received approval for storing clinical, biological and imaging data for research purpose. Patients give informed consent for their data to be stored in the database and used for research, in France and abroad.

Study population

All patients diagnosed with RRMS, starting natalizumab or fingolimod for the first time on or after the 1st January 2011 and treated for at least 90 days were included in our study, except if they had received a treatment from the excluded treatment list (alemtuzumab, haematopoietic stem cell transplantation, mitoxantrone, ocrelizumab, rituximab, cladribine, daclizumab, cyclophosphamide) or enrolled in a randomized clinical trial. Several criteria regarding follow-up were required: a

recorded visit within six months before or one month after treatment start with disability assessment (the baseline visit), one follow-up visit while on natalizumab or fingolimod and at least two follow-up visits after the start of index therapy, at least 180 days apart.

Outcomes

The outcomes of interest were: (1) The count of relapses, (2) the time to first relapse, (3) the time to first disability worsening and (4) the time to first disability improvement from treatment start (baseline) to censoring. Baseline was treatment initiation. End of study or period of assessment for count outcome depended of the management of censoring.

Disability was measured using the Expanded Disability Status Scale (EDSS)[22]. Disability worsening was defined as an increase by at least 1.5 point of EDSS when EDSS at baseline was 0 or 1, at least 1 point when EDSS was between 2 and 6, and at least 0.5 point when EDSS was more than 6, confirmed over at least 6 months and sustained for the remainder of the follow-up. Definition of disability worsening improvement followed the same rules, but with a decrease in EDSS scores.

Propensity score matching and weighting

PS is a balancing score defined as the probability of being assigned to a given treatment conditionally on observed baseline variables [16,23]. PS makes four assumptions: consistency, exchangeability, positivity, and no misspecification of the propensity score model. Consistency means that a patient's potential outcome under the treatment actually received is equal to the patient's observed outcome [24]. Exchangeability refers to the fact that there are no unmeasured confounders. Positivity assumption refers to the fact that all patients have a non-zero probability of receiving either treatment. Misspecification refers to error in the PS model such as omitted variable, irrelevant variable or dependant variable included, etc.

Using PS, two treatment effects can be estimated: the average treatment effect for the treated group which is the average treatment effect on those treated patients, or the average treatment effect for the entire population which is the average effect on population level [25]. In the present work, we were interested in estimating both effects.

PS was estimated using a logistic regression based on the following baseline characteristics: sex, age, MS duration (from first MS symptom to baseline), EDSS score, number of relapses in the past 12 months, number of previous treatment starts, and the nature of clinical activity recorded during the previous 12 months (disability progression, relapses, both or no activity). Country was added as random effect.

Propensity score matching

One-to-one, greedy, nearest neighbour, random matching was used. Matching caliper values of 0.1 (used in the original studies), 0.2 (recommended by literature [26]) and 0.02 standard deviations of the PS (to prioritise close matching) were used. PS matching allows for estimating the average treatment effect for treated only.

Propensity score weighting

Two weighting procedures were evaluated: weighting with Inverse Probability of Treatment Weighting (IPTW) and odds. Using IPTW, the weight for a treated patient and for a control are respectively defined as follows $w_i = \frac{1}{p_i}$ and $w_i = \frac{1}{1-p_i}$ where p_i is the PS for a patient i. For IPTW, the weights were stabilized by multiplying them by the marginal probability of receiving the treatment actually received, to manage the instability due to very large weights [27]. The stabilized IPTW is referred to as sIPTW. IPTW as well as sIPTW allow for estimating average treatment effect for the entire population. Using odds, the weight for a treated patient is 1 and the weight for control is defined as follows: $w_i = \frac{p_i}{1-p_i}$. Odds allow for estimating the average treatment effect for treated.

Management of censoring

Intention-to-treat analysis includes all matched or weighted patients in the group of initial treatment until the earliest between last clinical information date and outcome date, regardless of their following exposure. Per-protocol analysis includes all matched or weighted patients until the earliest between treatment stop date, last clinical information date and outcome date. Pairwise-censoring is another technique which can be associated with matching only. In each pair, patient's follow-up is censored at the shortest follow-up time of the pair.

Statistical analysis

Characteristics of the patients included in the analyses as well as those excluded by the matching procedure were described – overall and by treatment groups, and before and after propensity score matching/weighting. Standardized mean differences or Mahalanobis distances were computed for all variables, with 10% considered to be an acceptable difference [24]. Incidence of relapses was evaluated using a negative binomial model. To take into account the duration of the follow-up, an offset term was used. For the cumulative hazards of first relapse, first EDSS improvement and first EDSS worsening, Cox proportional hazards models were used. Proportional hazards assumption was

assessed with Schoenfeld residuals test and violation of the assumption was corrected by introducing an interaction term with a time variable.

The probability of disability worsening and improvement events is associated with the frequency of EDSS scores. Therefore, time to disability outcomes models were adjusted for annualized visit density [28]. Robust estimation of variance was used [29]. Moreover, to take into account the pairing of patients within the matched data, a cluster term was used in the negative binomial model and frailty terms were used in the Cox models.

All analyses were conducted in both the intention-to-treat and the per-protocol frameworks. Analysis based on matching were competed in the pairwise-censoring frameworks. The analyses were performed using R-software (R 3.4.0).

Table 1 gives an overview of all the analytical approaches considered in the present work.

Results

Patients' characteristics

Overall, 5,148 patients were included in this study (Figure 1). The median follow-up duration from treatment start to the last clinical visit was 3.1 years (interquartile range (IQR): 2.0-4.5). Of the included patients, 1989 (39%) patients were treated with natalizumab and 3159 with fingolimod.

Patient's characteristics are described in Table 1. The median age at treatment start was 37.7 years with a median MS duration of 6.9 years. Most of patients had an active disease as shown by the number of relapses in the 12 months before treatment start. Seventy percent of patients had an EDSS score at treatment equal or superior to 2. Most of patients had been previously treated by another treatment before natalizumab or fingolimod start.

The median duration of follow-up per groups are shown in Table 2. The median duration of natalizumab and fingolimod treatments were 2.00 (1.3-3.1) and 2.2 (1.2-3.6) respectively (p=0.002).

Patient's characteristics after balancing procedures (matching or weighting)

The PS was defined as the probability of being treated with natalizumab, conditional on patient and disease characteristics at treatment start. The distributions of PS showed a good overlap between the two groups, except in the tails, as shown in Figure 2.

The use of three different caliper values for PS-matching led to three similar matched data sets (Table 1). The characteristics of the matched groups were comparable to the

characteristics of the overall studied sample. Between 1890 and 1916 patients were excluded, depending of the caliper value, with the largest numbers of patients excluded with the smallest caliper. The excluded patients tended to experience less disease activity.

Table 3 presents patients' characteristics by treatment group. Overall, 35% of patients treated with fingolimod had a low EDSS score at treatment start while it was 22% in the group treated by natalizumab. The matching procedure improved the balance between the compared groups for some variables but taking into account the covariance with Mahalanobis distance show quite high value for register and number of previous MS treatments. Moreover, matching favoured inclusion of patients with active disease.

The Table 4 presents patient characteristics by treatment group after weighting on sIPTW or odds. The two groups were well balanced, with standardized mean differences or Mahalanobis distances around 10% for all patient characteristics, except for the number of previous MS treatments, as natalizumab tended to be prescribed as first-line treatment more frequently than fingolimod).

Note that, in intention-to-treat analysis, data from time when patients switched to other therapies from fingolimod or natalizumab were also included in the analysis. Exposure after first natalizumab- or fingolimod-exposure is shown in Table S1.

Comparison of treatment effectiveness between natalizumab and fingolimod

Figure 3 summarises the results of all analyses comparing effectiveness of natalizumab and fingolimod. Whether it was using intention-to-treat analysis, on treatment or pairwise censoring, all confidence intervals overlapped, meaning that the different analyses led to similar conclusions even if the value of the punctual estimate could vary.

For all outcomes except time to first EDSS worsening, intention-to-treat analyses give deflated results in comparison to per-protocol or pairwise-censoring analysis. In particular, studying count of relapses, the intention-to-treat analysis using matching gave differential results to the rest of the analyses leading to significant result in favour of natalizumab using weighting, but non-significant difference using matching. Per-protocol or pairwise censoring analyses confirmed the significant result in favour of natalizumab.

Per-protocol analysis and pairwise censoring analysis gave similar results although the width of the confidence interval was in some cases impacted. Regarding count of relapses outcome, confidence interval was smaller while there

were wider for disability outcome. Regarding time to first relapse, it looks stable. This stability was also observed for intention-to-treat analysis and comparing weighting and matching. Thus, all analyses led to significant results in favour of natalizumab in delay first relapse after treatment start.

All results appeared non-significant for the time to first EDSS worsening.

Intention-to-treat looks more sensitive than per-protocol or pairwise censoring giving unstable results according to the balance method in particular regarding count of relapses and first EDSS improvement. That instability led to unclear conclusion about first EDSS improvement using intention-to-treat. Per-protocol analysis showed that natalizumab was more beneficial for disability improvement. Results were non-significant using pairwise-censoring.

Discussion

Comparison of natalizumab and fingolimod effectiveness was performed by several groups in the past[6–11]. Two main methodological differences were identified, i.e. the way to control confounding by using propensity score method and the way to manage censoring. In this empirical study, we explored how methodological choices may influence the results of comparative effectiveness studies in a complex chronic disease with long-term follow-up and several measured disease outcomes.

All patients treated by natalizumab or fingolimod regardless the therapeutic context (switch, treatment naïve, etc.) were studied. Overall, whatever the methods used, conclusions were similar as confidence intervals overlapped and in keeping with the original studies and the pooled analysis [9,10,13]. Nevertheless, some variations in punctual values have been observed. Most differences were related to the censoring management since intention-to-treat gave unstable and often deflated results than the per-protocol analysis. These fluctuations were observed when outcome was count of events and time to medium-term events but not when studying time to short-term event (relapse). Per-protocol and pairwise censoring analyses gave similar results although the confidence interval's width could indicate a lack of power using pairwise censoring. Weighting and matching procedures give consistent results confirming the good performance of both methods although the most important difference concerned the type of effect[17] which is estimated. The caliper value did not have much impact on the results confirming than 0.2 is good enough[26]. Moreover, a smaller caliper would result in

increasing the number of excluded patients (as unmatched) which can contribute to a loss of statistical power.

In this work, both intention-to-treat and per-protocol analyses were performed. Intention-to-treat analysis tends to reduce the measure of treatment effectiveness. This observation was expected since intention-to-treat is a conservative approach. This feature is an asset in the analysis of superiority since we seek to avoid accepting the assumption of superiority wrongly. Patients and neurologists may be more interested in the per-protocol effect than in the intention-to-treat one. Nevertheless, per-protocol results should be interpreted with caution. Indeed, there is a risk of attrition bias and informed censoring in the per- protocol approach. In particular, per-protocol analysis led to compare only patients who adhere to their assigned treatment[30]. That could led to selection bias and so, impact the comparability[31]. Moreover, a second drawback of the per protocol analysis is the selection of "treatment responders" [32] since patients who well respond to treatment are likely to remain treated, which biased the result in favor of the treatment. In an intention-to-treat approach, the strength of the randomization is preserved since patients are analyzed according to the treatment group assigned at baseline and treatment stop or switch ignored. Methodology is currently under development to take into account pre- and post-randomization variables and so improve per-protocol analysis [33]. Moreover, marginal structural models enable continuous re-evaluation for time-dependent confounders after treatment has been allocated would be a solution[34]. In this study, we showed that the fluctuation in intention-to-treat and per-protocol results were different according the outcomes. In particular, time to short and frequent event such as relapse were less impacted by the censoring management since the studied period were closer.

In this study, the pairwise censoring procedure was also used. It present similar limitations to the per-protocol analysis since the comparability could be impacted by all actions performed after baseline. Moreover, the delay before reaching outcome or censoring was already took into account in the Cox model.

Four methods using PS exist: stratification, matching, weighting and adjustment. It has been demonstrated that matching and weighting induced greater balance on baseline characteristics than stratification or adjustment [35], the reason why we did not consider these two procedures in the present work. Moreover, these methods presented better performances in terms of bias minimization and variance estimation [36–38]. The stability of results in our study seems

argue in this way. Nevertheless, these two methods are not equivalent, in particular in term of estimated average treatment effect[39]. By construction, the matched sample was similar to the natalizumab-treated population. It was also the case using weighting by the odds since all treated patient got a weight of one. Using IPTW, the weighted sample was similar to the entire study population. In particular, despite similar indications of these two treatments, the population treated by natalizumab and the one treated by fingolimod were different, that could explain some differences in the punctual effect. Nevertheless, the result got using weighting by the odds were often inflated in comparison to the ones got with matching. Even if the same effect is estimated, matching led to exclude patient that could explain this small difference. Indeed, although matching allowed for approaching average treatment effect for treated, due to exclusion, matching did not allow for estimating the average treatment effect for "all" treated. In the context of exclusion in matching, the comparability between the matched cohort and the target population was checked. The context and aim of a given research should determine which approach and so which effect provides greater utility and is the most appropriate. In our context, we explore all methods but estimate the average treatment effect for the entire population seem more appropriate to our questions. Indeed, both treatments could be prescribed without any restriction such as a restricted number of drugs available. Since all the target population could be treated by one of the two treatments or the other one, there is no apparent reason to restrict the analysis to the natalizumab-treated patients or fingolimod-treated patients.

PS is a frequently used tool for estimating the effects of treatments when using observational data. The methodology relative to PS responds to good practices, in particular to test or discuss the four PS assumptions. Although some of them, such as consistency and exchangeability, cannot be tested, the violation of these assumptions should be limited. A key point is the definition of the inclusion criteria. In one of the 3 original studies, there was no limitation in term of study period [10]. However, including patients treated by natalizumab before fingolimod availability makes the positivity assumption not respected. That is why in the present study, inclusion period was defined from fingolimod availability date. Regarding variables to be used in the balance procedure, there is a lack of consensus but it has been recommended to consider variables associated with the outcome in the PS [40,41]. It has also been suggested to include all measured baseline

characteristics [42], this was done in two of the 3 studies [10,14], as well as in the present work. Moreover, propensity score should include only variables that are measured at baseline and not post baseline [17].

The use of PS does not guarantee the comparability of the two groups. Balance and common support assessments are required. In the three previous studies, usual comparison tests or standardized mean differences such as Cohen's distance were performed. It is now recommended to compare the two treatment groups using standardized differences [17] as statistical significance testing was judged as a limitation approach in the context of balance diagnostic [43]. In the present study, standardized mean differences showed good balance between natalizumab and fingolimod groups giving confidence in balance on measured covariates. The register variable using matching gave high Mahalanobis distance using matching whereas values were low using weighting. The common support is assessed according to the degree of overlap between the two distribution of propensity scores [27]. Our study present a good overlap. Nevertheless, no fingolimod-treated patients had very high propensity of being natalizumab-treated probably linked to a different used of treatments.

From the methodological perspective, the limitations of this study include the heterogeneity of drug availability included in the MS Base cohort. Another limitation concern the choice of variables included in the PS. The inclusion of all relevant variables could lead to include potential instrumental variables that it is not recommended. From the perspective of the subject of the analysis, this study was limited by unavailability of MRI data universally across the studied cohorts. MRI is a known prognostic marker in MS [45]. The previous studies that included MRI information or combination of available MRI information with multiple imputation of missing MRI data showed similar results [9,10]. Finally, this study combined data from multiple specialist MS centres. Heterogeneity of data in multisite registries at multiple countries (with differences in regional practice, health care systems and treatment access) may represent a problem [46]. We have mitigated the potential heterogeneity in the present dataset by including country as a random term in the models of propensity score.

This study provided several estimates of treatment effects based on different statistical strategies confirming that the method affect the results. Note that, this study did not allow for comparing performance and robustness of methods that

can be done only with simulation studies. The difference in literature could be understood giving more details in methods section and explanation of these choices. To conclude, we recommend describing precisely the methodology as it affects the conclusion.

Reference

1. Rudick R a., Stuart WH, Calabresi P a., et al. Natalizumab plus Interferon Beta-1a for Relapsing Multiple Sclerosis. *N Engl J Med*. 2006;354(9):911-923. doi:10.1056/NEJMoa044396
2. Polman CH, O'Connor PW, Havrdova E, et al. A Randomized, Placebo-Controlled Trial of Natalizumab for Relapsing Multiple Sclerosis. *J new Engl Med*. 2006;354(9):899-910. doi:10.1056/NEJMoa1508749
3. Kappos L, Radue EW, O'Connor P, et al. A Placebo-Controlled Trial of Oral Fingolimod in Relapsing Multiple Sclerosis Ludwig. *New Engl J*. 2010;362(5):387-401. doi:10.1056/NEJMoa1613683
4. Cohen J a., Barkhof F, Comi G, et al. Oral Fingolimod or Intramuscular Interferon for Relapsing Multiple Sclerosis. *N Engl J Med*. 2010;362(5):402-415. doi:10.1056/NEJMoa0907839
5. Vargas DL, Tyor WR. Update on disease-modifying therapies for multiple sclerosis. *J Investig Med*. 2017;65(5):883-891. doi:10.1136/jim-2016-000339
6. Koch-Henriksen N, Magyari M, Sellebjerg F, Soelberg Sørensen P. A comparison of multiple sclerosis clinical disease activity between patients treated with natalizumab and fingolimod. *Mult Scler J*. 2017;23(2):234-241. doi:10.1177/1352458516643393
7. Lorscheider J, Benkert P, Lienert C, et al. Comparative analysis of natalizumab versus fingolimod as second-line treatment in relapsing–remitting multiple sclerosis. *Mult Scler J*. 2018;24(6):777-785. doi:10.1177/1352458518768433
8. Prosperini L, Saccà F, Cordioli C, et al. Real-world effectiveness of natalizumab and fingolimod compared with self-injectable drugs in non-responders and in treatment-naïve patients with multiple sclerosis. *J Neurol*. 2017;264(2):284-294. doi:10.1007/s00415-016-8343-5
9. Barbin L, Rousseau C, Casey R, Edan G, Lebrun-frenay C. Comparative efficacy of fingolimod

vs natalizumab A French multicenter observational study. *Neurology*. 2016;86:771-778. doi:10.1212/WNL.0000000000002395

10. Kalincik T, Horakova D, Spelman T, et al. Switch to natalizumab versus fingolimod in active relapsing-remitting multiple sclerosis. *Ann Neurol*. 2015;77(3):425-435. doi:10.1002/ana.24339
11. Baroncini D, Ghezzi A, Annovazzi PO, et al. Natalizumab versus fingolimod in patients with relapsing-remitting multiple sclerosis non-responding to first-line injectable therapies. *Mult Scler*. 2016;22(10):1315-1326. doi:10.1177/1352458516650736
12. Gajofatto A, Bianchi MR, Deotto L, Benedetti MD. Are natalizumab and fingolimod analogous second-line options for the treatment of relapsing-remitting multiple sclerosis? A clinical practice observational study. *Eur Neurol*. 2014;72:173-180. doi:10.1159/000361044
13. Sharmin S, Lefort M, Andersen J, et al. Comparative effectiveness of natalizumab and fingolimod in different subgroups of patients with relapsing-remitting multiple sclerosis. In: ECTRIMS, ed. Stockolm: ECTRIMS; 2019.
14. Barbin L, Rousseau C, Casey R, Edan G, Lebrun-frenay C. Comparative efficacy of fingolimod vs natalizumab: A French multicenter observational study. *Neurology*. 2016;87(18):1961. doi:10.1212/WNL.0000000000003325
15. Csizmadi I, Collet J-P. Bias and Confounding in Pharmacoepidemiology. *Textb Pharmacoepidemiol*. 2006.
16. Austin PC. A tutorial and case study in propensity score analysis: An application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate Behav Res*. 2011;46(1):119-151. doi:10.1080/00273171.2011.540480
17. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399-424. doi:10.1080/00273171.2011.568786
18. Kalincik T, Jokubaitis V, Izquierdo G, et al. Comparative effectiveness of glatiramer acetate and interferon beta formulations in relapsing–remitting multiple sclerosis. *Mult*

Scler J. 2015;21(9):1159-1171. doi:10.1177/1352458514559865

19. Butzkueven H, Chapman J, Cristiano E, et al. MSBase: An international, online registry and platform for collaborative outcomes research in multiple sclerosis. *Mult Scler*. 2006;12(6):769-774. doi:10.1177/1352458506070775
20. Magyari M, Koch-Henriksen N, Sørensen PS. The Danish multiple sclerosis treatment register. *Clin Epidemiol*. 2016;8:549-552. doi:10.2147/CLEP.S99500
21. Vukusic S, Casey R, Rollot F, et al. Observatoire Français de la Sclérose en Plaques (OFSEP): A unique multimodal nationwide MS registry in France. *Mult Scler J*. 2018:1-5. doi:10.1177/1352458518815602
22. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33(11):1444-1452. doi:10.1212/WNL.33.11.1444
23. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects Effects. *Biometrika*. 1983;70(1):41-55.
24. Austin PC, Stuart E a. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679. doi:10.1002/sim.6607
25. Hernán M, Robins J. *Causal Inference*. Boca Raton.; 2019.
26. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat*. 2011;10(2):150-161. doi:10.1002/pst.433
27. Vaughan AS, Kelley CF, Luisi N, Del Rio C, Sullivan PS, Rosenberg ES. An application of propensity score weighting to quantify the causal effect of rectal sexually transmitted infections on incident HIV among men who have sex with men Data analysis, statistics and modelling. *BMC Med Res Methodol*. 2015;15(1):1-9. doi:10.1186/s12874-015-0017-y
28. Kalincik T, Cutter G, Spelman T, Jokubaitis V, Havrdova E, Horakova D. Defining reliable disability outcomes in multiple sclerosis. *Brain*. 2015;138(11):3287-3298.
29. Cole SR, Hernán MA. Adjusted survival curves with inverse

probability weights. 2004:45-49. doi:10.1016/j.cmpb.2003.10.004

30. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1):48-55. doi:10.1177/1740774511420743

31. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology*. 2017;28(1):54-59. doi:110.1016/j.bbi.2017.04.008

32. Sormani MP, Bruzzi P. Can we measure long-term treatment effects in multiple sclerosis? *Nat Rev Neurol*. 2014;11(3):176-182. doi:10.1038/nrneurol.2014.237

33. Murray EJ, Hernán MA. Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials. *Clin Trials*. 2016;13(4):372-378. doi:10.1177/1740774516634335

34. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560. doi:10.1097/00001648-200009000-00011

35. Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Med Decis Mak*. 2009;29(6):661-677. doi:10.1177/0272989X09341755

36. Austin PC. The performance of different propensity score methods for estimating marginal odds ratios. 2007;26:3078-3094. doi:10.1002/sim

37. Austin PC. The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. *Stat Med*. 2010;29(20):2137-2148. doi:10.1002/sim.3854

38. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med*. 2013;32(16):2837-2849. doi:10.1002/sim.5705

39. Pirracchio R, Carone M, Rigon MR, Caruana E, Mebazaa A, Chevret S. Propensity score estimators for the average treatment effect and the average treatment effect on the treated may yield very different estimates. *Stat Methods Med Res*. 2013;25(5):1938-1954. doi:10.1177/0962280213507034

40. Brookhart M alan, Schneeweiss S,

Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149-1156. doi:10.1093/aje/kwj149

41. Witte J, Didelez V. Covariate selection strategies for causal inference: Classification and comparison. *Biometrical J*. 2018;(December 2017):1-20. doi:10.1002/bimj.201700294

42. Nayan M, Hamilton RJ, Juurlink DN, Finelli A, Kulkarni GS, Austin PC. Critical appraisal of the application of propensity score methods in the urology literature. *BJU Int*. 2017;120(6):873-880. doi:10.1111/bju.13930

43. Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: A systematic review and suggestions for improvement. *J Thorac Cardiovasc Surg*. 2007;134(5). doi:10.1016/j.jtcvs.2007.07.021

44. Rosenbaum P. *Observational Studies*. Pringer-Ve. New-York; 2002.

45. Tintore M, Rovira À, Río J, et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain*. 2015;138(7):1863-1874. doi:10.1093/brain/awv105

46. Bovis F, Signori A, Carmisciano L, et al. Expanded disability status scale progression assessment heterogeneity in multiple sclerosis according to geographical areas. *Ann Neurol*. 2018;84(4):621-625. doi:10.1002/ana.25323

Tables and Figures

Table 1: Overview of all evaluated analytical approaches

Outcome	PS method	Analytical model
Counts of relapses[a]	Weighting[b]	Weighted negative binomial model adjusted for treatment status
	Matching[c]	Negative binomial model with cluster for treatment status
Time to first relapse[a]	Weighting[b]	Weighted Cox model adjusted for treatment status
	Matching[c]	Frailty Cox model adjusted for treatment status
Time to first EDSS worsening[a]	Weighting[b]	Weighted Cox model adjusted for treatment status and visit density
	Matching[c]	Frailty Cox model adjusted for treatment status and visit density
Time to first EDSS improvement[a]	Weighting[b]	Weighted Cox model adjusted for treatment status and visit density
	Matching[c]	Frailty Cox model adjusted for treatment status and visit density

[a]Analyses were conducted in intention-to-treat, per-protocol and pairwise-censoring (matching only) frameworks

[b]Three values of calipers were considered (0.02,0.1,0.2)

[c]Two type of weights were considered (inverse probability weighting and weighting by the odds)

Table 2: Follow-up duration according outcomes of interest in years

Outcome	Intention-to-treat analysis	Per-protocol analysis
Counts of relapses[a]	3.17 (2.01-4.59)	2.09 (1.24-3.41)
Time to first relapse[b]	3.11 [3.05; 3.18]	2.27 [2.21; 2.31]
Time to first EDSS worsening[b]	3.16 [3.10; 3.23]	2.11 [2.07; 2.16]
Time to first EDSS improvement[b]	3,20 [3,13; 3.27]	2.08 [2.02; 2.12]

[a] Median length of follow-up (median (Interquartile range))

[b] Median survival time of the reverse Kaplan-Meier, taking into account the length and the completeness of follow-up (median (95% confidence interval)

Table 3: Characteristics at baseline of the overall study population, as well as the subgroups of matched and unmatched patients depending on the caliper value

	Overall	Matching with caliper=0.1		Matching with caliper=0.2		Matching with caliper=0.02	
	ALL N=5148	**Matched N=3258**	**Excluded N=1890**	**Matched N=3278**	**Excluded N=1870**	**Matched N=3232**	**Excluded N=1916**
Sex[a]							
Female	3698 (72%)	2342 (72%)	1356 (72%)	2352 (72%)	1346 (72%)	2332 (72%)	1366(71%)
Male	1450 (28%)	916 (28%)	534 (28%)	926 (28%)	524 (28%)	900 (28%)	550 (29%)
Age at treatment start[b]	37.7 (30.1-44.7)	37.3 (30.1-44.3)	38.5 (31.8-45.6)	37.2 (30.1-44.4)	38.7 (31.8-45.4)	37.4 (30.2-44.5)	38.3 (31.5-45.1)
MS duration at treatment start[b]	6.9 (3.1-12.5)	6.4 (2.7-2.0)	7.9 (4.0-13.4)	6.4 (2.6-2.0)	7.9 (4.0-13.4)	6.3 (2.6-12.0)	8.1 (4.0-13.3)
EDSS at treatment start[a]							
Less than 2	1556 (30%)	782 (24%)	774 (41%)	810 (25%)	746 (40%)	789 (24%)	767 (40%)
Between 2 and 3.5	2384 (46%)	1609 (49%)	775 (41%)	1593 (49%)	791 (42%)	1588 (49%)	796 (42%)
4 or more	1208 (23%)	867 (27%)	341 (18%)	875 (27%)	33 (18%)	855 (26%)	353 (18%)
Number of relapses in the previous 12 months[a]							
0	1857 (36%)	1063 (33%)	794 (42%)	1085 (33%)	772 (41%)	1059 (33%)	798 (42%)
1	2021 (39%)	1290 (40%)	731 (39%)	1268 (39%)	753 (40%)	1276 (39%)	745 (39%)
2	975 (19%)	690 (21%)	285 (15%)	707 (22%)	268 (14%)	696 (22%)	279 (15%)
3 or more	295 (6%)	215 (7%)	80 (4%)	218 (7%)	77 (4%)	107 (7%)	94(5%)
Number of previous MS treatments[a]							
0	836 (16%)	582 (18%)	254 (13%)	580 (18%)	256 (14%)	584 (18%)	252 (13%)
1	2559 (50%)	1594 (49%)	965 (51%)	1597 (49%)	962 (51%)	1558 (48%)	1001 (52%)
2	1187 (23%)	738 (23%)	449 (24%)	744 (23%)	443 (24%)	748 (23%)	439 (23%)
3 or more	566 (11%)	344 (11%)	222 (12%)	357 (11%)	209 (11%)	342 (11%)	224 (12%)
MS activity before treatment start[a]							
None	1438 (28%)	776 (24%)	662 (35%)	782 (24%)	656 (35%)	764 (24%)	674 (35%)
Progression	419 (8%)	287 (9%)	132 (7%)	303 (9%)	116 (6%)	295 (9%)	124 (6%)
Relapse	2159 (42%)	1395 (43%)	764 (40%)	1398 (43%)	761 (41%)	1397 (43%)	762 (40%)
Relapse and progression	1132 (22%)	800 (25%)	332 (18%)	795 (24%)	337 (18%)	776 (24%)	356(19%)
Data source							
MS Base	3293 (64%)	1874 (58%)	1419 (75%)	1882 (57%)	1411(75%)	1852 (57%)	1441 (75%)
DMSR	1444 (28%)	1167 (36%)	277 (15%)	1179 (36%)	265 (14%)	1153 (36%)	291 (15%)
OFSEP	411 (8%)	217 (7%)	194 (10%)	217(7%)	194 (10%)	227 (7%)	184 (10%)

[a]N(%); [b]Median (IQR)

Table 4: Characteristics at baseline according to treatment group of the overall population and the three matched cohorts

	Overall N=5148			Matching (caliper=0.1) N=3258			Matching (caliper =0.2) N=3278			Matching (caliper=0.02) N=3232		
	natalizumab N=1989	**fingolimod** N=3159	**SMD[c]**	**natalizumab** N=1629	**fingolimod** N=1629	**SMD[c]**	**natalizumab** N=1639	**fingolimod** N=1639	**SMD[c]**	**natalizumab** N=1616	**fingolimod** N=1616	**SMD[c]**
Sex[a]			4%			1%			2%			0.8%
Female	1451 (73%)	2247 (71%)		1175 (72%)	1167 (72%)		1183 (72%)	1169 (71%)		1169 (72%)	1163 (28%)	
Male	538 (27%)	912 (29%)		454 (28%)	462 (28%)		456 (28%)	470 (29%)		447 (28%)	453 (28%)	
Age at treatment start[b]	36.6 (29.3; 43.9)	38.5 (31.6; 45.4)	13%	37.2 (29.8-44.4)	37.4 (30.3-44.2)	0.5%	37.2 (29.7-44.4)	37.4 (30.3-44.3)	1%	37.3 (29.7-44.4)	37.6 (30.5-44.5)	3%
MS duration at treatment[b]	6.3 (2.4; 11.8)	7.4 (3.6; 13.0)	13%	6.3 (2.3-12.2)	6.5 (2.9-11.8)	2%	6.2 (2.3-12.2)	6.6 (2.9-11.9)	2%	6.3 (2.3-12.2)	6.4 (2.8-11.9)	2%
EDSS at treatment start[a]			32%			5%			9%			7%
2 or less	434 (22%)	1122 (35%)		374 (23%)	408 (25%)		377 (23%)	433 (26%)		372 (23%)	417(26%)	
Between 2 and 4	981 (49%)	1403 (44%)		822 (50%)	787 (48%)		826 (50%)	767 (47%)		813 (50%)	775 (48%)	
4 or more	574 (29%)	634 (20%)		433 (27%)	434 (27%)		436 (25%)	439 (27%)		431 (27%)	424 (26%)	
Number of relapses in the previous 12 months[a]			37%			8%			6%			6%
0	570 (29%)	1287 (41%)		543 (33%)	520 (32%)		545 (33%)	540 (33%)		541 (33%)	518 (32%)	
1	752 (38%)	1269 (40%)		620 (38%)	670 (41%)		623 (38%)	645 (39%)		618 (38%)	658 (41%)	
2	484 (24%)	491 (15%)		346 (21%)	344 (21%)		351 (22%)	356 (22%)		350 (22%)	346 (21%)	
3 or more	183 (9%)	112 (3%)		120 (7%)	95 (6%)		120 (7%)	98 (6%)		107 (7%)	94 (6%)	
Number of previous MS treatments[a]			17%			17%			17%			16%
0	401 (20%)	435 (14%)		334 (21%)	248 (15%)		337 (21%)	243 (15%)		334 (21%)	250 (15%)	
1	924 (46%)	1635 (52%)		742 (46%)	852 (52%)		746 (46%)	851 (52%)		732 (4%)	826 (51%)	
2	457 (23%)	730 (23%)		367 (23%)	371 (23%)		370 (23%)	374 (23%)		365 (23%)	383 (24%)	
3 or more	207 (10%)	359 (11%)		186 (11%)	158 (10%)		186 (11%)	171 (10%)		185 (11%)	157 (10%)	
MS activity before treatment start[a]			29%			4%			2%			3%
None	410 (21%)	1028 (32%)		393 (24%)	383 (24%)		395 (24%)	387 (24%)		390 (24%)	374 (23%)	
Progression	160 (8%)	259 (8%)		150 (9%)	137 (8%)		150 (9%)	153 (9%)		151(9%)	144 (9%)	
Relapse	886 (44%)	1273 (40%)		686 (42%)	709 (44%)		694 (42%)	704 (43%)		690 (43%)	707 (44%)	
Relapse and progression	533 (27%)	599 (19%)		400 (25%)	400 (25%)		400 (24%)	395 (24%)		385 (24%)	391 (24%)	
Data source[a]			28%			18%			16%			12%
MS Base	1141 (57%)	2152 (68%)		949 (58%)	925 (57%)		957 (58%)	925 (56%)		935 (58%)	917 (57%)	
DMSR	607 (31%)	837 (26%)		607 (37%)	560 (34%)		607 (37%)	572 (35%)		593 (37%)	560 (35%)	
OFSEP	241 (12%)	170 (5%)		73 (4%)	144 (9%)		75 (5%)	142 (9%)		88 (5%)	139 (9%)	

[a]N(%); [b]Median (IQR); [c]SMD: standardized mean differences or Mahalanobis distances between Natalizumab treated patients and Fingolimod treated patients

Table 5: Characteristics at baseline by treatment group in the overall study sample, and cohorts weighted on sIPTW and odds

	Unweighted N= 5148			Weighting using sIPTW			Weighting using the odds		
	natalizumab N=1989	**fingolimod** N=3159	**SMD[c]**	**natalizumab**	**fingolimod**	**SMD[c]**	**natalizumab**	**fingolimod**	**SMD[c]**
Sex[a]			4%			1%			1%
Female	1451 (73%)	2247 (71%)		71%	72%		73%	73%	
Male	538 (27%)	912 (29%)		29%	28%		27%	27%	
Age at treatment start[b]	36.6 (29.3; 43.9)	38.5 (31.6; 45.4)	13%	37.6 (30.4-45.2)	37.9 (30.7-44.8)	2%	36.6 (29.3-43.9)	36.8 (29.8-43.9)	1%
MS duration at treatment[b]	6.3 (2.4; 11.8)	7.4 (3.6; 13.0)	13%	6.8 (2.7-12.9)	7.0 (3.2-12.4)	2%	6.2 (2.4-11.8)	6.2 (2.7-11.6)	1%
EDSS at treatment start[a]			32%			12%			7%
2 or less	434 (22%)	1122 (35%)		26%	31%		22%	23%	
Between 2 and 4	981 (49%)	1403 (44%)		50%	45%		49%	46%	
4 or more	574 (29%)	634 (20%)		24%	24%		29%	31%	
Number of relapses in the previous 12 months[a]			37%			4%			3%
0	570 (29%)	1287 (41%)		35%	36%		29%	29%	
1	752 (38%)	1269 (40%)		39%	39%		38%	37%	
2	484 (24%)	491 (15%)		20%	18%		24%	23%	
3 or more	183 (9%)	112 (3%)		6%	6%		9%	10%	
Number of previous MS treatments[a]			17%			15%			14%
0	401 (20%)	435 (14%)		19%	14%		20%	15%	
1	924 (46%)	1635 (52%)		46%	52%		46%	52%	
2	457 (23%)	730 (23%)		23%	23%		23%	23%	
3 or more	207 (10%)	359 (11%)		12%	11%		10%	10%	
MS activity before treatment start[a]			29%			3%			3%
None	410 (21%)	1028 (32%)		26%	28%		21%	20%	
Progression	160 (8%)	259 (8%)		8%	8%		8%	9%	
Relapse	886 (44%)	1273 (40%)		42%	42%		44%	43%	
Relapse and progression	533 (27%)	599 (19%)		23%	22%		27%	27%	
Data source[a]			28%			1%			8%
MS Base	1141 (57%)	2152 (68%)		30%	29%		57%	53%	
DMSR	607 (30%)	837 (26%)		62%	62%		30%	34%	
OFSEP	241 (12%)	170 (5%)		8%	8%		12%	13%	

[a]N(%); [b]Median (IQR); [c]SMD: standardized mean differences and Mahalanobis distance between natalizumab treated patients and fingolimod treated patients

Figure 1: Flowchart of the study population

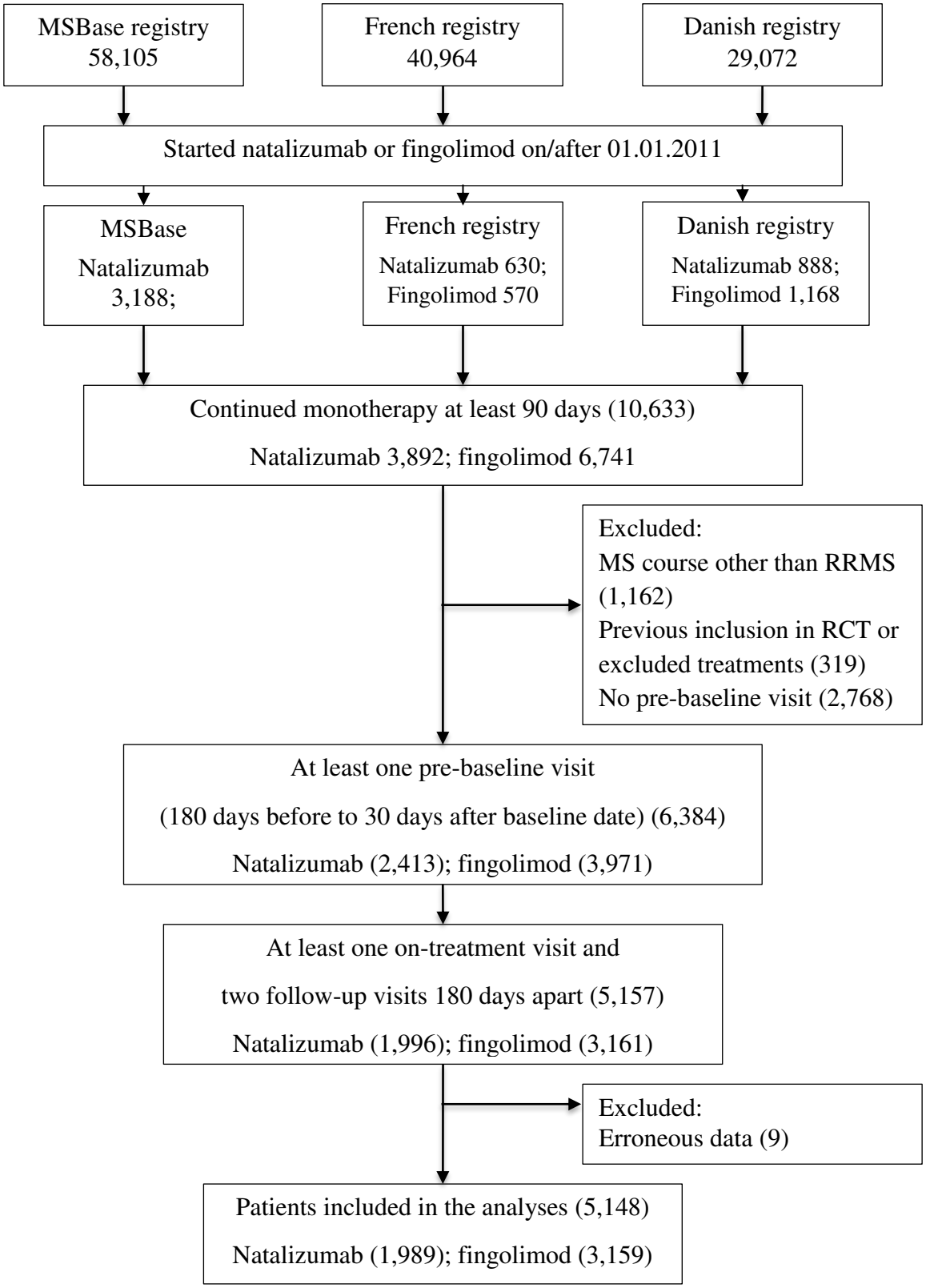


Figure 2: Propensity scores by treatment group (probability of being treated by natalizumab)

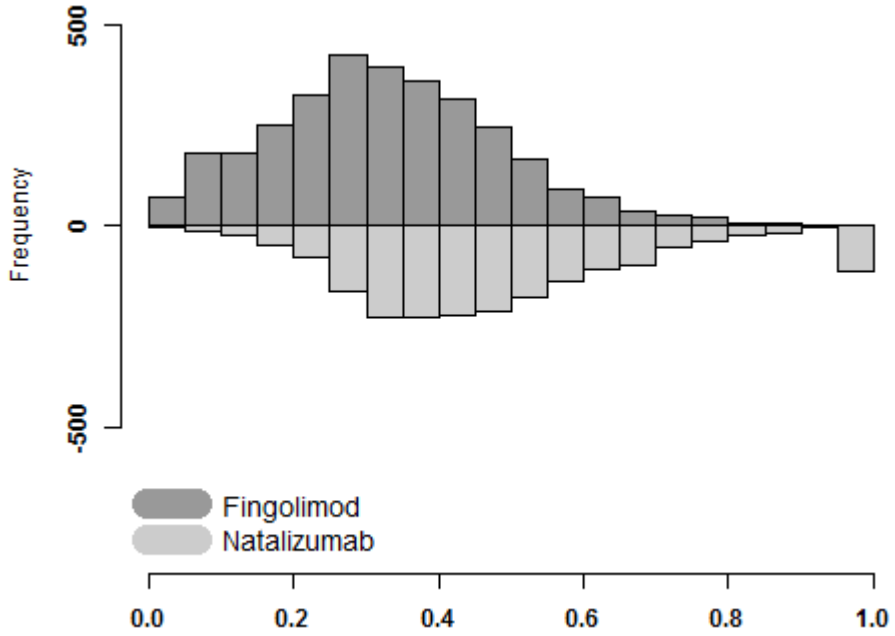
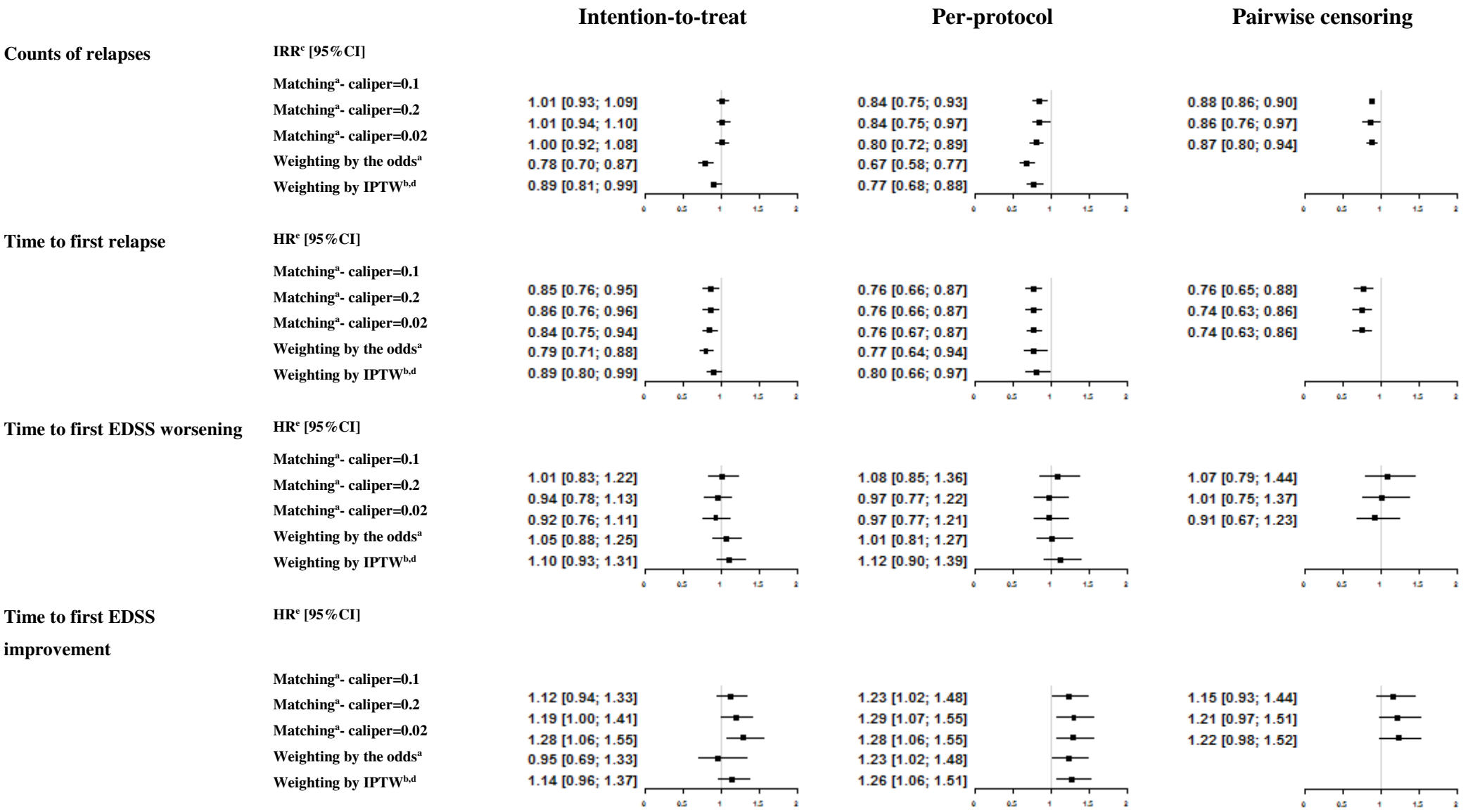


Figure 3: Forest plot of treatment effect for the four outcomes, according to the 52 analyses (3 matching + 2 weighting * Intention-to-treat/ per-protocol/pairwise censoring)



[a]Average Treatment effect for treated; [b]Average Treatment effect for the entire population; [c]Incidence Rate Ratio, [d]inverse Probability of treatment weights, [e]hazard ratio & 95%Confidence Interval. For counts of relapses, time to first relapse, time to first EDSS worsening: a value less than one means an effect in favor of natalizumab in comparison to fingolimod. For time to first EDSS improvement: a value more than one means an effect in favor of natalizumab in comparison to fingolimod

Supplementary materials

Table S1: Treatment exposure after natalizumab or fingolimod start during the recorded follow-up.

Fingolimod-treated patients (N= 3159)		Natalizumab-treated patients (N= 1989)	
Sequence[a]	**N(%)**	**Sequence[a]**	**N(%)**
Unknown due to ongoing fingolimod or loss of follow-up	2577 (81%)	Unknown due to ongoing natalizumab or loss of follow-up	1311 (66%)
Natalizumab	222 (7%)	Fingolimod	393 (20%)
Fingolimod	93 (3%)	Natalizumab	101 (5%)
Dimethyl Fumarate	42 (1%)	Dimethyl Fumarate	23 (1%)
Teriflunomide	29 (0.9%)	Glatiramer acetate	14 (0.7%)
Interferon	15(0.5%)	Fingolimod, Natalizumab	12 (0.6%)
Glatiramer acetate	15 (0.5%)	Rituximab	11 (0.5%)

[a]sequence observed for 0.5% or more patients

Table S2: characteristics at baseline of unmatched cohort by treatment group

	Overall N=5148		Unmatched – caliper=0.1 N=1890		Unmatched – caliper=0.2 N=1870		Unmatched – caliper=0.02 N=1916	
	natalizumab N=1989	**fingolimod** N=3159	**natalizumab** N=360	**fingolimod** N=1530	**natalizumab** N=350	**fingolimod** N=1520	**natalizumab** N=373	**fingolimod** N=1543
Sex[a]								
Female	1451 (73%)	2247 (71%)	276 (77%)	1356 (71%)	268 (77%)	1078 (71%)	282 (76%)	1084 (70%)
Male	538 (27%)	912 (29%)	84 (23%)	534 (29%)	82 (23%)	442 (29%)	91 (24%)	459 (30%)
Age at treatment start[b]	36.6 (29.3-43.9)	38.5 (31.6-45.4)	37.2 (29.8- 44.4)	37.4(30.3-44.2)	37.2 (29.7-44.4)	37.3 (30.3-44.3)	37.3 (29.8-44.4)	37.6 (30.5-44.5)
MS duration at treatment[b]	6.3 (2.4-11.8)	7.4 (3.6-13.0)	6.3 (2.3-12.2)	6.5(2.9-11.8)	6.2 (2.3-12.2)	6.6 (2.9-11.9	6.3 (2.33-12.2)	6.4 (3.8-11.9)
EDSS at treatment start[a]								
2 or less	434 (22%)	1122 (35%)	60 (17%)	714 (47%)	57 (16%)	689 (45%)	62 (17%)	705 (46%)
Between 2 and 4	981 (49%)	1403 (44%)	159 (44%)	616 (40%)	155 (44%)	636 (42%)	168 (45%)	628 (41%)
4 or more	574 (29%)	634 (20%)	141 (39%)	200 (13%)	138 (39%)	195 (13%)	143 (38%)	210 (14%)
Number of relapses in the last 12 months[a]								
0	570 (29%)	1287 (41%)	27 (8%)	767 (50%)	25 (7%)	747 (49%)	29 (8%)	769 (50%)
1	752 (38%)	1269 (40%)	132 (37%)	599 (39%)	129 (37%)	624 (41%)	134 (36%)	611 (40%)
2	484 (24%)	491 (15%)	138 (38%)	147 (10%)	133 (38%)	135 (9%)	134 (36%)	145 (9%)
3 or more	183 (9%)	112 (3%)	63 (18%)	17 (1%)	63 (18%)	14 (1%)	76 (20%)	18 (1%)
Number of previous treatment[a]								
0	401 (20%)	435 (14%)	67 (19%)	187 (12%)	64 (18%)	192 (13%)	67 (18%)	185 (12%)
1	924 (46%)	1635 (52%)	182 (51%)	783 (51%)	178 (51%)	784 (52%)	192 (51%)	809 (52%)
2	457 (23%)	730 (23%)	90 (25%)	359 (23%)	87 (25%)	356 (23%)	92 (25%)	347 (22%)
3 or more	207 (10%)	359 (11%)	21 (6%)	201 (13%)	21 (6%)	188 (12%)	22 (6%)	202 (13%)
MS activity before treatment start[a]								
None	410 (21%)	1028 (32%)	17 (5%)	645 (42%)	15 (4%)	641 (42%)	20 (5%)	654 (42%)
Progression	160 (8%)	259 (8%)	10 (3%)	122 (8%)	10 (3%)	106 (7%)	9 (2%)	115 (7%)
Relapse	886 (44%)	1273 (40%)	200 (56%)	564 (37%)	192 (55%)	569 (37%)	196 (53%)	566 (37%)
Relapse and progression	533 (27%)	599 (19%)	133 (37%)	199 (13%)	133 (38%)	204 (13%)	148 (40%)	208 (13%)
Register[a]								
MS Base	1141 (57%)	2152 (68%)	192 (53%)	1227 (80%)	184 (53%)	1227 (81%)	206 (55%)	1235 (80%)
DMSR	607 (31%)	837 (26%)	0 (0%)	277 (18%)	0 (0%)	265 (17%)	14 (4%)	277 (18%)
OFSEP	241 (12%)	170 (5%)	168 (47%)	26 (2%)	166 (47%)	28 (2%)	153 (41%)	31 (2%)

[a]N(%); [b]Median (IQR); [c]SMD: standardized mean differences between Natalizumab treated patients and Fingolimod treated patients

Chapitre 5

V. Evaluation des stratégies thérapeutiques

A. Contexte **166**
1. Contexte scientifique 166
1.1 Le choix de la stratégie thérapeutique 166
1.2 Le moment de l'initiation du traitement 167
2. Contexte logistique 169
B. Description des stratégies thérapeutiques **170**
C. Efficacité de l'initiation précoce d'un traitement : résultats préliminaires **187**
1. Contexte 187
2. Patients et méthodes 188
2.1 Population d'étude 188
2.2 Exposition 189
2.3 Critères de jugement 189
2.4 Analyse statistique 189
3. Résultats 190
4. Discussion et perspectives 194
D. Comparaison de l'efficacité des stratégies thérapeutiques : perspectives **197**

A. Contexte

Dans cette partie du travail, nous nous sommes concentrés sur les pratiques d'utilisation des traitements de fond de la SEP en France. Deux questions peuvent notamment émerger dans la décision thérapeutique. Quel traitement utilisé au regard de la discussion avec le patient et de l'activité de la maladie ? A quel moment l'instaurer ?

Ainsi, dans cette partie, toujours dans un objectif d'évaluation d'efficacité, nous nous sommes concentrés sur les utilisations de traitements au regard de ces deux questionnements.

1. Contexte scientifique

1.1 Le choix de la stratégie thérapeutique

Comme évoqué dans la partie introductive de cette thèse, les options thérapeutiques dans la SEP rémittente sont devenues plus nombreuses ces dernières années. Les recommandations établies incitent à considérer l'ensemble des traitements disponibles dans le traitement de la SEP au regard du rapport bénéfice/risque et des caractéristiques des patients, incluant les caractéristiques individuelles, les comorbidités et la sévérité de la maladie (Montalban *et al.*, 2018). Dans un tel contexte, trois stratégies thérapeutiques peuvent être identifiées et définies comme suit :

(1) La stratégie d'escalade dans laquelle on débute par un traitement de première ligne (interférons-β, acétate de glatiramère, etc.) avant de changer verticalement, si le patient ne répond plus au premier traitement

(2) La stratégie intensive caractérisée par l'indication d'un traitement de deuxième ligne (natalizumab, fingolimod, etc.) d'emblée

(3) La stratégie d'induction consistant en l'administration d'emblée d'un traitement de troisième ligne sur un court terme, suivi d'un relais par un traitement de première ligne

Bien que dans cette thèse nous nous sommes principalement intéressés à des questions portant sur l'efficacité des traitements, il convient d'étudier les traitements dans leur globalité. En particulier, leur profil de tolérance est un élément important dont le neurologue tient compte

dans le cadre de sa décision thérapeutique. Parmi les trois stratégies évoquées, les deux dernières citées (intensive et induction) exposent les patients à des risques d'effets indésirables rares mais graves. Ces risques peuvent rendre la décision thérapeutique difficile à prendre pour le neurologue comme pour le patient, devant tous deux évaluer cette décision au regard du rapport bénéfice/risque. Des études sur l'efficacité des traitements de fond dits de haute efficacité ont déjà été conduites et ont prouvé leur efficacité, voire leur supériorité par rapport aux traitements de première ligne (Putzki *et al.*, 2010*a*, *b*; Khatri *et al.*, 2011; Cohen *et al.*, 2013; Agius *et al.*, 2014; Izquierdo *et al.*, 2017; Buttmann, 2018; Chartier *et al.*, 2018). En particulier, un travail franco-italienne a étudié un groupe de patients présentant des formes actives de SEP pendant trois ans dans un essai randomisé comparant le traitement par mitoxantrone pendant 6 mois suivi par un traitement d'interférons-β à un traitement d'interféron-β d'emblée (Edan *et al.*, 2011). Cet essai a démontré le rôle bénéfique que présente le traitement d'induction dans la prise en charge des patients atteints de SEP agressives. En revanche, à notre connaissance, aucune étude observationnelle n'a cherché à comparer l'efficacité des stratégies chez des patients éligibles à la stratégie d'induction, réservée à des formes très sévères de SEP. Certains évoquent même le manque d'essai randomisé sur les effets de l'induction (D'Amico *et al.*, 2017).

1.2 Le moment de l'initiation du traitement

Si différentes stratégies thérapeutiques peuvent être envisagées, le moment de l'initiation du traitement semble faire consensus, toute sévérité confondue. En effet, aujourd'hui, nombreux sont les arguments qui incitent à agir au plus tôt dans le développement de la maladie. Surtout que les nouveaux critères diagnostiques permettent d'identifier la SEP de plus en plus tôt, en tenant compte des paramètres fournis par l'IRM.

Un des premiers arguments en faveur de l'initiation précoce d'un traitement concerne le mécanisme d'action des traitements. En effet, les études conduites à ce jour suggèrent que les traitements ont un effet plus efficace durant le processus inflammatoire de la maladie que durant la phase progressive, marquée par le processus neurodégénératif (Wolinsky *et al.*, 2007; Kieseier, 2011; Baldassari and Fox, 2018). Ceci suggère, de fait, qu'il y a une fenêtre d'opportunité de traitement : celle de la phase inflammatoire, première phase de développement de la maladie (Freedman *et al.*, 2014). Cet argument corrobore une étude ayant démontré

l'évolution de la maladie en deux phases. La première phase, s'étendant du début clinique à la survenue d'un score EDSS irréversible de 3, apparait être de durée variable et correspond à une inflammation focale (Leray et al. 2010). La seconde phase conduit à un score EDSS irréversible de 6 dans une durée constante, indépendamment du temps d'atteinte du premier palier. Ces arguments conduisent alors à penser que l'objectif est de contrôler au mieux la première phase de maladie afin de retarder au maximum l'atteinte de l'EDSS 3. En effet, puisque les traitements approuvés ciblent l'inflammation, on cherchera à retarder le handicap le plus possible en les utiliser tôt, on cherchera à retarder le handicap le plus possible.

Un autre argument est fourni par les extensions d'essais thérapeutiques. En particulier, cinq études ont été conduites sur l'efficacité des interférons-β et de l'acétate de glatiramère chez des patients présentant un syndrome cliniquement isolé (communément appelé CIS pour « clinical isolated syndrom »). Le CIS est un premier épisode de démyélinisation d'origine inflammatoire touchant le système nerveux, autrement dit une poussée (Miller *et al.*, 2012). Cet épisode ponctuel représente un risque important de conversion en SEP. Dans les essais de traitement de première ligne, CHAMPS (interféron-β), ETOMS (interféron-β), BENEFIT (interféron-β), PreCISe (acétate de glatiramère) et REFLEX (interféron-β), les chercheurs se sont concentrés sur le risque de développement de la SEP après ce premier épisode (Jacobs *et al.*, 2000; Comi *et al.*, 2001*a*, 2009, 2012; Kappos *et al.*, 2006*a*; Freedman *et al.*, 2014). Ces cinq études ont démontré que chez les patients présentant un CIS, le risque de conversion en SEP était réduit chez les patients traités par rapport aux non traités de 52% à 39%, selon les études. Cet argument est d'ailleurs renforcé par les extensions de ces études originales (Fujihara, 2014; Kappos *et al.*, 2016). Des résultats tendant à montrer un bénéfice de l'initiation précoce des traitements dits de haute-efficacité sont également disponibles (Merkel *et al.*, 2017) .

D'autres arguments immunopathologiques et histopathologiques conduisent également à préférer l'initiation précoce du traitement (Tintoré, 2008, 2009). Il a en effet été montré que les dommages axonaux irréversibles qui sous-tendent l'accumulation du handicap se produisent très tôt dans l'évolution de la maladie. Des études d'histoire naturelle démontrent, de plus, que le nombre de poussées survenues au cours des premières années de la maladie est associé au temps écoulé avant la survenue du handicap (Confavreux *et al.*, 2003; Leray *et al.*, 2010). Dans le même sens, l'accumulation des lésions dans les premières années de SEP est un facteur pronostique de l'évolution du handicap à long terme.

L'ensemble de ces arguments conduit à penser que plus le traitement est instauré tôt, meilleur sera le résultat. Ainsi, agir tôt est désormais recommandé (Filippini *et al.*, 2017; Montalban *et al.*, 2018) conduisant même à certaines directives. Par exemple en Angleterre (Scolding et al. 2015), il est conseillé d'instaurer le traitement dans les 12 mois suivant l'apparition de la SEP.

Dans un tel contexte, des premières études observationnelles ont été conduites pour démontrer l'effet de l'initiation précoce du traitement sur le handicap mais les résultats obtenus à ce jour méritent d'être complétés (Freedman *et al.*, 2014; Merkel *et al.*, 2017). Nous y reviendrons plus en détails dans une des parties suivantes.

Dans ce contexte, il paraît nécessaire de compléter la littérature sur l'efficacité des usages à long terme, en particulier sur la progression du handicap. Ainsi, les objectifs de ce travail étaient triples : (1) Décrire les patients traités en France sur la période s'étendant de 1996 à 2013 et les traitements reçus, (2) Evaluer le bénéfice de l'initiation précoce d'un traitement sur la progression du handicap au global et en fonction des caractéristiques des patients, (3) Décrire les patients recevant un traitement par induction et évaluer l'efficacité du traitement d'induction en comparaison au traitement d'escalade.

2. Contexte logistique

Cette étude a fait l'objet d'une demande de financement (25 000€) auprès de la fondation EDMUS et de la fondation ARSEP dans un appel d'offre en septembre en 2017. L'évaluation de la demande a été conduite par un jury international. La rédaction du projet scientifique m'a permis d'appréhender la rédaction d'un protocole d'étude dans le cadre d'une demande de financement. De plus, outre cette demande, il s'agissait également de fournir les éléments nécessaires pour obtenir l'accès aux données OFSEP. La décision favorable nous a été transmise mi-décembre 2017. Pour des raisons administratives, les données nous ont été fournies en janvier 2019. Après un premier nettoyage, une mise à jour a été demandée en raison de critères d'inclusion finalement trop stricte (livrée en février 2019). Nous reviendrons sur cet élément par la suite.

B. Description des stratégies thérapeutiques

Le premier objectif de ce travail comportait la description des stratégies thérapeutiques dans les centres experts français. Ce travail a été réalisé dans les derniers mois de thèse et consiste en une étude descriptive des patients dont le traitement s'inscrit dans une des stratégies thérapeutiques suivantes : d'escalade, d'induction et intensive. Ce travail a fait l'objet de la rédaction d'un article que nous proposons ci-après. Cette version de l'article n'a pas encore été soumise pour relecture à l'ensemble des co-auteurs et est présentée dans son état d'avancement. En particulier, la période d'étude de cette étude descriptive est remise en question. Nous envisageons de l'étendre jusqu'à la date d'extraction. De plus, dans le cadre d'une étude descriptive, il pourra être plus pertinent d'inclure plus de centres. Dans sa version finalisée tenant compte des corrections des co-auteurs, il sera proposé dans une revue à visée clinique, éventuellement spécifique à la SEP.

Use of therapeutic strategies in multiple sclerosis expert centers in France

Authors:

Mathilde Lefort[(1)(2)], Sandra Vukusic[(3-6)], Romain Casey[(3-6)], Gilles Edan[(2)(7)], and Emmanuelle Leray[(1)(2)(8)] on behalf of OFSEP investigators

Affiliations:

[(1)] Rennes University, EHESP, REPERES – EA 7449, F-35000 Rennes, France

[(2)] Univ Rennes, CHU Rennes, Inserm, CIC 1414 [(Centre d'Investigation Clinique de Rennes)], F-35000 Rennes, France

[(3)] Hospices Civils de Lyon, Service de Neurologie, sclérose en plaques, pathologies de la myéline et neuro-inflammation, F-69677 Bron, France

[(4)] Observatoire Français de la Sclérose en Plaques, Centre de Recherche en Neurosciences de Lyon, INSERM 1028 et CNRS UMR 5292, F-69003 Lyon, France

[(5)] Université de Lyon, Université Claude Bernard Lyon 1, F-69000 Lyon, France

[(6)] Eugène Devic EDMUS Foundation against multiple sclerosis, state-approved foundation, F-69677 Bron, France

[(7)] Department of Neurology, CHU Pontchaillou, Rennes, France

[(8)] METIS Department, EHESP French school of public health, Rennes, France

Abstract

Introduction

Treatment options for relapsing-remitting multiple sclerosis (RRMS) have increased these last two decades, offering greater choice to neurologists affecting practices. Several arguments argue to act soon after the diagnosis. The choice of treatment is made according to patients' characteristics, disease activity and drug safety leading to three main therapeutic strategies: escalation, intensive treatment and induction. Our objective was to describe the therapeutic strategies, in terms of initiation delay and patients' profiles.

Patients and methods

This study was based on data collected from ten MS expert centers contributing to OFSEP (« Observatoire Français de la Sclérose en plaques »), a French network of MS expert centers. All incident RRMS cases starting for the first time a treatment from 01/01/1996 to 31/12/2013. The first exposure was used to classify patients into three groups: escalating strategy (starting first line treatment), intensive strategy (second line) and induction (third line). Initial Treatment duration and following exposure were described. The time before being newly treated was estimated using Kaplan-Meier method.

Results

Among the 6,894 patients included in this study, the median delay from MS clinical onset to treatment start was 1.6 (0.7-2.8). This time decreased over time, as well as the width of the interquartile range. Most of patients were treated in an escalating strategy (94%). The median time before the second-line treatment exposure was 17.0 years, 95% confidence interval: [15.8; 18.6]. Among patients starting and stopped natalizumab in intensive strategy, the median washout delay before starting another treatment was respectively 7.6 weeks [3.0; 26.4] and 4.6 weeks [2.6; 10.9] before and after 2014. Finally, among patients in induction strategy (3%), the median delay from MS onset to treatment start was 1 year (0.3-2.3). After the first exposure to mitoxantrone (median numbers of infusion of 6.5), patients were either untreated or treated by a first line treatment during a median time of 10.6 years.

Conclusion

This study described practices in French expert centers and highlights changes over time, especially early treatment.

Keywords: therapeutic strategy, early treatment, multiple sclerosis

Introduction

In the last twenty years, treatment options for relapsing-remitting multiple sclerosis (RRMS) have increased, with a larger number of approved disease-modifying therapies (DMTs), with various ways of administration (injections, pills, or infusions), a better knowledge of the risk-benefit balance and an increasing experience in their use by neurologists in real life settings. In this context, the choice and the timing of the initial treatment are more complex and a matter of debate.

There is accumulating evidence to act soon once the diagnosis is set (Tintoré, 2009). Natural history and immunopathological studies have shown that what happens early during the inflammation phase of the disease influences the long-term outcome, in particular the development of irreversible neurological disability (Tintoré, 2009; Leray *et al.*, 2010). Moreover, after a certain state of irreversible disability, the disease progression is constant and seems independent of the early disease activity. As approved MS treatments target inflammation, it can be worth using them early. Moreover, extensions of randomized clinical trials pointed out that early treatment was beneficial to reduce the time of conversion from a first attack to a clinically definite MS, compared with delayed treatment (Comi *et al.*, 2013; Fujihara, 2014; Kappos *et al.*, 2016). Moreover, the benefit was demonstrated persistent in the long term regarding conversion to clinically definite MS as well as time to first relapse (Hartung *et al.*, 2019). Therefore, an early action is now strongly recommended by the European Committee of Treatment and Research in Multiple Sclerosis (ECTRIMS) and the European Academy of Neurology (EAN) (Montalban *et al.*, 2018).

Beyond the timing of initiation, the choice of treatment in the context of an increasing therapeutic arsenal is a crucial question. For RRMS, this choice is made according to patient's characteristics, comorbidities, disease severity, drug safety profile and accessibility of drug (Montalban *et al.*, 2018). Three main therapeutic strategies emerged: escalation, intensive treatment and induction.

MS drugs are classified as first, second or third line, according to risk/benefit considerations. First line therapies (interferon-β, glatiramer acetate, teriflunomide or dimethyl fumarate) are considered to have a milder efficacy and a good safety profile. Second line therapies (natalizumab, fingolimod) are more active but their use is restricted due to a potentially worse risk profile. They can be used when a

patient fails a well-conducted first-line DMTs or in naïve patients with highly active MS. Third-line therapies (mitoxantrone, alemtuzumab, hematopoietic autologous stem cell transplant) are usually restricted to very aggressive disease course. Anti-CD20 antibodies (rituximab, ocrelizumab) are more difficult to classify, they are indicated in both first and second line indication.

Conceptually, there are two ways of considering MS treatments. Escalation, the most frequently used option, consists in starting with a first-line DMT and switch to more active treatments only in case of treatment failure (occurrence of a relapse or new MRI lesions). For patients with rapidly evolving severe RRMS, neurologist can suggest to start with a second line treatment as first treatment (Committee for Medicinal Products for Human Use, 2016*a*, 2018). This strategy can be called intensive strategy. Finally, for patient with very active disease, induction aims at using a very active drug first that might induce a profound and sustained change in the immune system such as mitoxantrone or alemtuzumab, followed by maintenance with a first-line DMT.

Therefore, the objective of this study is to describe the therapeutic strategies, in terms of initiation delay and patients' characteristics in a population of patients with incident MS followed in France in the 1996-2013 period.

Patients and methods

Data collection

Data were collected from OFSEP (« Observatoire Français de la Sclérose en plaques »), a French network of MS expert centers using the European database for Multiple Sclerosis (EDMUS) software in routine practice to collect prospective standardized data on all their MS patients (Vukusic *et al.*, 2018). Data from the ten oldest (data collection started in 2001 or before) MS expert centers (Lyon, Nancy, Bordeaux, Rennes, Toulouse, Nice, Clermont-Ferrand, Dijon, Besançon, Nîmes) were extracted at December 15th 2018, leading to an initial cohort of 34,582 patients.

Study population

All incident RRMS cases starting, for the first time, a treatment from January1st 1996 to December 31st 2013 were included in our study. A minimum of three visits and a follow-up of 5 years from MS onset were required. The first DMT (interferon beta-1b) was approved in France in 1996, start of the study period. It stops in

2013 to get sufficient follow-up after treatment initiation.

Exposure

All patients exposed at least one day to one of the DMTs available during the study period were considered as treated. The first exposure was used to classify patients into three groups. If a patient started a first line treatment (interferons-β, glatiramer acetate), then he/she was considered being in an intention of escalating strategy. If a patient started a second line (natalizumab and fingolimod), then he/she was considered being in an intention of intensive strategy. Finally, if a patient started mitoxantrone, then he/she was considered being in an intention of induction strategy. Three other treatments may be also observed in the follow-up since they become available after the inclusion period (teriflunomide, diméthylfumarate, alemtuzumab). An interruption of less than 3 months for interferons-β and glatiramer acetate was not considered as a treatment stop.

Variables

Following variables were described: sex, age, year of MS clinical onset, number of relapses in the 12 months preceding treatment start and EDSS score within 12 months before treatment start. The annual relapses rate was calculated as follows: number of relapses from MS onset to treatment start divided by the disease duration at treatment initiation.

Statistical analysis

Most of results were descriptive using usual indicators such as median and inter-quartile range for continuous variables and percentages for categorical variables. The three strategies were described separately. In each group, duration of the first treatment and following exposure were described. For patients treated by natalizumab, the washout delay before the switch to another DMT was studied.

Among patients in escalating strategy, the median delay before being treated with a second line treatment was estimated using Kaplan-Meier methods. The times from the start of first-line treatment were right-censored at the earliest of initiation of a third line treatment or the most recent clinical information. Among patients treated by natalizumab, the median delay before being treated by another treatment (washout delay) was estimated using Kaplan-Meier methods. In induction group, the median delay before being treated by a high-efficacy treatment (natalizumab, fingolimod, mitoxantrone, alemtuzumab) was estimated using Kaplan-Meier methods. Analysis was performed using R software (R 3.4.0).

Approval and patients' consent

All patients enrolled in the OFSEP database provide informed consent for participation and for their data to be stored and used for research. Data confidentiality and safety are ensured according to the recommendations of the French "Commission Nationale Informatique et Libertés" (CNIL). OFSEP has received approval for storing clinical, biological and imaging data for research purpose. The present study did not require any additional procedure, according to the French law. OFSEP has been registered to clinicaltrials.gov NCT02889965.

Data availability

The individual data from the present study can be obtained upon request and after validation from the OFSEP scientific committee (see website: http://www.ofsep.org/fr/)

Results

Overall population

Overall, 6,894 patients were included in the study (Figure 1). Among them, 75% were women. The median age at MS clinical onset was 30 years (Interquartile range (IQR)= 24-38). MS onset occurred before 2002 in 38%, between 2002 and 2005 in 33% and 2007-2013 in 29%. The median follow-up duration from MS onset to last clinical information was 12.4 years (IQR= 8.8-16.3).

The median delay from MS onset to treatment start was 1.6 years (IQR= 0.7-3.8). As shown in Figure 2, the median delay from MS onset to treatment start decreased over time, from 2.9 years (IQR= 1.2-5.8) before 2002 to 0.9 year (IQR= 0.4-1.8) after 2007. Furthermore, the decrease in IQR suggested the homogenization of practices among neurologists.

Overall, 6,019 (87%) patients started treatment with a first-line treatment and so belong to the escalating strategy group. Likewise, 191 (3%) patients were considered in an intensive therapeutic strategy, and 184 (3%) patients initiated mitoxantrone as first treatment, so were considered in the induction strategy. Remaining patients (N=500) were treated with an off-label treatment (7%). As shown in Figure 3, each newly available treatment was used as soon as it was available to neurologists. Consequently, the part of off-label medications reduced over time.

Characteristics of patients according to the three therapeutic strategy groups and the period are described in Table 1, excluding patients started with an off-label treatment.

Escalating strategy

In the escalating group, the median follow-up duration was 12.3 years (IQR= 9.0-16.2). The median age at treatment start was 33 years (IQR= 27-41). In the year preceding treatment start, 34% of patients presented at least two relapses. The annual relapse rate since MS onset was 1.3 (IQR: 0.7-2.5) relapse per person-year. The median delay from MS onset to treatment start was 1.7 years (IQR= 0.7-3.9)

Among the 6,019 patients treated, 4,954 (82%) and 1,065 (18%) started interferon-β and glatiramer acetate for a median duration of 3.0 (IQR= 1.3-6.2) and 2.4 years (IQR= 1.0-5.1) respectively. For 862 patients (14%), this treatment was ongoing at the last clinical information. The Kaplan-Meier estimate of the delay before to the second line treatment start showed that the median time was 17 years [15.8; 18.6] (Figure S1). At five years, the probability of being free from second line treatment was 0.78 [0.77; 0.80] and it decreased to 0.62 [0.61; 0.63] at ten years.

Intensive strategy

Regarding the intensive strategy (N=187), the study period started at natalizumab arrival (2007-2013). The median follow-up duration was 8.7 years (IQR= 6.4-11.9). . The median age at treatment start was 33 (27-41) and 77% patients were women. In the year preceding treatment start, 62% of patients presented at least two relapses. Among patients with an EDSS score available in the year before treatment start (N=140 (75%)), 43% had an EDSS of 3 or more before starting treatment. The median delay from MS onset to treatment start was 1.5 years (IQR= 0.6-4.8).

Most of them started natalizumab (131 (70%)), others had fingolimod (56 (30%)). The median number of natalizumab courses was 47 (IQR= 25- 73) corresponding to a median duration of 3.6 years (IQR= 1.9 - 5.6). The median duration of fingolimod was 3.0 years (IQR= 1.1; 5.1). For 51 patients (27%), the treatment was ongoing at the last clinical information.

Among patients initially treated by natalizumab, 6% stopped after a treatment duration comprised between 2 and 2.5 years. Among all natalizumab stoppers (N=104), the median washout delay before starting another treatment was 5.7 weeks [3.1; 10.9]. It was respectively 7.6 weeks [3.0; 26.4] and 4.6 weeks [2.6; 10.9] before and after 2014. After natalizumab, 56 (54%) patients were treated by fingolimod.

Induction strategy

In this group (N=184), the median follow-up duration was 14.4 years (IQR= 8.9-17.7), 70% were women, 81% of

patients were less than 40 years old at MS onset and the median age at treatment start was 31 years (IQR= 24-39). In the year before treatment start, 56% of patients presented two relapses or more. The annual relapse rate since MS onset was 3.1 (IQR: 1.3-5.3). Among patients with an EDSS score available in the year before treatment start, 72% had an EDSS of 3 or more before starting treatment. The median delay from MS onset to treatment start was 1.0 years (IQR= 0.3-2.3).

The median number of mitoxantrone infusions was 6.5 (IQR= 3.3–7.5). After this first exposure period, 43 (23%) were exposed a second time to a third line treatment (40 to mitoxantrone or 3 to alemtuzumab). The Kaplan-Meier estimate of the delay before a high-efficacy treatment start indicated that, after the first exposure to mitoxantrone, patients were either untreated or treated by a first line treatment during a median time of 10.6 years (Figure S2). The probability of being free from a new highly efficacy treatment was 0.84 [0.78; 0.89] at one year and 0.64 [0.58; 0.72] at five years.

Discussion

This work aimed to describe practices in treatment uses in France, in particular the time from MS clinical onset to first initiation and patients' characteristics according to the three main therapeutic strategies (escalation, intensive, induction).

One of our main findings was the earlier initiation over time. High-efficacy treatment as first intention (induction or intensive strategies) was started soon in the disease development since half of patient started in year after MS onset.

As expected, we showed that most patients were treated in an escalating strategy. This was consistent with recommendations (Montalban *et al.*, 2018). Most of the time, the first-line treatment looked being sufficient during a while. Indeed, the median delay before starting a second line treatment was 17 years given that patients could have been treated by same or other first line treatment or untreated before the need of second line treatment. Some patients (around 15%) were treated despite any new relapse in the year before treatment start. This could be explained by the presence of others signs of activity, especially on MRI, but unavailable in the present study.

As soon as second line treatments were available in France, treatment initiations in a goal of intensive strategy were observed. We found that most of patients started by natalizumab but it could be explained by the earlier availability of

natalizumab (2007) compared with the fingolimod availability date (2011). Moreover, around 40% initiating natalizumab were then treated by fingolimod. It probably corresponds to patients who are JC+ who switched to another treatment as their risk of progressive multifocal leukoencephalopathy was high after 2 years of natalizumab exposure (McGuigan *et al.*, 2016). The median washout was around 8 weeks before 2014 and around 5 weeks given that it is recommended that washout should be no longer than 2 months (Giovannoni and Naismith, 2014). In particular, we observed that the third quartile diminished after 2014 (from 26 weeks to 11 weeks).

A first initiation by mitoxantrone concerned few patients. That appears consistent with literature since this strategy should be reserved for patients with highly active RMMS (Committee for Medicinal Products for Human Use, 2016*b*) and a well-defined patient profile (Edan and Le Page, 2013; D'Amico *et al.*, 2017). As expected, patients were young and presented severe disease with, in particular, a high relapse activity. The initial exposure was 3 infusions for 25% of patients and around 6 infusions for 50%, corresponding to recommendations and allowing a second mitoxantrone treatment later in the disease course, if needed (Le Page *et al.*, 2015).

The main limitation of this study was missing data on EDSS score at treatment start and MRI that did not permit to analyze factors associated with the choice of the strategy. MRI was not systematically collected during the study period. Overtime, the importance of MRI in therapeutic decision increased and the lack of MRI in study should be limited.

To conclude, patients are treated early in the French expert centers, as recommended. The expansion of therapeutic options offers the opportunity of choosing according to patient characteristics and risk of disability worsening. Further work will be done to compare patients outcomes according to the three initial strategy.

References

Comi G, Martinelli V, Rodegher M, Moiola L, Leocani L, Bajenaru O, et al. Effects of early treatment with glatiramer acetate in patients with clinically isolated syndrome. Mult Scler 2013; 19: 1074–83.

Committee for Medicinal Products for Human Use. Assessment report Tysabri. 2016

Committee for Medicinal Products for Human Use. Novantrone EMEA-H-A-30-1399. 2016; 44

Committee for Medicinal Products for Human Use. Assessment report Gilenya. 2018

D'Amico E, Ziemssen T, Cottone S. Induction therapy for the management of early relapsing forms of multiple sclerosis. A critical opinion. Expert Opin Pharmacother 2017; 18: 1553–1556.

Edan G, Le Page E. Induction Therapy for Patients with Multiple Sclerosis : Why ? When ? How ? 2013: 403–409.

Fujihara K. BENEFIT 8-year results provide further support for the long-term value of early treatment of multiple sclerosis. J Neurol Neurosurg Psychiatry 2014; 85: 1179–1179.

Giovannoni G, Naismith RT. Natalizumab to fingolimod washout in patients at risk of PML: When good intentions yield bad outcomes. Neurology 2014; 82: 1196–1197.

Hartung HP, Graf J, Kremer D. Long-term follow-up of multiple sclerosis studies and outcomes from early treatment of clinically isolated syndrome in the BENEFIT 11 study. J Neurol 2019

Kappos L, Edan G, Freedman M, Montalban X, Hartung HP, Hemmer B, et al. The 11-year long-term follow-up study from the randomized BENEFIT CIS trial. Neurology 2016; 87: 978–987.

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: 1900–1913.

McGuigan C, Craner M, Guadagno J, Kapoor R, Mazibrada G, Molyneux P, et al. Stratification and monitoring of natalizumabassociated progressive multifocal leukoencephalopathy risk: Recommendations from an expert group. J Neurol Neurosurg Psychiatry 2016; 87: 117–125.

Montalban X, Gold R, Thompson AJ, Otero-Romero S, Amato MP, Chandraratna D, et al. ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis.

Mult Scler 2018; 24: 96–120.

Le Page E, Deburghgraeve V, Lester MA, Cardiet I, Leray E, Edan G. Alemtuzumab as rescue therapy in a cohort of 16 aggressive multiple sclerosis patients previously treated by Mitoxantrone: an observational study. J Neurol 2015; 262: 1024–1034.

Tintoré M. New options for early treatment of multiple sclerosis. J Neurol Sci 2009; 277: 9–11.

Vukusic S, Casey R, Rollot F, Brochet B, Pelletier J, Laplaud DA, et al. Observatoire Français de la Sclérose en Plaques (OFSEP): A unique multimodal nationwide MS registry in France. Mult Scler J 2018: 1–5.

Tables and Figures

Figure 1: Flowchart

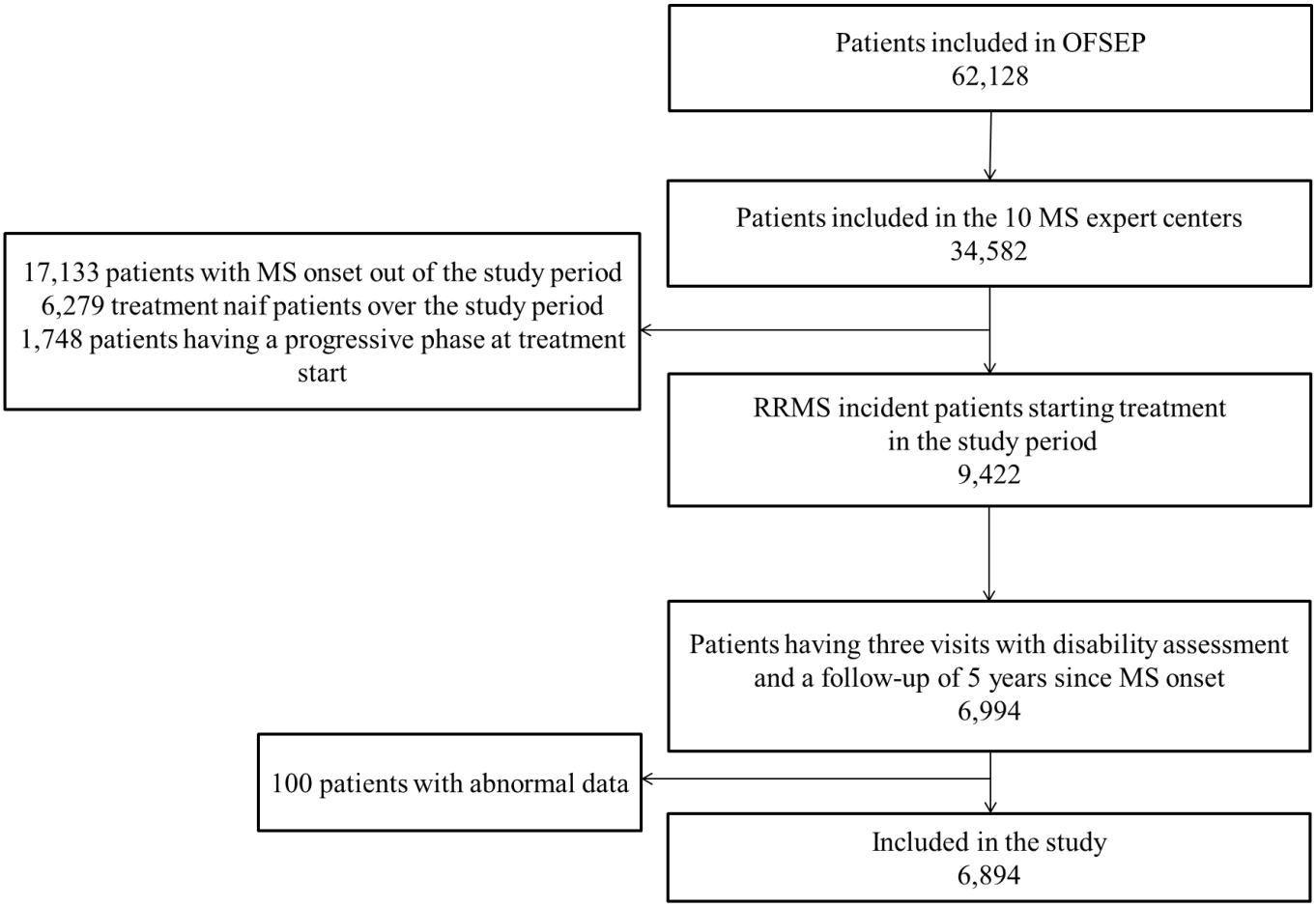


Figure 2: Delay from MS onset to treatment start according to year of MS onset

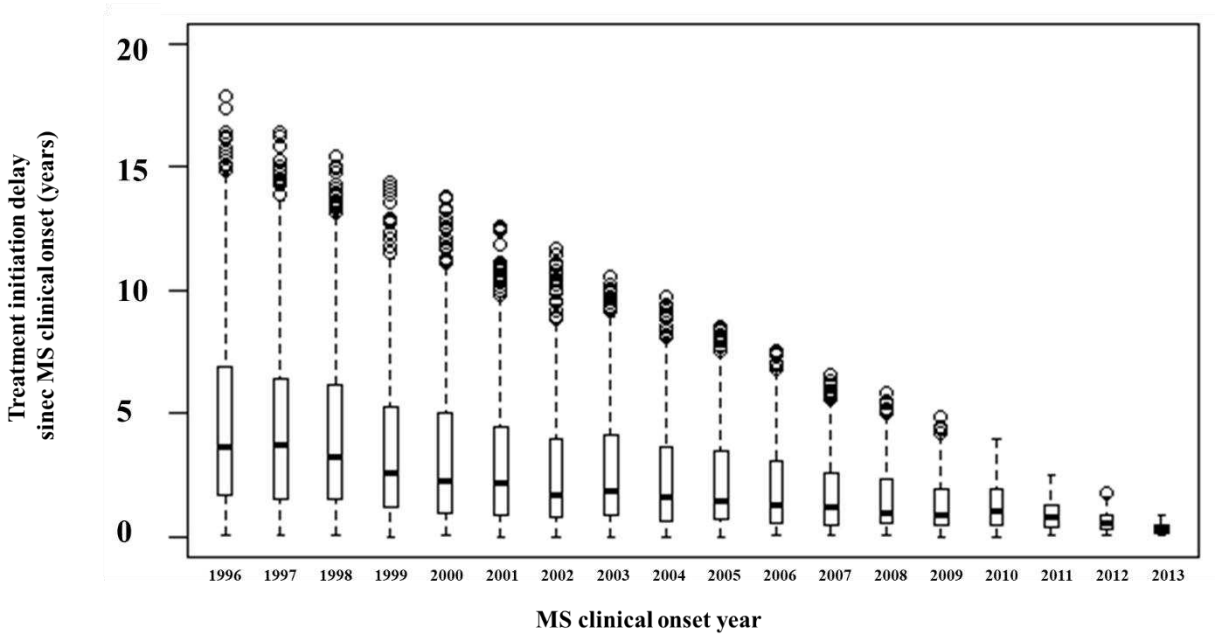
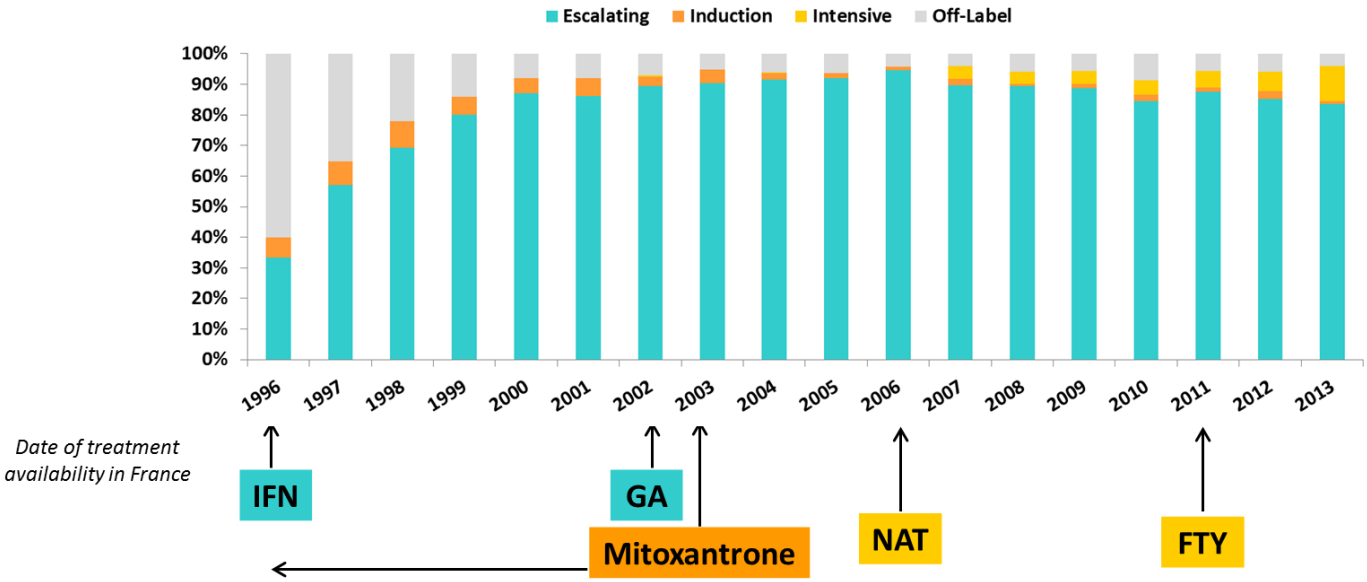


Figure 3: Distribution of the three therapeutic strategies over time (1996-2013) according to the year of first treatment initiation



IFN= Interferons-β, GA= Glatiramer Acetate, NAT= natalizumab, FTY=Fingolimod

Table 1: Patient's characteristics at first treatment initiation by therapeutic strategy and period of first treatment initiation

	Period 1996 - 2013		Period 2007-2013		
	Escalating N= 6019	Induction N= 184	Escalating N=2923	Intensive N=187	Induction N= 53
Sex					
Female	4518 (75.1%)	128 (69.6%)	2200 (75.3%)	144 (77%)	35 (66%)
Male	1501 (24.9%)	56 (30.4%)	723 (24.7%)	43 (23%)	18 (34%)
Age at MS clinical onset	30.3 (24.2-38.2)	29.5 (23.7-37.2)	30.3 (24.3-38.6)	29.8 (24.5-37)	32.4 (24.6-39.9)
Age at treatment start	33.4 (26.8-41.1)	30.6 (24.5-38.9)	34.1 (27.1-41.9)	33 (27.2-41.4)	34.8 (26.1-42.2)
MS duration at treatment start	1.7 (0.7-3.9)	1 (0.4-2.3)	1.8 (0.7-4.8)	1.6 (0.6-4.8)	0.9 (0.3-3)
Number of relapses within 12 months before treatment start					
0	880 (14.6%)	15 (8.2%)	482 (16.5%)	21 (11.2%)	3 (5.7%)
1	3065 (50.9%)	65 (35.3%)	1595 (54.6%)	50 (26.7%)	23 (43.4%)
2	1611 (26.8%)	63 (34.2%)	688 (23.5%)	83 (44.4%)	18 (34%)
3	357 (5.9%)	32 (17.4%)	129 (4.4%)	27 (14.4%)	9 (17%)
4	82 (1.4%)	5 (2.7%)	22 (0.8%)	6 (3.2%)	0 (0%)
5 and more	24 (0.4%)	4 (2.2%)	7 (0.2%)	0 (0%)	0 (0%)
EDSS within 12 months before treatment start					
0	747 (12.4%)	2 (1.1%)	439 (15%)	11 (5.9%)	0 (0%)
1	1164 (19.3%)	12 (6.5%)	624 (21.3%)	22 (11.8%)	5 (9.4%)
2	984 (16.3%)	19 (10.3%)	512 (17.5%)	33 (17.6%)	8 (15.1%)
3	387 (6.4%)	34 (18.5%)	169 (5.8%)	27 (14.4%)	7 (13.2%)
4	210 (3.5%)	17 (9.2%)	114 (3.9%)	25 (13.4%)	7 (13.2%)
5 and more	77 (1.3%)	36 (19.6%)	32 (1.1%)	22 (11.8%)	7 (13.2%)
NA	2450 (40.7%)	64 (34.8%)	1033 (35.3%)	47 (25.1%)	19 (35.8%)
Annual relapse rate since MS onset	1.3 (0.7-2.5)	3 (1.3-5.3)	1.2 (0.6-2.4)	1.7 (0.7-3.7)	3.1 (1.3-5.3)

Supplementary material

Figure S1: Kaplan-Meier estimates of the time before switching with a second line treatment for patients in escalating strategy

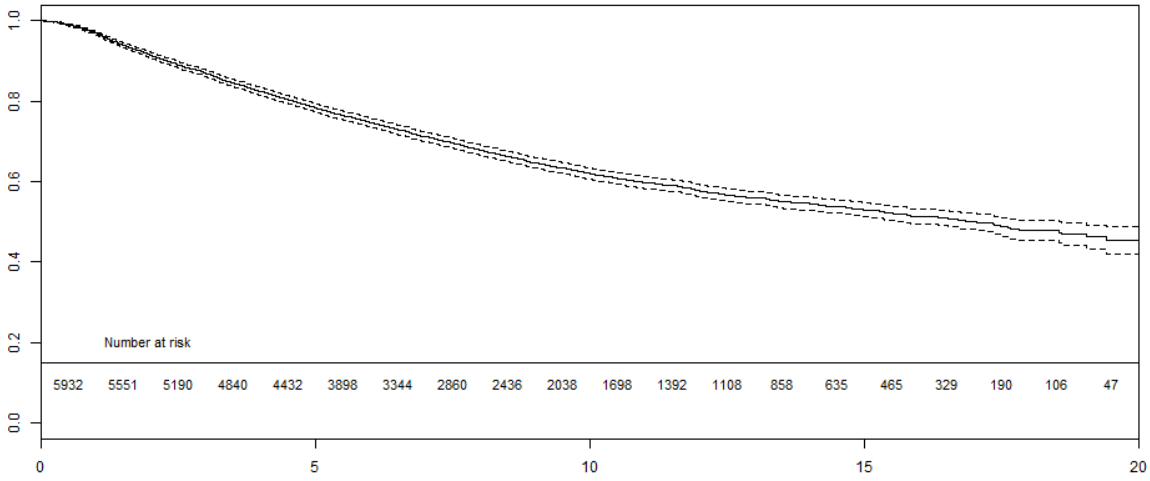
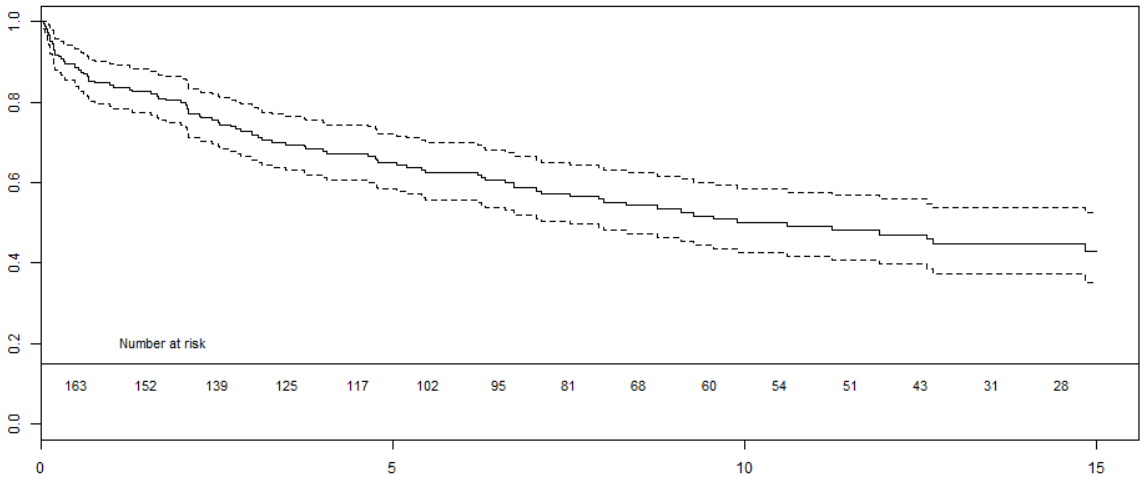


Figure S2: Kaplan-Meier estimates of the time before being exposed a second time to high-efficacy treatment for patients in induction strategy



C. Efficacité de l'initiation précoce d'un traitement : résultats préliminaires

Le deuxième objectif de ce travail porte sur l'évaluation de l'efficacité de l'initiation d'un traitement précoce en vie réelle. Il s'agit de résultats préliminaires qui seront complétés par la suite avant rédaction et soumission d'un article sur le sujet.

1. Contexte

Comme évoqué dans le contexte de ce chapitre, la prise en charge thérapeutique précoce des patients atteints de SEP semble aujourd'hui être devenu le « gold standard ». Néanmoins, l'intérêt d'une action précoce sur la progression du handicap reste à confirmer par des études en vie réelle (Freedman *et al.*, 2014; Fernández, 2017; Merkel *et al.*, 2017). Toutefois, certaines études observationnelles ont déjà fourni des résultats tendant à montrer que l'initiation précoce du traitement est bénéfique pour limiter la progression du handicap (Trojano *et al.*, 2009; Cocco *et al.*, 2015; Kavaliunas *et al.*, 2016; Landfeldt *et al.*, 2018). En particulier, une étude réalisée en Suède a montré que le risque d'atteindre un score EDSS de 4 augmentait de plus de 7% pour chaque année de retard du traitement après le début de la SEP (Kavaliunas *et al.*, 2016).

Outre la question même de l'efficacité, se pose la question de la définition d'une initiation précoce. En effet, les travaux proposés à ce jour, que ce soit des études issues d'essais randomisés ou des études observationnelles, fournissent des définitions variées de la précocité d'initiation.

Dans les extensions d'essais randomisés, les patients sont suivis à l'aveugle dans le groupe issu de la randomisation pendant un certain temps. Une fois l'essai clinique conclu, les patients ayant reçu le placebo peuvent être traités, permettant une comparaison des deux groupes traités dans des délais différents. Le délai de retard est de l'ordre de 6 à 24 mois par rapport à la randomisation (Merkel *et al.*, 2017).

Dans les études observationnelles, ce délai peut être allongé : de 1 à 5 ans par exemple (Trojano *et al.*, 2009). D'autres études, en utilisant le délai depuis le début de la SEP en continu, estiment l'effet associé à une année de maladie supplémentaire sans traitement (Kavaliunas *et al.*, 2016). D'autres caractéristiques, associées à l'activité de la maladie, peuvent également être

utilisées pour définir un traitement initié tôt, au regard du développement du handicap (avant et après un stade donné) (Cocco *et al.*, 2015) ou de l'âge (Merkel *et al.*, 2017), par exemple.

Dans cette étude, plusieurs définitions ont été ou seront envisagées pour évaluer l'efficacité de l'initiation précoce d'un traitement. Dans un premier temps, les traitements seront analysés dans un seul et même groupe, puis, dans un second temps, des analyses stratifiées sur le type de traitement initié seront conduites.

Cette étude est présentée dans son état d'avancement et sera complétée avant la rédaction de la publication. Nous présentons ici quelques résultats préliminaires et les perspectives envisagées.

2. Patients et méthodes

Cette partie du travail est basée sur une population proche de celle étudiée dans le travail descriptif présenté précédemment. Nous rappelons brièvement quelques éléments en ajoutant les compléments spécifiques à cette partie de l'analyse.

2.1 Population d'étude

A partir de la sélection de 10 centres participant à l'OFSEP (Lyon, Nancy, Bordeaux, Rennes, Toulouse, Nice, Clermont-Ferrand, Dijon, Besançon, Nîmes), les patients incidents, présentant une SEP évoluant par poussée et initiant un traitement entre le 1er janvier 1996 et le 31 décembre 2013, ont été analysés. Comme précédemment, 3 visites aux centres experts pour lesquelles une mesure d'EDSS était disponible étaient requises pour l'inclusion, ainsi qu'une durée de SEP d'au moins 5 ans.

Il s'ajoute aux critères précédemment définis que, pour être inclus, les patients ne devaient pas avoir atteint un score EDSS 3 confirmé à 6 mois et maintenu jusqu'aux dernières nouvelles au moment de l'initiation du traitement.

2.2 Exposition

Tout patient exposé au moins un jour à l'un des traitements disponibles sur la période d'étude, était considéré traité. Sur la période 1996-2013, étaient disponibles : les interférons-β, l'acétate de glatiramère, le natalizumab, le fingolimod et la mitoxantrone.

Dans un premier temps, la variable d'intérêt a été définie en utilisant la durée de la SEP. Ainsi, ont été considérés comme des traitements initiés tôt dans trois analyses séparées :

- Définition #1 : un traitement initié dans l'année suivant le début de SEP
- Définition #2 : un traitement initié dans les 2 ans suivant le début de SEP
- Définition #3 : un traitement initié dans les 3 ans suivant le début de SEP

2.3 Critères de jugement

A ce stade, trois critères de jugement ont été considérés : (1) le temps entre l'initiation de traitement et l'atteinte de l'EDSS 3 confirmé à 6 mois et maintenu jusqu'aux dernières nouvelles, (2) le temps entre l'initiation du traitement et l'atteinte de l'EDSS 6 confirmé (6 mois) et maintenu jusqu'aux dernières nouvelles et (3) le temps entre l'initiation du traitement et la survenue d'une nouvelle poussée.

2.4 Analyse statistique

Dans le but de contrôler le biais d'indication, un score de propension a été utilisé. Il a été construit par régression logistique en intégrant les variables associées au critère de jugement. Ainsi, pour chaque critère de jugement, un score de propension spécifique était défini. Les variables suivantes ont été considérées : sexe, âge à l'initiation du traitement, le nombre de poussées dans les 12 mois précédant l'initiation, la réalisation d'un premier examen en centre expert et le centre de prise en charge. Le nombre de poussées était défini comme une variable à deux catégories : une poussée ou aucune dans l'année précédant l'initiation et deux poussées ou plus. En raison d'un nombre important de données manquantes sur le score EDSS avant l'initiation, celui-ci n'a pu être utilisé comme variable de confusion.

Les critères de jugement ont été étudiés en utilisant un modèle de Cox pondéré par les poids sIPTW, avec un estimateur robuste de la variance. Les analyses ont été réalisées sur le logiciel R (R 3.4.0).

3. Résultats

Après application des critères d'inclusion, 5987 patients ont été inclus. En considérant la définition de traitement précoce avec le seuil à un an (définition #1), 2178 (36%) patients étaient considérés comme initiant le traitement tôt. Le seuil de deux ans (définition #2) conduisait à considérer 3393 (57%) patients traités tôt et le seuil de 3 ans (définition #3) à 4122 (69%) patients.

Globalement, considérant notre critère de jugement principal (EDSS 3 confirmé et maintenu), la durée médiane de suivi était de 9,46 ans [9,31 ; 9,62]. La durée médiane de suivi dans chaque groupe et pour chaque critère de jugement, puisqu'elles ont été déterminées en utilisant la méthode de Kaplan-Meier inversé, sont résumées dans le Tableau 5.

Tableau 5 : Durées de suivi selon la définition d'un traitement initié tôt et des critères de jugements considérés dans l'analyse de l'efficacité d'un traitement précoce

	Etude du temps d'atteinte de EDSS 3		**Etude du temps d'atteinte de EDSS 6**		**Etude du temps avant la première poussée**	
	Tôt	Tardif	Tôt	Tardif	Tôt	Tardif
Definition #1	9,7 [9,4 ; 9,9]	9,3 [9,1; 9,5]	9,7 [9,4 ; 9,9]	9,2[9,0 ; 9,5]	8,8 [8,3 ; 9,5]	8,5 [8,0 ; 8,9]
Definition #2	9,8 [9,6 ; 10,0]	8,9 [8,7 ; 9,2]	9,8 [9,6 ; 10,0]	8,9 [8,7 ; 9,1]	8,9 [8,6 ; 9,3]	8,2 [7,7 ; 8,6]
Definition #3	9,8 [9,6 ; 10,0]	8,6 [8,2 ; 8,9]	9,8 [9,6 ; 10,0]	8,6 [8,2 ; 8,8]	8,9 [8,6 ; 9,3]	7,9 [7,4 ; 8,3]

Les caractéristiques des patients dans chacun des groupes sont présentées dans le Tableau 6. Le sexe semble équitablement réparti selon que les patients soient traités tôt ou non avec un ratio hommes-femmes similaire à celui que l'on connait pour la SEP. Les patients traités tôt avaient en moyenne 32 ans à l'initiation du traitement et présentaient à 55% deux poussées ou plus. Quelle que soit la définition du groupe « tardif », environ 75% des patients de ce groupe ont présenté aucune poussée ou une poussée dans l'année précédant l'initiation.

L'étude des associations entre les variables considérées et les critères de jugement a conduit à considérer deux scores de propension. Le délai d'atteinte d'un score EDSS confirmé et maintenu de 3 était associé aux variables suivantes : le sexe, l'âge à l'initiation du traitement, le nombre de poussées dans les 12 mois précédant l'initiation, la réalisation d'un examen au centre expert et le centre de prise en charge. Il en était de même pour le délai avant la survenue d'une nouvelle poussée après l'initiation du traitement. En revanche, le délai d'atteinte d'un score EDSS 6 confirmé et maintenu n'était pas associé au nombre de poussées dans les 12 mois précédant l'initiation. Cette variable n'a par conséquent pas été incluse dans le modèle estimant le score de propension relatif à ce critère de jugement.

Afin d'évaluer la qualité du chevauchement de la distribution des scores de propension selon le groupe étudié, des histogrammes en miroir sont proposés Figure 21. Pour chacune des analyses, un bon chevauchement des distributions était observé.

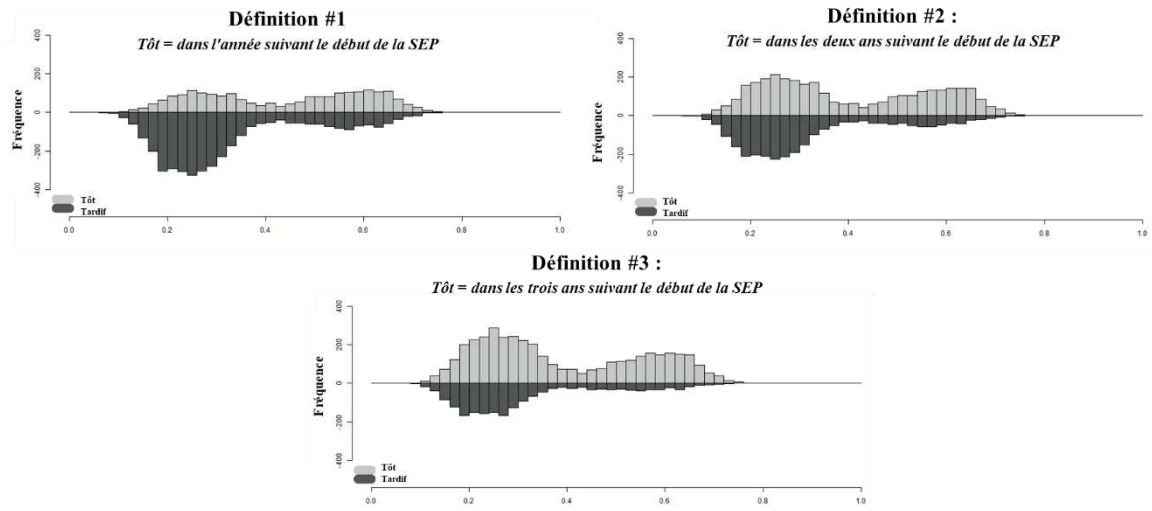


Figure 21 : Distribution des scores de propension relatifs aux critères de jugement EDSS 3 et de poussées pour chaque définition de groupes

Tableau 6 : Caractéristiques des patients selon les définitions d'initiation précoce

	Total	Définition #1 : *Tôt = dans l'année suivant le début de la SEP*		Définition #2 : *Tôt = dans les deux ans suivant le début de la SEP*		Définition #3 : *Tôt = dans les trois ans suivant le début de la SEP*	
	N= 5987	Tôt N= 2178	Tardif N=3809	Tôt N=3393	Tardif N=2594	Tôt N=4122	Tardif N=1865
Sexe							
Femme	4522 (75,5%)	1658 (76,1%)	2864 (75,2%)	2569 (75,7%)	1953 (75,3%)	3120 (75,7%)	1402 (75,2%)
Homme	1465 (24,5%)	520 (23,9%)	945 (24,8%)	824 (24,3%)	641 (24,7%)	1002 (24,3%)	463 (24,8%)
Age à l'initiation du traitement		32 (9,6)	35 (9,7)	32 (9,7)	36 (9,5)	32 (9,7)	37 (9,3)
Visite en centre expert							
Au moins une	3570 (59,6%)	1258 (57,8%)	2312 (60,7%)	1990 (58,7%)	1580 (60,9%)	2424 (58,8%)	1146 (61,4%)
Aucune	2417 (40,4%)	920 (42,2%)	1497 (39,3%)	1403 (41,3%)	1014 (39,1%)	1698 (41,2%)	719 (38,6%)
Poussées							
Une poussée ou aucune	3838 (64,1%)	977 (44,9%)	2861 (75,1%)	1865 (55%)	1973 (76,1%)	2419 (58,7%)	1419 (76,1%)
Deux poussées ou plus	2149 (35,9%)	1201 (55,1%)	948 (24,9%)	1528 (45%)	621 (23,9%)	1703 (41,3%)	446 (23,9%)
Centre							
centre 1	307 (5,1%)	119 (5,5%)	188 (4,9%)	181 (5,3%)	126 (4,9%)	213 (5,2%)	94 (5%)
centre 2	601 (10%)	211 (9,7%)	390 (10,2%)	308 (9,1%)	293 (11,3%)	387 (9,4%)	214 (11,5%)
centre 3	443 (7,4%)	164 (7,5%)	279 (7,3%)	246 (7,3%)	197 (7,6%)	294 (7,1%)	149 (8%)
centre 4	215 (3,6%)	88 (4%)	127 (3,3%)	128 (3,8%)	87 (3,4%)	151 (3,7%)	64 (3,4%)
centre 5	1668 (27,9%)	619 (28,4%)	1049 (27,5%)	993 (29,3%)	675 (26%)	1190 (28,9%)	478 (25,6%)
centre 6	400 (6,7%)	162 (7,4%)	238 (6,2%)	237 (7%)	163 (6,3%)	291 (7,1%)	109 (5,8%)
centre 7	575 (9,6%)	213 (9,8%)	362 (9,5%)	333 (9,8%)	242 (9,3%)	407 (9,9%)	168 (9%)
centre 8	662 (11,1%)	238 (10,9%)	424 (11,1%)	380 (11,2%)	282 (10,9%)	464 (11,3%)	198 (10,6%)
centre 9	339 (5,7%)	124 (5,7%)	215 (5,6%)	188 (5,5%)	151 (5,8%)	226 (5,5%)	113 (6,1%)
centre 10	777 (13%)	240 (11%)	537 (14,1%)	399 (11,8%)	378 (14,6%)	499 (12,1%)	278 (14,9%)

Les différences standardisées avant et après pondération sont présentées en Figure 22. Pour chaque étude, la qualité de l'équilibre obtenu sur les variables introduites dans les modèles spécifiant les scores de propension était assez satisfaisante, réduisant les différences observées initialement entre les groupes. On note, tout de même, une difficulté à équilibrer la variable relative au nombre de poussées entre les deux groupes lorsqu'ils sont définis selon la définition #3.

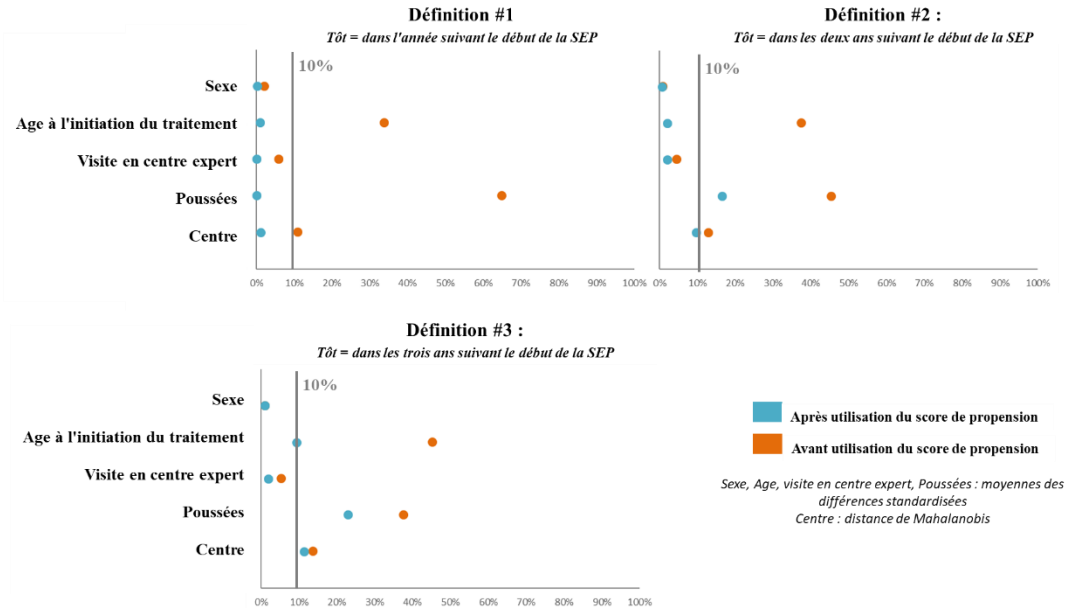


Figure 22 : Différences standardisées entre les groupes selon les trois définitions (score de propension associée aux critères de jugement EDSS 3 et poussées)

Pour ne pas multiplier les résultats, nous présentons seulement les résultats d'évaluation de l'équilibre obtenu avec le score relatif au critère de jugement de l'EDSS 3 ou de poussée. Les données équivalentes pour le critère de jugement relatif à EDSS 6 sont présentées en annexe F. On note également une évaluation du score satisfaisante.

La Figure 23 résume les résultats obtenus selon chacune des définitions et critères de jugement étudiés.

Ces résultats (HR<1) tendent à montrer la réduction du risque d'atteinte de l'EDSS 3 en débutant le traitement au plus tôt.

Un résultat similaire est observé concernant le temps d'atteinte d'un score EDSS 6 pour la première définition de la précocité. Dans les autres cas, les intervalles de confiance, plus larges témoignent peut-être d'un manque de puissance.

Enfin, en étudiant le délai avant la première poussée, aucune différence n'a pu être mise en évidence.

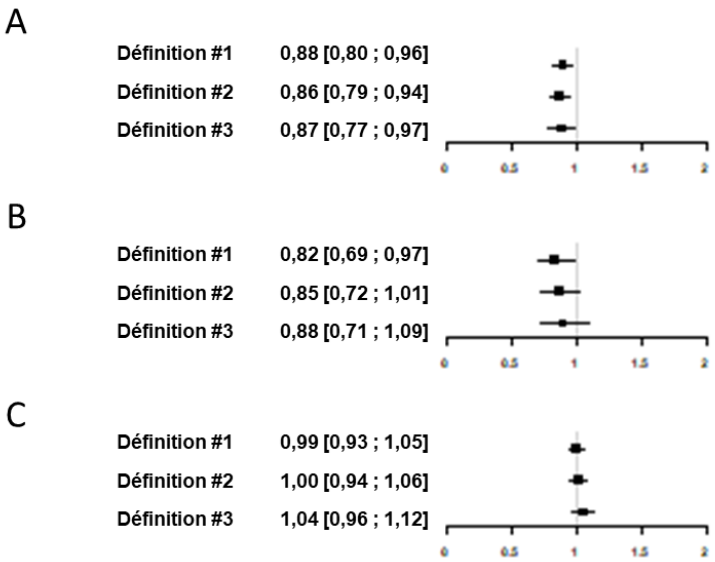


Figure 23 : Risque de survenue du critère de jugement associé à l'initiation précoce du traitement
A : EDSS 3, B : EDSS 6, C : nouvelle poussée

4. Discussion et perspectives

Ces résultats préliminaires tendent à montrer un bénéfice de l'initiation du traitement précoce sur le handicap que ce soit dans l'année, les deux ans ou les trois ans suivant le début de SEP. Nous ne sommes pas en mesure de conclure sur le risque de survenues d'une nouvelle poussée. De plus, les analyses portant sur le temps d'atteinte de l'EDSS 6 confirmé et maintenu manquent de puissance.

Les résultats obtenus sur la progression du handicap convergent avec ceux obtenus dans les études précédemment conduites (Trojano *et al.*, 2009; Kavaliunas *et al.*, 2016). Le groupe italien (Trojano *et al.*, 2009) a étudié le temps d'atteinte, depuis l'initiation du traitement, d'un score EDSS 4 et 6 dans deux groupes de patients : l'un débutant les interférons-β dans l'année suivant le début de la SEP, l'autre au-delà. Les résultats tendaient à montrer un bénéfice mais n'étaient pas significatifs (EDSS 4 : HR= 0,63 [0,40 ;1,01] et EDSS 6 : HR= 0,58 [0,25 ; 1,33]).

En revanche, l'étude de la première progression d'un point du score EDSS démontrait une réduction du risque de progression dans le groupe traité tôt (HR= 0,61 [0,46; 0,82]). Le groupe suédois (Kavaliunas *et al.*, 2016) avait également observé un bénéfice du traitement précoce puisque, depuis le début de la SEP, les patients débutant le traitement trois ans après le début de la SEP présentaient un risque plus élevé d'atteindre un EDSS 4 plus vite en comparaison aux patients démarrant le traitement dans l'année (HR= 2,64 [1,71 ; 4,08]). L'étude de l'initiation entre un et trois ans ne permettait pas de montrer de différence significative par rapport au démarrage du traitement avant un an (HR= 1,37 [0,80 ; 2,35]).

Concernant l'étude de la survenue d'une nouvelle poussée, aucune différence n'a pu être démontrée. Pour ce type de critères de jugement témoignant d'une nouvelle activité après la mise sous traitement (nouvelle poussée, nouvelle lésion, progression d'un point sur l'échelle du handicap), il est parfois proposé de retarder le temps d'évaluation de quelques mois afin de ne pas tenir compte des premiers mois de traitements qui n'ont pas un effet immédiat sur l'activité de la maladie. Dans ce contexte il est recommandé de « décaler » la date d'évaluation, le terme « rebaseline » est employé. Des temps plus ou moins longs sont proposés selon le traitement considéré (3 à 6 mois pour les interférons-β, le diméthyle fumarate, le tériflunomide et le natalizumab ; 9 à 12 mois pour le l'acétate de glatiramère par exemple). Bien qu'une telle modification semble justifiée, on exerce une modification après la mise en route du traitement ce qui impacte la comparabilité faite au moment de l'initiation du traitement. En revanche, en analyse de sensibilité, on pourrait éventuellement envisager de mesurer les variables à une date postérieure à l'initiation pour rendre les patients comparables à ce moment-là.

Après utilisation de la pondération sIPTW, les deux groupes présentaient un équilibre satisfaisant bien qu'il soit parfois difficile d'équilibrer sur la variable relative au nombre de poussées. Dans la suite, nous chercherons à améliorer cet aspect. De plus, il est important de faire remarquer que le score ne tenait pas compte de facteurs tels que le score EDSS ponctuel à l'initiation du traitement ou les données d'imagerie. Les données d'imagerie sont souvent manquantes dans nos études. Il était prévu d'en tenir compte dans ce travail. Malheureusement, le manque de données ou l'hétérogénéité des données disponibles dans l'extraction fournie par l'OFSEP nous ont conduit à ne pas les étudier dans l'immédiat. Une étude descriptive plus poussée serait nécessaire pour clarifier ces données et ne pas risquer d'introduire un biais de sélection et/ou des problèmes d'ajustement en les utilisant.

Une autre des limites de ce travail concerne la sélection des centres. Au regard de nos critères de jugement (notamment EDSS 6), il nous a semblé préférable d'inclure des centres

pour lesquels les données de suivi de patients étaient collectées depuis suffisamment de temps (2001). Ainsi, les 10 centres ont été choisis par rapport à la durée de leur collecte de données. En revanche, faisant ce choix, nous accentuons peut-être le biais de recrutement déjà présent dans l'OFSEP.

Dans cette étude préliminaire, nous avons étudié le fait d'être traité tôt, tout type d'exposition confondu. Dans la suite, nous réaliserons des analyses stratifiées par type de traitements. Un sous-groupe initiant un traitement d'escalade sera considéré, ainsi qu'un sous-groupe débutant un traitement dit de haute efficacité (natalizumab, fingolimod, mitoxantrone).

Au-delà du type de traitement, la période d'étude a pu avoir un impact sur notre résultat. En effet, cette étude a été conduite sur l'ensemble de la période (1996-2013), sans tenir compte d'un effet période probablement important au regard de l'évolution thérapeutique marquante des années étudiées. Initialement, nous avions souhaité étudier une période plus restreinte (2001-2010). Des données supplémentaires ont été demandées à l'OFSEP car nous étions dans l'impossibilité de réaliser l'analyse. En effet, sur cette période et cette population, le délai médian d'initiation de traitement était de moins d'un an, la plupart des patients étant traités dans les 2 ans. Avec peu de patients « contrôles », nos analyses manquaient de puissance. Notons également que sur cette période, les pratiques apparaissaient très homogènes dans les 10 centres considérés. Dans la suite, nous conduirons des analyses stratifiées sur la période d'initiation.

Enfin, les résultats seront complétés en utilisant d'autres critères de jugement, ainsi que d'autres variables d'intérêts. En effet, à ce jour, les définitions du traitement précoce n'ont été établies qu'à partir du délai de la SEP. Dans la suite, nous envisageons d'analyser également une variable d'intérêt déterminée à partir de l'âge du patient au démarrage du traitement. Plusieurs seuils pourront être envisagés.

Cette partie du travail sera poursuivie après la soumission du manuscrit de thèse. Nous avons pour projet de la publier au cours de l'année 2020.

D. Comparaison de l'efficacité des stratégies thérapeutiques : perspectives

1. Contexte et objectif

La dernière partie de ce travail à laquelle nous n'avons pas pu répondre à ce jour concerne la comparaison de l'efficacité des stratégies thérapeutiques. Nous faisons le choix de l'évoquer malgré tout car la situation semble être intéressante concernant la question de l'analyse des données observationnelles et la discussion de l'intégralité de cette thèse.

Comme évoqué précédemment, les stratégies thérapeutiques n'ont, à notre connaissance, pas encore été comparées. En particulier, il est formulé dans la littérature, un manque de données sur l'efficacité d'un traitement d'induction (D'Amico *et al.*, 2017). Au regard du profil de tolérance de ce traitement, nous pourrions même parler d'un manque de démonstration de supériorité. Ainsi, il nous semble important de le documenter à partir d'une étude observationnelle.

L'objectif de cette partie du travail était donc de comparer les patients traités par induction à des patients comparables, au sens « éligibles » à l'induction, qu'ils aient été traités dans une stratégie d'escalade et/ou intensive. Cette comparaison porte essentiellement sur des critères d'efficacité, mais sera mise au regard, dans la mesure de leur disponibilité, d'éléments relatifs à la tolérance des traitements et aux effets indésirables.

Cette comparaison nous semble n'avoir de sens que sur une population très spécifique de patients présentant une SEP sévère que l'on pourrait considérer « éligible » au traitement par induction. Pour sélectionner ces patients, il est de fait nécessaire de disposer d'informations précises sur l'état clinique et IRM des patients au moment de l'initiation. Comme nous avons déjà pu le constater dans les deux parties précédentes, les données que nous avons actuellement à disposition, ne nous permettent pas de répondre de manière optimale à cet objectif. De fait, nous avons préféré revoir le protocole d'étude. Nous pensons notamment qu'une relecture des dossiers médicaux et la collecte du nombre de lésions IRM à l'initiation du traitement sont nécessaires pour répondre au mieux aux objectifs de ce travail. Nous proposons ci-après les étapes du protocole envisagé.

2. Protocole d'étude

L'étape clé du protocole de cette étude réside dans la définition de « l'éligibilité » au traitement d'induction. Nous envisageons la définition basée sur les critères suivants :

(1) Présenter une SEP évoluant par poussée
(2) Être âgé de moins de 40 ans au début de la SEP
(3) Présenter au moins 2 poussées dans l'année précédant l'initiation
(4) Présenter au moins 2 poussées en T2 ou avec prise de gadolinium dans l'année précédant l'initiation
(5) Ayant un EDSS < 4 un an avant l'initiation du traitement
(6) Présentant un score EDSS ≥ 4 à l'initiation

Comme évoqué précédemment, il apparait important de documenter plus précisément les données cliniques et d'imagerie pour les patients inclus dans cette étude. Nous proposons donc de sélectionner quelques centres avec lesquels nous pourrons travailler pour compléter les données. En particulier, la distribution des stratégies thérapeutiques dans les 10 centres initialement inclus, sur la période 2007-2013, est présentée en Figure 24.

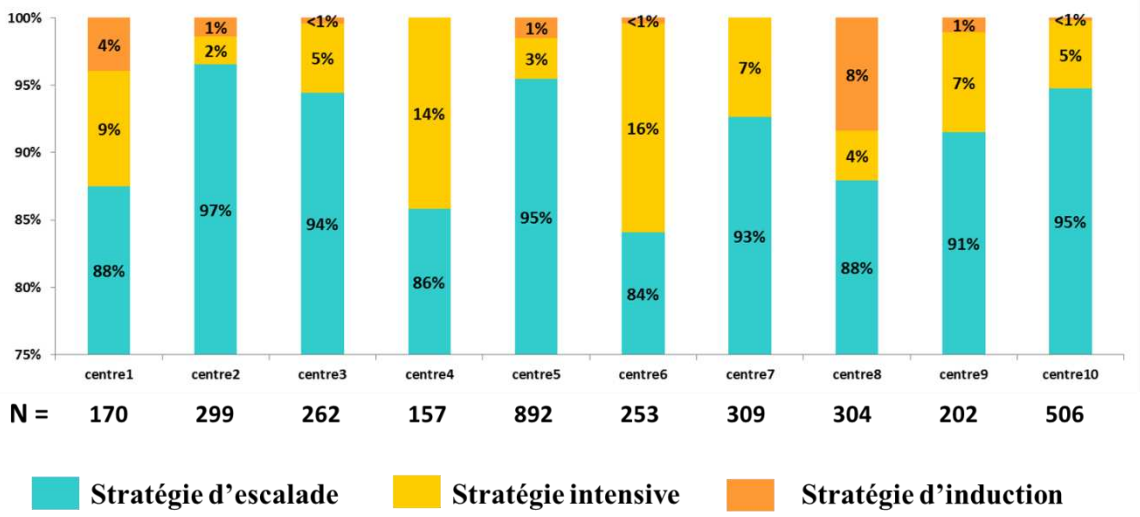


Figure 24 : Distribution des stratégies thérapeutiques initiées par centre sur la période 2007-2013

On constate que dans un centre (centre 8), la pratique d'induction est plus largement utilisée que dans les autres. De plus, le centre 6 présente très peu de patients traités par induction

mais une proportion importante de prise en charge intensive. Enfin, le centre 2, présente peu d'utilisation de traitements dits de haute efficacité.

Bien que les populations prises en charge dans les différents centres experts en France puissent être différentes, il ne semble pas y avoir de raison de penser que les patients présentant des formes sévères de SEP soient concentrés dans certaines régions de France. Ainsi, nous faisons l'hypothèse que ces différences sont essentiellement le fruit de différences de pratiques. Après vérification d'une telle hypothèse, en considérant notamment les effectifs de patients inclus selon les critères d'inclusion précédemment proposés, nous envisageons de reprendre tous les dossiers des patients inclus afin d'obtenir une collecte précise du score EDSS au moment de l'initiation du traitement et de la charge lésionnelle pour l'ensemble des patients répondants aux critères d'inclusion.

A partir de ces données révisées, en faisant le nécessaire dans le but de contrôler le potentiel biais d'indication subsistant malgré une sélection importante des patients, nous analyserons divers critères de jugement à court, moyen et long terme portant sur l'activité de poussées et la progression du handicap.

Chapitre 6

VI. Discussion générale

A. Les résultats principaux 202
1. Efficacité d'un traitement en vie réelle 203
2. Comparaison d'efficacité de traitements en vie réelle 203
3. Evaluation des stratégies thérapeutiques 204
B. La matière première : les connaissances, les données et les méthodes 205
1. Les données observationnelles : une richesse non exhaustive 206
1.1 OFSEP : des données longitudinales, rapportant l'état clinique du patient 206
1.2 OFSEP : des données collectées essentiellement en centre expert 207
1.3 OFSEP : des données non exhaustives 210
2. L'utilisation des méthodes 212
2.1 DAG de la relation entre les traitements de fond sur la progression du handicap dans la SEP rémittente 212
2.2 Les critères d'inclusion 215
2.3 Utilisation d'un score de propension 216
C. Les résultats et conclusions 219
1. L'exposition, une variable dépendante du temps 219
2. Les analyses en intention de traiter et en per protocol 222
3. Question d'efficacité à long terme en vie réelle : entre volonté et faisabilité 223
D. Un travail au carrefour de deux disciplines 224

Ce dernier chapitre a pour objectif de proposer une vision plus transversale des trois chapitres précédents. Mes idées se mélangent à deux sentiments paradoxaux. Si sur le plan individuel, il me semble atteindre une forme d'aboutissement, sur le plan professionnel, me guette un sentiment d'inachevé. Il est probable qu'il soit l'essence d'un travail de recherche puisqu'il n'y a pas une réponse mais des éléments de réponses qui doivent s'interpréter et se compléter dans un contexte global et discuté.

Je choisis de démarrer ma discussion ainsi car c'est certainement une des différences les plus marquantes entre la personne qui a débuté cette thèse il y a trois ans et celle qui la discute aujourd'hui. Perfectionniste, j'ai débuté ma thèse avec un calendrier bien organisé, à chaque trimestre un objectif, tout devait se dérouler par étape mais ce fut loin d'être le cas. J'ai découvert ce qu'était un travail de recherche et que le perfectionnisme résidait parfois dans la reconnaissance des imperfections et des limites.

J'ai organisé cette discussion en trois sous-parties portant tout d'abord sur la matière première, incluant les données et la connaissance contextuelle et méthodologique, puis les réponses et les conclusions apportées, enfin, des éléments plus transversaux sur le positionnement de ce travail au carrefour de plusieurs disciplines seront apportés. En premier lieu, je propose de revenir sur les principaux résultats de cette thèse.

A. Les résultats principaux

Ce travail de recherche avait pour objectif d'évaluer l'effet des stratégies thérapeutiques en conditions réelles d'utilisation sur la progression du handicap chez les patients atteints de SEP rémittente. Dans ce contexte, quatre objectifs ont été explorés ainsi qu'un objectif secondaire qui portait sur l'exploration des méthodes de score de propension. Ce dernier a été exploré de manière transversale. Nous revenons dans cette partie sur les principaux résultats présentés dans les trois chapitres précédents.

1. Efficacité d'un traitement en vie réelle

Le premier axe, abordé dans le chapitre 3, portait sur l'étude de l'efficacité des traitements de première ligne injectables sur la progression du handicap en conditions réelles d'utilisation. Ce travail a été réalisé sur les données collectées au centre expert de Rennes.

L'état de l'art a conduit à constater que malgré l'utilisation de plus en plus importante des scores de propension, il reste souvent des problèmes méthodologiques liés au biais d'indication. Le manque de contrôle du biais d'indication réside essentiellement dans la difficulté à établir le moment de la décision thérapeutique, c'est-à-dire la date d'entrée dans une étude comparative d'efficacité. Après de nombreuses réflexions sur le groupe contrôle à définir et la manière d'utiliser le score de propension, nous avons proposé d'utiliser une méthode jamais utilisée dans la SEP consistant en un appariement sur un score de propension dépendant du temps.

L'étude du délai d'atteinte d'un niveau de handicap modéré a conduit à montrer une faible différence entre les groupes traité et « pas encore traité », en faveur du second. Aucune différence n'a pu être observée sur l'aggravation du handicap.

Ce résultat inattendu est probablement lié à des facteurs de confusion non mesurés, en particulier provenant des données d'imagerie. De plus, la nature de la collecte des données utilisées a pu également conduire à sous-estimer le potentiel bénéfice du traitement. Nous pensons néanmoins que ce score de propension dépendant du temps pourrait être un outil utile, méritant d'être pris en compte dans les études d'efficacité du traitement dans le contexte d'une maladie chronique, en particulier dans le cadre d'une collecte de données régulière et moins dépendante du suivi.

2. Comparaison d'efficacité de traitements en vie réelle

Le deuxième axe, abordé dans le chapitre 2, concernait la comparaison de l'efficacité de deux traitements de deuxième ligne, le fingolimod et le natalizumab. Ce travail a été réalisé dans le cadre d'une mobilité internationale, à partir de 3 cohortes observationnelles (MSBase, Registre danois, OFSEP).

Les critères d'inclusion, les populations et les méthodes statistiques ont été identifiés comme principales différences dans les études originales qui ont conduit à des résultats

discordants. Ces trois éléments ont été explorés afin de mieux comprendre leur impact sur les résultats.

Parmi les résultats de ce travail, nous avons montré que malgré l'uniformisation des critères d'inclusion, l'analyse des cohortes séparées a confirmé les résultats discordants précédemment obtenus. En parallèle, l'analyse de la cohorte poolée a montré que le natalizumab présentait une efficacité supérieure au fingolimod dans la réduction du risque de poussée et l'amélioration du handicap. Par contre, aucune différence dans le risque d'aggravation du handicap n'a pu être montrée entre les deux traitements. Ce premier travail a indiqué que les différences d'efficacité pouvaient donc dépendre des caractéristiques des cohortes de patients étudiées, ce qui a conduit à mener des analyses stratifiées selon les caractéristiques démographiques et cliniques des patients.

L'impact des méthodes statistiques a également été exploré, discutant les choix méthodologiques faits dans les études originales que ce soit pour le contrôle du biais d'indication ou le type d'analyse longitudinale. Les analyses comparatives ont finalement conduit à des résultats proches. Cependant, dans un tel contexte (traitements dont la disponibilité n'est pas contrainte), l'étude de la population entière est adaptée. Nous concluons donc avec le résultat obtenu par la pondération sIPTW (mesurant l'effet ATE) et montrant que le natalizumab apparait supérieur au fingolimod dans la réduction des poussées. En intention de traiter, l'étude n'a pas permis de montrer de différence sur la progression ou l'amélioration du handicap.

3. Evaluation des stratégies thérapeutiques

Enfin, le dernier axe de cette thèse concernait l'évaluation des stratégies thérapeutiques. Cette partie portait sur la description des usages des stratégies thérapeutiques, puis l'effet de l'initiation précoce d'un traitement sur la progression du handicap et enfin la comparaison des stratégies thérapeutiques sur la progression du handicap. Ce travail, réalisé à partir des données de l'OFSEP, n'a pas encore été complètement exploré.

Le principal résultat de l'étude descriptive a indiqué que le premier traitement était initié de plus en plus tôt sur la période s'étendant de 1996 à 2013. De plus, l'ensemble de l'arsenal thérapeutique disponible sur la période d'étude était utilisé dès la disponibilité d'un nouveau traitement.

L'étude de l'efficacité de l'initiation précoce du traitement a conduit à montrer son bénéfice par rapport au fait d'être traité plus tard sur le délai d'atteinte d'un niveau de handicap modéré. Ces analyses seront complétées après la soumission du premier article descriptif. Nous discuterons notamment la définition d'un traitement précoce selon d'autres critères comme l'âge ou le handicap.

L'étude de l'efficacité du traitement d'induction par rapport à un traitement administré dans une stratégie d'escalade ou intensive n'a pas pu être conduite. A ce jour, un protocole d'étude a été proposé et sera mis en œuvre plus tard.

B. La matière première : les connaissances, les données et les méthodes

Comme dans tout travail de recherche, l'origine de ce travail était la revue de la littérature. En particulier, dans les chapitres 3 et 4, les études précédemment conduites fournissaient des résultats divergents quant à l'efficacité des traitements de première et deuxième lignes. Ces variations et donc ces conclusions contradictoires empêchent aujourd'hui de fournir un message clair aux neurologues et aux patients sur l'efficacité des traitements en vie réelle. Cela crée même une forme de doute sur la production de connaissances à partir des données observationnelles. En effet, les caractéristiques des patients étudiés et les choix méthodologiques impactent les résultats, fournissant plusieurs réponses. Dans un contexte de vie réelle, la conclusion ne semble possible que dans l'établissement d'un faisceau de preuves fait de multiple réponses. En revanche, dans le cas présent il apparait qu'il manquait certaines clés de lecture pour mieux comprendre ces différences. En effet, chaque étude apporte un nouveau résultat mais il n'est pas toujours évident de comprendre en quoi les choix effectués l'ont fait varier des autres, que ce soit en termes d'ordre de grandeur ou de direction. Etant donnée la complexité pour un méthodologiste, on peut imaginer que pour des professionnels de santé, de telles variations troublent la compréhension.

1. Les données observationnelles : une richesse non exhaustive

Les études observationnelles permettent d'étudier les conditions réelles d'utilisation d'un médicament. Dans les études de comparaison d'efficacité, nous cherchons à démontrer l'effet causal que le traitement a eu sur le devenir des patients. La difficulté réside dans l'estimation de cet effet causal puisque pour qu'il le soit, il faudrait avoir mesuré l'ensemble des facteurs de confusion. De tels manques occasionnent l'estimation d'une association biaisée entre le fait d'avoir été traité et le devenir, modifiant l'effet « réel » en masquant ou en créant une association. Pour réaliser ce type d'étude, nous faisons appel à des bases de données dites observationnelles. Ces données sont la plupart du temps collectées indépendamment de la question de recherche et de fait non spécifiques à l'étude. Afin de limiter le manque de données importantes, on tend à recueillir, pertinemment, des données variées et ce, durant plusieurs années.

1.1 OFSEP : des données longitudinales, rapportant l'état clinique du patient

Dans le cas spécifique de l'OFSEP, les données sont collectées au cours du suivi médical du patient, principalement en centre expert. Cette collecte particulière de données confère à l'OFSEP une certaine richesse dans l'information. On dispose ainsi d'informations individuelles et cliniques riches, qui sont pertinentes à prendre en compte dans des analyses. Cette richesse réside à la fois dans la nature de la donnée et dans son caractère longitudinal.

La nature des données permet d'obtenir des données spécifiques reflétant l'état clinique du patient (poussées, handicap, IRM), bénéficiant en plus de l'expertise du neurologue dans le recueil. Ces informations sont importantes pour étudier l'efficacité d'un traitement puisqu'elles sont directement liées à la décision thérapeutique.

L'aspect longitudinal permet de mesurer l'état clinique au cours du temps et de fait d'observer l'évolution de la maladie. La mesure des données au cours du temps est primordiale dans un contexte de maladies chroniques, présentant des évolutions à court, moyen et long termes. En particulier, l'évolution de la SEP n'est pas linéaire et peut même être très variable d'un patient à l'autre (Leray *et al.*, 2010). Si l'âge au début de la SEP a souvent été démontré comme un facteur pronostique important, l'évolution de la maladie peut être plus ou moins lente, les poussées plus ou moins sévères laissant des séquelles importantes ou non et les lésions

plus ou moins nombreuses (Myhr *et al.*, 2001; Trojano *et al.*, 2002; Tremlett *et al.*, 2006; Vukusic and Confavreux, 2007; Scalfari *et al.*, 2011).

En revanche, comme nous l'avons vu dans ce travail, ce recueil présente des limites. L'un des premiers points à discuter concerne l'origine de la collecte.

1.2 OFSEP : des données collectées essentiellement en centre expert

La collecte de données, effectuée en centres experts, occasionne probablement la sélection d'une population de patients spécifique (Vukusic *et al.*, 2018). En effet, au regard de la prise en charge de la SEP en France, les patients suivis en centre expert présentent des profils de maladie certainement plus sévères ou complexes, nécessitant l'intervention d'un neurologue expert (Kerbrat *et al.*, 2015). Des différences de profils entre les patients suivis en centre expert ou non ont d'ailleurs été observées au Canada (Mckay *et al.*, 2016) et en région Lorraine (Debouverie *et al.*, 2009). En France, il n'existe finalement que le registre ReLSEP en Lorraine qui inclut tous les patients de la zone géographique, donc des patients pris en charge en centre expert et/ou en ville, regroupant de fait des patients aux profils de maladie plus variés [23]. A l'échelle Européenne, nous avons évoqué le registre danois dans ce travail de thèse. Ces données ont été utilisées pour l'étude comparative du natalizumab et du fingolimod. Par définition, cette cohorte présente la caractéristique d'être exhaustive de la population puisque l'inscription au registre est obligatoire au Danemark. Elle fournit donc des résultats généralisables aux patients atteints de SEP au Danemark.

La prise en charge en centres experts peut également contribuer à des différences de traitement. En effet, bien qu'en France, tout neurologue peut prescrire un traitement spécifique SEP, il n'est pas exclu que l'usage des traitements soit différent que l'on soit expert de la pathologie (« sepologue ») ou non. On peut imaginer qu'un neurologue expert de la SEP soit plus enclin à traiter très tôt et éventuellement soit plus éclairé sur le rapport bénéfice/risque, fournissant un accès à des thérapeutiques plus agressives. Le résultat obtenu dans l'étude descriptive du chapitre 5 sur l'initiation précoce peut d'ailleurs être mis au regard des critères d'inclusion. En effet, non seulement, les patients sont pris en charge en centres experts mais l'un des critères d'inclusion porte sur la régularité du suivi au centre. En se concentrant

[23] *https://www.reseau-chu.org/article/relsep-unique-en-france-le-registre-nanceen-sur-la-sclerose-en-plaques/*

uniquement sur des patients suivis au moins 5 ans depuis le début de la SEP avec au moins trois mesures d'EDSS disponibles, nous décrivons une prise en charge assez spécifique aux centres experts. On peut s'interroger sur le maintien d'un tel critère, en tout cas sur ce qu'il représente en terme de validité externe.

Dans l'étude portant sur le bénéfice des traitements de première ligne injectables présentée dans le chapitre 3, nous avons pris en compte une variable reflétant le délai depuis la première prise en charge en centre expert. En outre, si cette première consultation en centre expert donne l'occasion au neurologue de se positionner sur la prise en charge de la maladie, c'est également l'occasion de revenir sur l'historique médical du patient. En particulier, dans son discours le patient peut évoquer des symptômes passés, qui n'ont à l'époque pas été attribués à la SEP ou n'ont même pas fait l'objet d'une consultation chez un médecin généraliste ou neurologue. Ces symptômes reconnus comme les premiers symptômes de SEP marquent le début de la maladie telle que définie par le neurologue et saisie dans l'OFSEP. Dans une étude publiée en 2011 portant sur des patients suivis en centre expert en France, il a été montré qu'un tiers des patients étaient dans ce cas (Gout *et al.*, 2011). En moyenne, le délai entre les premiers symptômes non reconnus et les nouveaux symptômes poussant à consulter était d'environ 4 ans. On peut également faire référence à une étude d'incidence menée en Bretagne sur la période 2000-2001 pour laquelle la période d'identification a dû être étendue jusqu'en 2010. Cette extension assurait de collecter les cas incidents de SEP ayant consulté tardivement (Kerbrat *et al.*, 2015; Yaouanq *et al.*, 2015). Cette information apparait importante à préciser, en particulier dans l'étude présentée dans le chapitre 3. En effet, la date d'entrée de la modélisation du score de propension dépendant du temps étant le début de la SEP, nous ne pouvons mesurer l'impossibilité d'être traité pour une maladie encore non identifiée. Il est difficile d'établir l'impact que cela a pu avoir sur le résultat obtenu. On peut tout de même évoquer que les patients dans ce cas sont probablement des patients moins actifs, un long délai entre les deux premiers évènements étant un facteur de bon pronostic. Dans tous les cas, la première évaluation de données a été réalisée à la première visite pour laquelle des données étaient disponibles, la maladie était alors bien identifiée. Le moment où les patients ont une chance d'être traités correspond finalement au moment de la consultation avec un neurologue qu'il exerce en centre expert ou non. Cette date aurait été plus pertinente pour la modélisation du score de propension. En effet, durant une période entre le début de la SEP et la première visite au centre expert, le patient peut n'avoir aucune chance d'être traité puisqu'il se peut qu'il n'ait pas consulté de neurologue. Toutefois, la première visite avec un neurologue n'est pas disponible dans les données de l'OFSEP.

Ce format de la collecte a également eu un autre impact sur l'évaluation des données au cours du temps. En effet, l'intervalle de temps d'évaluation a été déterminé à partir de la disponibilité d'un EDSS. Ces derniers sont essentiellement mesurés au centre expert et très peu en ville ; lesquels sont de toutes façons rarement enregistrés dans la base. Entre le début de la SEP et la première visite avec mesure d'EDSS, aucune valeur n'est donc disponible. Pendant l'intervalle de temps précédant la mise à jour des données, le score de propension est celui calculé à partir des données mesurées à la première visite disponible. Pourtant, c'est une évolution de facteur qui a conduit à cette valeur de score de propension. Nous avons schématisé cette limite sur la Figure 25 : Evaluation des scores de propension dépendant du temps au cours du suivi. Ceci apporte un autre argument pour considérer que l'effet des traitements de première ligne a été sous-estimé en appariant un patient traité au moment de son traitement avec un patient du groupe « pas encore traité » qui, au même moment, présentait un score de propension similaire mais déterminé à partir de données mesurées après le moment de son appariement. De ce fait, il se peut qu'il y ait un risque de sous-estimation de la propension à être traité pour le patient traité et de surestimation pour le patient qui n'est pas encore traité. Cela est d'autant plus accentué que les initiations de traitements peuvent être collectées durant l'interrogatoire du patient mais les EDSS ne sont eux réalisés qu'au moment de la consultation. Ceci aurait pu être limité par une entrée dans l'étude au premier examen, mais qui ne correspondrait pas au même moment pour tous les patients, ou des mises à jour de données plus régulières. De plus, la première visite en centre expert n'est pas toujours bien renseignée dans l'OFSEP ou de façon non uniforme dans les centres. En effet, la date de première visite fait parfois référence à la date de première consultation au centre expert ou parfois à la date de première consultation neurologique quel que soit le lieu d'exercice.

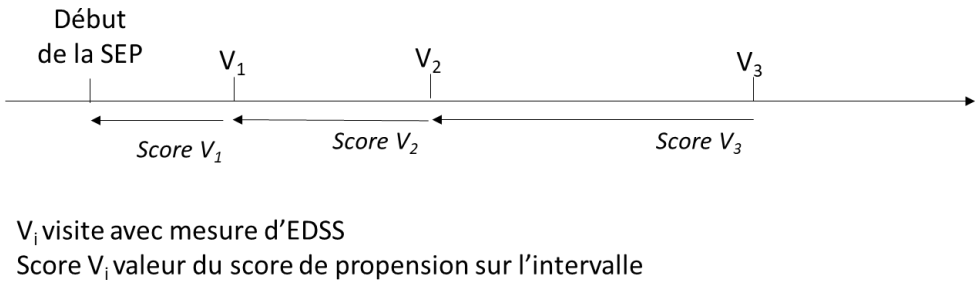


Figure 25 : Evaluation des scores de propension dépendant du temps au cours du suivi

1.3 OFSEP : des données non exhaustives

Finalement, de façon non spécifique à l'OFSEP, les données observationnelles sont souvent marquées par l'existence de données manquantes ou encore de variables non mesurées. Dans l'OFSEP et dans ce travail, cela se traduit par deux limites importantes.

En premier lieu, il s'agit de l'absence de certaines données qui semblent pourtant primordiales dans les analyses. En effet, les études proposées dans ce travail souffrent de l'absence des données d'imagerie. Dans le chapitre 3, il apparait évident que le résultat est impacté par une confusion résiduelle ayant influencé le résultat en défaveur du traitement. Si la démonstration de l'effet bénéfique du traitement sur l'accumulation du handicap n'est pas claire en vie réelle, ce résultat montrant un effet négatif laisse à supposer un biais résultant d'un problème d'équilibre. En effet, malgré l'appariement, les SEP des patients traités paraissent être plus actives que celles des patients du groupe « pas encore traités ». Dans l'étude présentée dans le chapitre 4, les données d'imagerie sont également manquantes. Il faut noter que dans l'étude originale française, les données étaient disponibles mais celles-ci n'ont pu être étudiées car partielles ou manquantes dans les autres cohortes (MSBase ou le registre danois respectivement). Dans l'étude originale française, les données d'imagerie avaient pu être utilisées grâce à un travail de révision de l'ensemble des IRM des patients inclus dans le but de compléter les données manquantes. Dans les données observationnelles, il y a parfois un écart entre ce dont on dispose et ce dont on a besoin pour répondre au mieux à la question de recherche. Cela nous amène à fournir des informations additionnelles, compléter les données, voire réviser nos protocoles. Avec le temps, l'accès à l'IRM est beaucoup plus systématique donc cette limite tend à diminuer. Sur des données plus récentes, ce n'est plus le manque de données qui pose problème mais leur hétérogénéité.

En second lieu, une des limites des données observationnelles, particulièrement inhérente aux données issues du suivi médical est la dépendance à la fréquence des consultations. En effet, l'objectif est de collecter dans les centres, les informations de tous les patients sans impacter le suivi ou la prise en charge. Les patients ne sont pas inclus dans l'OFSEP pour une étude spécifique, nécessitant un suivi particulier. Au-delà du fait que cela contribue probablement à l'hétérogénéité entre les centres, deux autres caractéristiques peuvent être soulevées. Il est tout d'abord probable que la fréquence de consultations soit liée à l'état du patient et donc à un potentiel besoin de prise en charge spécifique au moment de la consultation. Au biais de sélection dû à la collecte de données, s'ajoute donc dans certains

profils de prise en charge la dépendance à l'activité de la maladie. Ensuite, du fait de la collecte de données en centre expert, un manque d'exhaustivité dans le suivi peut être également observé en raison de consultation ponctuelle pour avis. Pour pallier à cela, il peut être envisagé de travailler dans un sous-groupe de patients pour lesquels des consultations régulières au centre expert ont été réalisées. Ce choix a été fait récemment dans un travail mené à Rennes portant sur la survenue de poussée en phase secondairement progressive (Ahrweiller *et al.*, 2019). Ce type de sélection accentue d'autant plus le biais de recrutement et la question de la généralisation des résultats mais ce biais, étant de toutes façons présent, ne devrions-nous pas privilégier la qualité des données ? Malgré tout, cette sélection pose la question d'un conditionnement sur le futur. En effet, dans l'étude d'une décision thérapeutique, le groupe étudié devrait être analysé sans modifier ce qui se déroule après cette décision puisque ce qui se passe ensuite est liée à cette décision thérapeutique, en particulier le suivi. Au-delà d'une telle sélection, le travail sur une population restreinte, comme une population monocentrique par exemple, offre tout de même la possibilité de compléter des informations manquantes nécessaires en accédant à l'information brute ou parfois, en recontactant le médecin traitant. Cela maximise les chances de travailler à partir de données complètes. Ceci est techniquement applicable à l'ensemble des données de l'OFSEP mais il faut évidemment le mettre au regard des moyens humains et financiers pour réaliser un tel travail. Ainsi, cela nous amène à considérer la quantité par rapport à la qualité, la représentativité et la généralisation de nos résultats dans un contexte où nous objectivons de mesurer un effet en vie réelle. Ce type de pratique conduit parfois à une recherche à deux niveaux, l'une sur des cohortes de grandes taille posant des questionnements méthodologiques complexes et nécessitant d'avoir conscience des limites inhérentes, l'autre qui fournit des résultats à partir de données précises mais difficilement généralisables car issu d'un groupe restreint.

Les données observationnelles sont nécessaires dans l'évaluation de traitement en vie réelle. Elles sont riches et souvent sans égal. En revanche, leur utilisation présente des limites qu'il est important de documenter. Malgré l'utilisation de méthodes statistiques complexes et adaptées, celles-ci pourront difficilement répondre au plus juste en présence de données clés non mesurées voire largement manquantes. C'est pourquoi l'absence de prise en compte de telles données se doit d'être discutée dans l'interprétation du résultat. Notons tout de même que nous évoquons ici certaines limites de l'OFSEP inhérente à une cohorte dont la collecte a débuté depuis longtemps. Plusieurs processus qualité sont mis en place et ce, depuis quelques années. Depuis le 15 juin 2013, un ensemble précis de données (les données de la fiche

minimale) est collecté de manière systématique : données personnelles et sociodémographiques, données cliniques (épisodes neurologiques, évaluations cliniques, handicap irréversible), IRM et de traitement de fond, afin d'harmoniser le recueil et d'améliorer la qualité des données.

2. L'utilisation des méthodes

Parmi les éléments à l'origine de travaux de recherche, il est essentiel d'aborder la conception méthodologique de l'étude. La définition d'un protocole d'étude épidémiologique se réalise par étape dont l'une des dernières seulement concerne le plan d'analyse statistique. Les critères d'inclusion, le choix du schéma d'étude et la définition des variables constituent des étapes qui sont primordiales au bon déroulement d'étude. On peut également envisager d'établir un graphique de type DAG pour formaliser nos hypothèses. Ma connaissance a évolué au cours de la thèse et ce n'est que tardivement que nous avons réfléchi à la formalisation d'un DAG. Il n'a donc pas été utilisé dans la réalisation de ces travaux mais il nous semble pertinent de présenter nos réflexions en discussion qui gagnerait à être confrontées au point de vue des neurologues.

2.1 DAG de la relation entre les traitements de fond sur la progression du handicap dans la SEP rémittente

Dans cette partie, nous faisons le choix de présenter un seul DAG car les questions de recherche posées dans les différents axes de cette thèse se rejoignent et nous semblent pouvoir être analysées à partir d'une même représentation. On pourrait l'intituler « DAG de la relation entre le traitement de fond et la progression du handicap irréversible dans la SEP rémittente ». Il est présenté en Figure 26 : DAG de la relation du traitement de fond sur la progression du handicap irréversible dans la SEP rémittente. Pour rappel des éléments évoqués dans la partie méthode sur les DAG (page 43), la variable « poussées », par exemple est dite cause commune du traitement et de la progression du handicap. Dans le but de mesurer l'effet direct du traitement sur la progression du handicap, on cherche à contrôler toutes les causes communes du traitement et de la progression du handicap. Parmi les potentielles variables causes communes du traitement de fond et de la progression du handicap irréversible, nous avons à ce jour intégré des variables démographiques et des variables relatives à l'activité de la maladie.

Concernant les caractéristiques démographiques, nous avons intégré le genre et l'âge. En ce qui concerne le genre, certains traitements présentant un risque tératogène (Houtchens and Kolb, 2013), la prise en charge peut s'avérer différenciée selon le sexe. D'ailleurs, la grossesse peut modifier le cours de la maladie (Voskuhl and Momtazee, 2017). De plus, dans la population masculine, il a été observé une évolution plus rapide du handicap (Debouverie *et al.*, 2008; Leray *et al.*, 2010; Kalincik, 2015) même si cela n'est pas toujours retrouvé (Runmarker and Andersen, 1993; Langer-Gould *et al.*, 2006). Concernant l'âge, il a été démontré que le handicap avait tendance à progresser plus vite lorsque la maladie débute plus tard (Leray *et al.*, 2010). De plus, il est pris en compte dans la stratégie thérapeutique, en particulier lorsqu'il est associé à une maladie sévère (Edan and Le Page, 2013). L'âge a un impact très important sur l'évolution de la SEP (Confavreux and Vukusic, 2006).

Concernant les variables relatives à l'activité de la SEP, la charge lésionnelle a été retenue comme cause commune. Comme nous l'avons déjà évoqué, l'IRM est un facteur pronostic important dans la SEP et est utilisée dans la décision thérapeutique. Enfin, les poussées et le handicap ponctuel ont été intégrés. Le terme « poussée » fait référence à la fois à la fréquence, au phénotype de la poussée et sa sévérité. Identifiés plus ou moins clairement comme facteurs pronostiques, ils impactent à la fois la progression de la maladie et la décision thérapeutique (Leone *et al.*, 2008; Leray *et al.*, 2010).

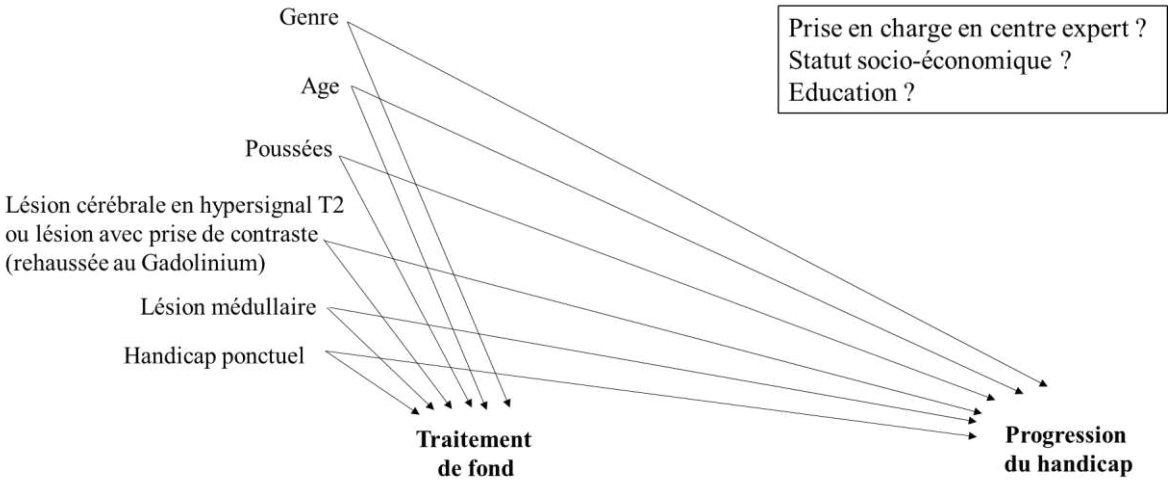


Figure 26 : DAG de la relation du traitement de fond sur la progression du handicap irréversible dans la SEP rémittente

Sur la figure, nous avons choisi de faire apparaitre sous forme de question des éléments dont la relation avec le traitement ou la progression du handicap semble moins évidente. Le statut socio-économique ou l'éducation n'a jamais été clairement démontré comme cause de l'un et/ou de l'autre (Goulden *et al.*, 2015). Une récente étude a tout de même démontré que les patients de statuts socio-économiques plus élevés (mesuré au niveau du lieu de résidence) avaient un accès facilité aux traitements de deuxième ligne, mais l'absence de certains facteurs d'ajustement amène à considérer ce résultat avec précaution (Calocer *et al.*, 2018). En revanche, on ne peut pas exclure qu'à l'instar de plusieurs autres pathologies cela peut être associé à la prise en charge, au travers par exemple de la relation médecin-patient (Fassin *et al.*, 2000; Kaminski *et al.*, 2008).

Un autre élément concerne la prise en charge en centre expert. Comme nous l'avons évoqué, le fait de consulter un neurologue en centre expert peut-être le témoin d'une certaine sévérité de SEP. De plus, l'usage des traitements peuvent être différencié en centre expert. Si nous devions placer cette variable sur le DAG, elle pourrait prendre deux places. La première serait une flèche entre une variable « prise en charge au centre » et la variable « traitement de fond » indiquant une association avec le traitement mais pas avec la progression du handicap. La seconde place serait sur le chemin de chaque variable relative à l'activité vers le traitement. Dans le premier cas, la prendre en compte pourrait conduire à un ajustement non nécessaire puisqu'il ne nous semble pas que la prise en charge en centre expert soit cause commune du traitement et du handicap. Dans le second cas, deux scénarios nous apparaissent possibles. Dans le scénario A, un potentiel sur ajustement peut être évoqué puisque nous ajusterions alors à la fois sur une variable reflétant l'activité de la maladie et la prise en charge en centre expert qui se trouverait sur le même chemin. Dans le scénario B au contraire, on pourrait voir cette variable de prise en charge comme un « proxy » d'un certain nombre de variables relatives à l'activité de la maladie que nous ne mesurons pas comme la sévérité de la poussée ou la charge lésionnelle. Dans ce cas, en tenir compte pourrait nous permettre de rectifier la confusion résiduelle liée à certaines variables non mesurées. Dans le travail étudiant l'efficacité des traitements de première ligne, nous avons intégré à notre analyse le délai depuis la première consultation en centre expert. L'hypothèse est qu'une consultation récente, ancienne ou inexistante peut avoir impacté la prise en charge ou être le témoin d'une maladie bien contrôlée. Après la formalisation de ce DAG, nous nous interrogeons sur la nécessité de prendre en compte une telle variable. Notons que l'analyse a aussi été réalisée sans tenir compte de cette variable et avait conduit à des résultats similaires.

2.2 Les critères d'inclusion

L'étape relative à la formulation des critères d'inclusion permet de définir la population à laquelle on pourra généraliser nos résultats. Ces critères se doivent d'être précis et de respecter des aspects fondamentaux associés à la comparaison de deux groupes. En particulier, au cours de cette thèse, l'hypothèse de positivité a parfois été questionnée. Pour rappel, l'hypothèse de positivité est l'hypothèse selon laquelle tout patient inclus dans l'étude comparative doit avoir une probabilité non nulle d'appartenir à chaque groupe.

Dans la littérature et certaines discussions, j'ai parfois eu le ressenti que l'ajustement ou les scores de propension étaient vus comme la solution à la comparabilité. Si ces méthodes permettent évidemment de tenir compte de la confusion, il convient de ne pas négliger l'étape de définition de la population. Cette recherche de comparabilité des groupes débute dès cette étape où nous cherchons à inclure des patients pour lesquels la comparaison est sensée et qui par construction de l'échantillon présentaient des chances similaires de recevoir le traitement.

En particulier, dans le chapitre 3, nous avons été amenés à ajouter une contrainte à la méthode d'appariement par score de propension dépendant du temps afin de respecter cette hypothèse de positivité qui a grandement impacté le résultat (page 74). Sur ce point, nous aurions pu également envisager d'inclure uniquement des patients incidents à partir de la disponibilité des traitements. Cependant, le questionnement initial visait à documenter l'effet des traitements de première ligne injectables chez tous les patients éligibles au traitement, incluant donc également les patients atteints de SEP depuis longtemps mais n'ayant pas été traités en raison de l'indisponibilité des traitements.

Dans le chapitre 4, les critères d'inclusion auraient probablement aussi mérités d'être mieux définis. Les pays participants à la cohorte MSBase ne disposent pas tous du natalizumab et du fingolimod. Ce point a été évoqué pendant les réunions sur place. Conscient de la violation de l'hypothèse de positivité, il semblait difficile d'envisager d'exclure certains pays. Des considérations plus « stratégiques » venaient ainsi se glisser dans nos réflexions méthodologiques.

Une fois la population définie, il apparait primordial de la décrire. Les caractéristiques des patients fournissent des informations importantes dans l'interprétation des résultats. Il est donc nécessaire de regarder le descriptif de base avant de réaliser des analyses plus approfondies. Par exemple, dans l'étude réalisée lors de la mobilité et présentée dans le chapitre

4, malgré des critères d'inclusion unifiés et une méthodologie similaire, les résultats restaient discordants laissant à penser que les différences sont plutôt le reflet d'une certaine hétérogénéité dans les pratiques dans les centres participants. La description des groupes pouvait nous interpeller constatant que les patients issus du registre danois traités par natalizumab et fingolimod initiaient le traitement plus tard et qu'ils présentaient moins de poussées que dans les autres cohortes. Dans le chapitre 5, nous proposons un article purement descriptif des stratégies thérapeutiques en France.

2.3 Utilisation d'un score de propension

Après la réalisation de la définition des critères d'inclusion et la description de l'échantillon, nous avons vu dans cette thèse que l'on peut avoir recours à des scores de propension pour équilibrer les deux groupes que l'on souhaite comparer.

Dans un premier temps, on peut s'interroger sur l'essor des scores de propension ces dernières années. Historiquement, les méthodes d'ajustement par régression étaient plus souvent utilisées que les scores de propension pour estimer l'effet d'un traitement à partir de données observationnelles. Comme évoqué dans le chapitre 2, l'utilisation d'un score de propension permet d'estimer l'effet marginal du traitement alors qu'un modèle de régression fournit un effet conditionnel. Dans le cas où l'on cherche à répondre à une question similaire à celle d'un essai randomisé, l'effet marginal est plus pertinent car il fournira un effet moyen du traitement sur la population (Austin, 2011*a*). Ce premier point justifie alors l'utilisation des scores de propension dans les études d'efficacité. Un ajustement par régression sera préféré dans la comparaison d'une caractéristique non modifiable comme l'étude de l'effet du sexe sur la survenu d'un évènement par exemple. De plus, des considérations plus pratiques peuvent aussi être évoquées (Austin, 2011*a*). L'évaluation de la comparabilité (chevauchement) est réalisée lorsqu'on utilise un score de propension mais plus difficile avec une approche par régression. De plus, lorsque que le nombre d'évènement réalisé est faible, l'approche par régression conduira à limiter le nombre de variables à introduire dans le modèles

Différentes méthodes utilisant des scores de propension existent et permettent d'estimer des effets différents. Cette différenciation entre les deux effets est clairement identifiée dans la littérature théorique mais l'est moins en pratique. Les résultats issus d'analyses permettant

d'estimer les effets ATE (relatif à la population entière) ou ATT (relatif à la population traitée) sont souvent comparés sans discuter le fait que les deux effets se réfèrent à des populations différentes. De plus, le choix de l'auteur n'apparait pas toujours clairement.

Deux méthodes utilisant les scores de propension ont été considérées dans ce travail : la pondération et l'appariement. L'appariement est largement utilisé dans la littérature, probablement en raison de sa ressemblance à un essai randomisé et à une formalisation plus accessible. La pondération présente de bonnes performances et limitent les exclusions de patients (Le Borgne, 2016). En vie réelle, nous avons tendance à préférer la pondération qui permet de ne pas exclure de patients et de fournir des résultats généralisables à la population cible. En effet, à cause des exclusions, l'échantillon apparié peut ne plus correspondre à la population cible initiale. L'étude du score de propension dépendant du temps nous a contraint à utiliser un appariement. Il s'agit d'une première étape vers une méthode que nous évoquerons dans la partie suivante. Il nous semblait important de préciser dans les résultats que malgré une exclusion importante de patients, la base de données étudiée était proche de la cohorte d'origine.

Dans tous les cas, bien qu'accessibles, les scores de propension répondent à une méthodologie statistique bien définie et donc à de bonnes pratiques d'utilisation. Comme pour d'autres méthodes sophistiquées accessibles par tous notamment via les logiciels, les scores de propension ne sont pas toujours utilisés ou présentés de manière rigoureuse. En particulier, on peut citer l'évaluation des hypothèses, souvent peu évoquée dans les articles utilisant les scores de propension, la manière dont les variables incluses dans le score ont été choisies ou encore la prise en compte du caractère aléatoire de l'appariement qui est rarement réalisée. Ce constat n'est pas spécifique à la SEP et est retrouvé dans d'autres disciplines (Nayan *et al.*, 2017). En pratique, comme nous l'avons vu, l'évaluation des hypothèses repose essentiellement sur la vérification du support commun graphiquement et de l'équilibre obtenu. L'évaluation de l'équilibre peut être établi à partir de diverses quantités : les différences standardisées, la distance de Cohen ou la distance de Mahalanobis qui tiendra compte de la covariance pour une variable catégorielle. En revanche, les tests de comparaison ne devraient pas être utilisés (Austin, 2007*a*). En effet, après appariement ou pondération, le fait d'avoir équilibré les groupes devient une propriété de l'échantillon. Cette propriété doit être prise en compte dans la mesure de l'évaluation ce qui n'est pas le cas dans un test de comparaison. De plus, le processus d'équilibre a impacté la taille de l'échantillon, or les tests de comparaison sont aussi influencés par la taille de l'échantillon. L'équilibre pourrait donc sembler meilleur qu'il ne l'est.

On peut s'interroger sur la manière de gérer un problème d'équilibre ou un manque de support commun. La littérature sur les scores de propension suggère de modifier le score en intégrant des interactions par exemple. Plus globalement, cette question se réfère à un objectif de réduction de la confusion. Dans nos études, nous avons été sujets à la confusion résiduelle en raison de données manquantes. Pour autant, nous avons cherché à réduire le plus possible cette confusion. En particulier, les critères d'inclusion étaient assez stricts dans l'ensemble des études. De plus, dans la comparaison présentée en chapitre 4, le choix de comparateurs actifs a probablement contribué à limiter la confusion (Yoshida *et al.*, 2015). En effet, ce type de comparaison permet d'augmenter le chevauchement et de réduire les biais liés aux potentiels facteurs de confusion non mesurés. Ainsi dans cette étude, nous évoquons comme limite la disponibilité des données IRM. Bien évidemment, cette donnée est primordiale dans la décision thérapeutique mais nous pouvons penser que nous avons réduit ce biais en comparant les patients traités par natalizumab à des comparateurs actifs, c'est-à-dire traités par fingolimod. Pour limiter les problèmes de confusion, un schéma d'étude où le patient est son propre témoin peut également être envisagé mais ce type de schéma n'est probablement pas adapté à l'évaluation de critère de jugement à long terme.

Sur certains points et notamment dans le choix des variables à intégrer dans le score de propension, la littérature est moins consensuelle. Ce manque de consensus a conduit à proposer dans ce travail des choix méthodologiques différents selon les études. En particulier, dans les deux études conduites en interne présentées dans le chapitre 3 et le chapitre 5, les variables liées au critère de jugement ont été utilisées pour modéliser le score de propension comme il semble être plutôt recommandé (Austin, 2011*c*). En revanche, dans l'étude collaborative (chapitre 4), toutes les variables mesurées ont été intégrées au score de propension. Cette pratique ne semble pas à écarter formellement, mais intégrer des variables qui n'avaient pas à être prises en compte risque d'impacter l'estimation du modèle du score de propension, notamment en augmentant la variance. De plus, on prend le risque d'intégrer des variables instrumentales (liées uniquement au traitement).

Concernant la sélection des variables liées au critère de jugement, celle-ci a été effectuée en étudiant l'association statistique entre les variables et le critère de jugement. Un test a donc été réalisé. Une telle pratique pourrait conduire à exclure à tort des variables, en raison d'un manque de puissance pour détecter l'association (Austin, 2011*a*). Ainsi, un déséquilibre peut être engendré sur des variables exclues à tort du score de propension.

Le dernier point que nous souhaitons évoquer concerne le caractère aléatoire de l'appariement qui semble être rarement pris en compte dans les études publiées. En effet, la méthode d'appariement recommandée dans la littérature et que nous avons utilisé dans ce travail est dite aléatoire. En ce sens, le premier patient traité auquel on cherchera un contrôle est choisi aléatoirement. Il existe d'autres méthodes respectant un ordre. On peut par exemple commencer à apparier par le patient présentant le plus haut score ou le plus faible. Ce caractère aléatoire peut modifier la base appariée. C'est pourquoi dans l'étude proposée dans le chapitre 3, nous avons utilisé le bootstrap afin de répéter l'analyse pour tenir compte de cette potentielle variabilité. Un graphique illustratif de cette variabilité, issu d'une étude non présentée dans la thèse, est proposé en annexe G. L'utilisation du bootstrap a également été réalisée dans le but de mieux estimer la variance (Austin, 2016).

Ces travaux conduits avec d'autres équipes m'ont montré qu'en l'absence de consensus, les habitudes de pratiques et une certaine conviction priment sur le contexte théorique. De plus, ces échanges m'ont aussi conduit à m'interroger sur l'accessibilité de la connaissance théorique. En effet, j'ai pu constater que la méthodologie des scores de propension, bien qu'utilisée, était parfois mal connue. Evidemment, ce constat est à mettre au regard d'une littérature qui s'est développée ces dernières années, se rendant de plus en plus accessible.

C. Les résultats et conclusions

1. L'exposition, une variable dépendante du temps

Dans cette thèse, nous nous sommes intéressés à l'efficacité des traitements et de leur utilisation en vie réelle. Deux types de comparaison ont été abordés : recevoir un traitement par rapport au fait de ne pas le recevoir ou recevoir un traitement A par rapport au fait de recevoir un traitement B. On pourrait penser que le premier point se rapprocherait du deuxième mais la gestion de ces deux études comparatives n'est finalement pas si proche (Yoshida *et al.*, 2015). En effet, dans le premier cas traité versus non traité, le groupe comparateur est un groupe non utilisateur et cela pose des questions spécifiques, notamment dans la gestion de la confusion. Les patients issus d'un groupe non utilisateurs du médicament ont des caractéristiques

différentes des patients utilisateurs, d'autant plus dans un contexte de traitement sans contrainte de disponibilité et ayant déjà prouvé son efficacité à court terme. La confusion est en revanche moins difficile à gérer dans le deuxième cas puisque comme nous l'avons déjà évoqué, deux groupes de patients traités par des médicaments aux indications similaires se ressemblent que ce soit en termes de caractéristiques mesurées mais également en termes de caractéristiques non mesurées.

Ainsi dans ce travail, l'évaluation de l'efficacité des premières lignes injectables nous a posé de nombreuses questions. Tout d'abord, nos premiers travaux portaient sur l'étude des interférons-β uniquement. L'acétate de glatiramère présentant une indication similaire aux interférons-β, il nous a semblé plus pertinent de regrouper dans un deuxième temps les deux traitements. D'ailleurs, avec la méthodologie finalement choisie, sans ce regroupement, l'analyse aurait probablement conduit à comparer des patients traités par interféron-β à des patients pas encore traités par interféron-β mais traités par acétate de glatiramère.

Dans nos premières réflexions sur cette comparaison, nous avons cherché à utiliser un schéma d'étude classique conduisant à comparer des patients traités avec des patients non traités. La difficulté d'une telle comparaison résidait dans la définition de l'entrée dans l'étude pour les patients non traités. Il était difficile d'identifier une date correspondant à la date de décision thérapeutique. La date théorique à laquelle le patient aurait pu être traité mais ne l'a pas été. De plus, comme évoqué précédemment, un groupe de patients non traités dans le contexte de prescription des traitements de première ligne injectables correspond à un groupe de patients spécifiques, présentant probablement des formes moins actives de SEP ou de meilleur pronostique. Cela est d'autant plus vrai que les traitements ont été démontrés efficaces dans les essais thérapeutiques et leurs extensions. On leur accorde de fait une plus-value par rapport au fait ne pas être traité. Enfin, un tel point est accentué lorsque l'on cherche à mesurer un effet sur un critère de jugement observé à long terme. Nous avons donc été amenés à réfléchir à une autre manière de comparer les patients.

La méthodologie que nous avons utilisée offre la possibilité de mieux tenir compte du caractère dépendant du temps de l'exposition et nous semble être un apport intéressant dans le champ des maladies chroniques qui nécessitent une prise en charge au long cours. Des méthodes tenant compte de la dépendance du temps, comme les modèles structuraux marginaux, commencent à être plus utilisées. Les modèles structuraux marginaux peuvent être utilisés pour

estimer l'effet causal d'une exposition dépendante du temps en présence de facteurs de confusion dépendant du temps impactés par les expositions précédentes au traitement (Robins *et al.*, 2000). Dans les modèles structuraux marginaux, contrairement à la méthode que nous avons utilisée ici, les pseudo-échantillons sont reconstruits à chaque pas de temps (tous les mois ou tous les 6 mois, etc.) en tenant compte de la mise à jour des variables mais également du potentiel changement d'exposition. Une telle méthode apparait pertinente dans le cas où une variable au temps t, prédit à la fois le critère de jugement et l'exposition à t et que l'exposition à t prédit cette variable au temps $t + 1$. On pourra alors mieux gérer la question des expositions cumulées puisque à chaque pas de temps le statut d'exposition sera mis à jour. Un des exemples en SEP a été proposé par l'équipe canadienne sur l'efficacité des interférons-β sur la progression du handicap (Karim *et al.*, 2014). Une schématisation de la relation causale est proposée en Figure 27 : Relation causale du traitement de la SEP avec trois temps d'évaluation proposée par Karim et al, 2014.

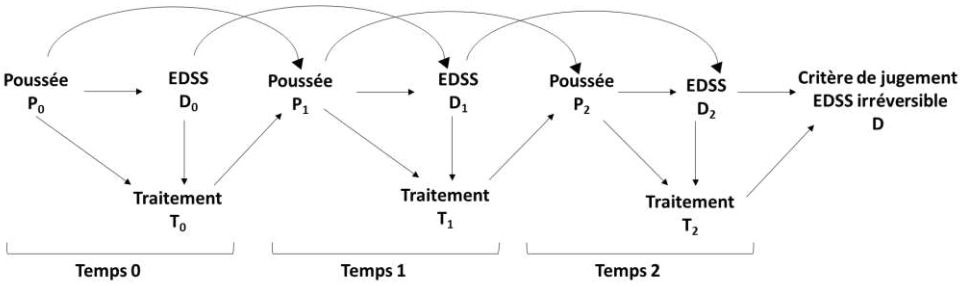


Figure 27 : Relation causale du traitement de la SEP avec trois temps d'évaluation proposée par Karim et al, 2014

Depuis, dans la SEP, la méthode a également été utilisée sur les données issuees du réseau Big MS data , regroupant la cohorte italienne, la cohorte suédoise, le registre danois, MSBase et l'OFSEP (Laffaldano *et al.*, 2019). Ces travaux ne sont pas publiés mais ont été présentés au congrès ECTRIMS 2019. Ces méthodes nécessitent de larges données et une certaine exhaustivité dans les données disponibles.

Dans notre travail, cette première approche avec le score proposé par Lu est un premier pas vers une comparaison au cours du temps. A l'inverse des modèles structuraux marginaux, l'équilibre n'est réalisé qu'à un instant donné, celui de l'initiation du traitement. De plus, une telle méthodologie fournit finalement un résultat portant sur le retard de l'initiation d'un traitement. En effet, le groupe « pas encore traité » n'est autre qu'un groupe dans lequel l'initiation du traitement a été retardée. Nous avons tout de même fait le choix de ne pas formuler nos conclusions ainsi. Parler de traitement précoce dans le contexte de la SEP ne

correspond pas au délai observé dans ce travail. En effet, les patients traités débutaient le traitement après une durée moyenne de SEP de 5 ans, ce qui, au regard des éléments présentés dans le chapitre 5, ne peut être qualifié de précoce.

Cette étude fournit un cadre intéressant pour formaliser les designs d'essais randomisés avec entrée retardée à partir de données observationnelles (Leber, 1997). Dans ce type d'essais, les patients sont randomisés pour être des patients démarrant le traitement d'emblée (patients à démarrage précoce) ou des patients démarrant le traitement plus tard après une période sous placebo (patients à démarrage retardé). Ainsi, à la lecture du chapitre 3 et la formulation des objectifs du chapitre 5, on peut s'interroger sur la non-utilisation du score dépendant du temps. L'utilisation d'une telle méthode nous permettrait effectivement de comparer l'initiation précoce à une initiation plus tardive et la différence de délais pourra être estimée. L'étude proposée en chapitre 5 ayant également pour objectif de mieux documenter ce délai et le risque de progression en fonction de celui–ci, il nous semble important de définir nos groupes en fonction du délai.

2. Les analyses en intention de traiter et en per protocol

Dans ce travail, le choix de l'une ou de l'autre analyse nous a posé de nombreuses questions. Si pour certains, l'analyse en intention de traiter est indiscutablement l'analyse la moins biaisée et devrait être conduite systématiquement, pour d'autres, l'analyse en per protocol apparait plus correcte car plus proche de la réalité et de ce que l'on souhaite mesurer.

Dans un cadre théorique, la réponse semble claire (European Agency for the Evaluation of Medicinal Products, 2000). Dans une analyse de supériorité, on préfère l'analyse en intention de traiter. En effet, l'analyse en intention de traiter à tendance à réduire la mesure de l'efficacité des traitements. De fait, la différence tend à diminuer. Ceci est un atout en analyse de supériorité puisque nous cherchons à éviter le fait d'accepter l'hypothèse de supériorité à tort. En revanche, en analyse de non infériorité, l'analyse en per protocol apparait aussi pertinente puisqu'elle a tendance à maximiser les chances de montrer une différence entre les deux groupes. En pratique, la conclusion d'une étude de non-infériorité est plus sûre si l'analyse en intention de traiter et l'analyse en per protocol donnent des résultats similaires (Sedgwick, 2013).

Pour autant en pratique, cela semble moins clair. Dans ce travail, l'analyse en per protocol a été proposée dans le chapitre 4 comparant le natalizumab et le fingolimod. D'ailleurs,

les résultats des deux premiers articles qui seront publiés en lien avec cette étude collaborative, seront les résultats obtenus avec l'analyse en per protocol. Dans nos discussions avec l'équipe sur place, il semblait difficile d'admettre que l'intention de traiter puisse être un résultat approprié. L'idée d'estimer l'effet de la décision thérapeutique n'apparaissait pas suffisamment satisfaisante car trop loin de la mesure de l'effet du traitement que certains pensent mieux approcher en analysant uniquement les périodes pendant lesquels on sait que le traitement a été pris. Le fait de considérer « à tort » un patient traité apparait alors incorrecte. Une telle sélection peut modifier les groupes étudiés et de fait perdre la comparabilité, fournissant alors un effet difficile à définir.

Finalement, dans une analyse en intention de traiter, on évalue l'effet d'une décision thérapeutique. Cette décision thérapeutique pourra effectivement occasionner une non-adhérence, des arrêts ou des changements de traitement. Ces éléments sont inhérents à la vie réelle alors finalement, n'était-ce pas là le vrai effet du traitement en vie réelle ?

Considérer la dépendance dans le temps de l'exposition, par exemple avec les modèles structuraux marginaux pourrait permettre de mieux mesurer l'effet du traitement, tout en conservant la comparabilité.

3. Question d'efficacité à long terme en vie réelle : entre volonté et faisabilité

Dans ce travail, nous souhaitions répondre à une question, par exemple : quelle est l'efficacité à long terme des traitements de première ligne injectables ? ou encore quelle est l'efficacité de l'initiation précoce d'un traitement à long terme et de quel ordre devrait être ce délai d'initiation ? Ces questions ont déjà trouvé réponses à court terme, en particulier au travers d'essais cliniques randomisés : les traitements de première ligne injectables ont démontré leur efficacité, notamment sur la réduction des poussées, les recommandations de prise en charge sont de traiter le plus tôt possible.

Ces premiers éléments de réponses obtenus à partir d'essais randomisés ont bien évidemment déjà impacté la prise en charge des patients et les pratiques des neurologues. Les résultats concernant les traitements de première ligne injectables ont conduit à les prescrire à un grand nombre de patients. De plus, le chapitre 5 a montré l'application des recommandations de traitement précoce. Nous sommes alors en présence d'un problème méthodologique pour l'analyse en vie réelle, celle du manque de patients contrôles. Dans l'étude portant sur les

premières lignes injectables, ils ont conduit à proposer un schéma d'analyse moins habituel. Dans l'étude portant sur l'initiation précoce, le manque de patients traités tardivement a conduit à élargir la période d'étude, mélangeant ainsi les pratiques.

D. Un travail au carrefour de deux disciplines

Face à une question de recherche, on souhaite apporter une réponse claire à laquelle on accorde pleine confiance mais la réalité est plus nuancée. La réponse est faite de discussion, contextualisation au regard des choix en particulier méthodologiques. D'ailleurs, on lit parfois dans des publications des phrases précautionneuses évoquant les potentiels biais de l'étude, sans évoquer ni beaucoup de détails ni l'impact et/ou le sens que ce biais peut avoir eu sur le résultat (sous ou sur estimer un effet, par exemple). Il semble même parfois qu'un résultat biaisé soit perçu comme incorrect. Aujourd'hui, il ne me semble pas qu'il y ait un résultat « optimal » mais des résultats dont l'obtention devrait pouvoir être évoquée en toute transparence et dans un objectif de reproductibilité. Bien évidemment, cela est posé dans une démarche de limitations des biais, notre mission restant de les réduire et les contrôler afin d'obtenir des résultats les moins altérés possibles. C'est l'ensemble de ces résultats qui nous fournira un faisceau d'arguments permettant de conclure. Cette transparence et reproductibilité, ces discussions formalisant en quoi le résultat peut avoir été impacté me semblent parfois difficiles dans le contexte actuel d'évaluation de la recherche, une recherche productive, dans laquelle la quantité des publications et l'impact factor font souvent loi. De plus, la production de résultats divergents avec trop peu d'éléments indiquant pourquoi et souvent dans une démarche naturelle de « vente » de son propre travail, a tendance à contribuer aux doutes dont les études observationnelles sont parfois victimes (Concato *et al.*, 2010; Marko and Weil, 2010). En ce sens, depuis 2004, il existe une liste de recommandations pour présenter les études observationnelles (STROBE pour « STrengthening the Reporting of OBservational studies in Epidemiology »)[24]. Plus récemment, un programme nommé REPEAT pour « Reproducible Evidence: Practices to Enhance and Achieve Transparency » visant à améliorer la transparence, la reproductibilité et la validité de la recherche à partir de données observationnelles est actuellement en cours de réalisation par un équipe de Harvard[25]. Le but

[24] https://www.strobe-statement.org/
[25] https://www.repeatinitiative.org/

de ce programme est d'établir des recommandations pour améliorer la transparence et la présentation des études, en particulier auprès des sociétés professionnelles et des éditeurs.

Comme dans tout travail de recherche, à l'origine, il y a le questionnement. Ce questionnement se place dans le contexte de la littérature existante. Ces connaissances déjà produites ont façonné notre hypothèse. La force de ces hypothèses dépend de la confiance que l'on accorde à l'étude. Cette confiance devrait se lire dans la discussion relative à chaque résultat produit. Pour autant, un résultat ne peut être discuté qu'au regard de l'état des connaissances au moment de sa production ou encore de leur accessibilité. Un des exemples marquants de cette thèse concerne l'étude de l'effet des traitements de première ligne injectables. En effet, bien que les résultats existants issus des études observationnelles soient parfois divergents, l'hypothèse d'un bénéfice reste présente. Dans ce travail, nous avons vu dans le chapitre 3 que plusieurs protocoles ont été envisagés. Les analyses ont systématiquement été conduites montrant des résultats divergents conduisant parfois à démontrer un bénéfice, parfois non, qu'il s'agisse d'une impossibilité de le montrer ou d'obtenir un résultat à l'encontre de l'hypothèse. Nous nous sommes efforcés de garder un regard objectif portant uniquement sur la méthodologie. Cependant, il est parfois difficile de détacher la confiance que l'on porte à un résultat obtenu de la comparaison au résultat qu'on attendait.

Dans cette thèse d'épidémiologie, des connaissances issues de la pharmaco-épidémiologie et de la clinique ont été mobilisées. Les discussions découlant de ce travail font écho à ces deux disciplines. La rencontre de ces connaissances n'a pas toujours été simple durant ce travail. Dans la littérature, j'ai parfois été surprise par le décalage que l'on peut observer entre les deux disciplines. Dans l'utilisation des scores de propension, il semble que parfois la littérature théorique bien que disponible ne soit pas suffisamment utilisée par les auteurs. On peut alors questionner l'accessibilité de la connaissance, non pas en termes de disponibilité mais en termes d'appropriation. Le questionnement peut même concerner l'interaction entre les disciplines. Plus d'interactions permettraient pourtant de conserver de bonnes pratiques sans se priver des contextes théoriques de l'une ou l'autre des disciplines, ni même des contraintes. Cela commence dès la collecte des données. En effet, l'épidémiologiste demande des données de bonne qualité, précises et régulières alors que le contexte clinique est souvent contraint par le temps et des demandes multiples pas toujours absorbables. J'ai d'ailleurs eu l'occasion d'assister à ce type de discussion lors des Assises de l'OFSEP. En effet, depuis juillet 2018, l'OFSEP a proposé une innovation de sa cohorte, en suivant de manière

annuelle et systématique une partie des patients inclus dans la cohorte mère, celle que nous avons utilisée dans cette thèse. En particulier, dans cette cohorte appelée OFSEP HD, un protocole spécifique de collecte des données d'imagerie a été défini afin que celle-ci soit réalisée dans un intervalle de temps proche des initiations de traitement. Il apparaissait lors de cette discussion de grande difficulté à réaliser les IRM dans les temps. De telles données, nous permettraient pourtant de réaliser des études plus robustes. Il faut alors réussir à faire coïncider la recherche et la faisabilité au quotidien. De plus, la communication entre les différents corps de métier est primordiale pour faciliter les interactions. Dans le but de clarifier la conception des études et la communication, le groupe REPEAT propose d'ailleurs des graphiques de conception d'une étude d'efficacité. Ils y présentent, la date d'entrée dans l'étude, les fenêtres d'évaluation des critères d'inclusion et d'exclusion, les fenêtre d'évaluation des variables et la fenêtre de suivi. Ces graphiques peuvent permettre d'aider à éclairer d'éventuels problèmes de violation d'hypothèses parfois observés. Un de leur exemple est proposé en annexe H.

Dans cette thèse, je me suis efforcée de tenir compte au mieux des aspects cliniques, épidémiologiques et statistiques. Bien évidemment, ma connaissance de la pathologie reste encore limitée, bien que non comparable à ce que je connaissais en début de thèse. De plus, sans formation en épidémiologie ou santé publique, plusieurs concepts basiques m'étaient inconnus. J'ai cependant eu la chance de profiter d'un accompagnement de qualité pour m'aider à trouver ma place dans cette discipline. Ma place, au carrefour de deux autres m'a d'ailleurs souvent questionnée. Ni clinicien, ni statisticien, l'épidémiologiste doit pourtant avoir suffisamment de connaissances dans les deux disciplines pour répondre au mieux aux objectifs tout en identifiant ses limites. Aujourd'hui, cette mission me semble donc être de réaliser de bonnes études faisant appel aux bons concepts statistiques, sans parfois pouvoir en maîtriser tous les aspects, tout en s'assurant de faire les bons choix sur le plan de la pathologie et de la prise en charge dans le but de répondre aux questions qui ont un sens pour le clinicien et les patients.

Références

Agius M, Meng X, Chin P, Grinspan A, Hashmonay R. Fingolimod therapy in early multiple sclerosis: An efficacy analysis of the transforms and freedoms studies by time since first symptom. CNS Neurosci Ther 2014; 20: 446–451.

Ahrweiller K, Rousseau C, Le Page E, Bajeux E, Leray E, Michel L, et al. Decreasing impact of late relapses on disability worsening in secondary progressive multiple sclerosis. Mult Scler J 2019: 1–12.

Altman DG, De Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. Br J Cancer 1995; 72: 511–518.

Austin P, Grootendorst P, Anderson G. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Stat Med 2007; 26: 734–753.

Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: A systematic review and suggestions for improvement. J Thorac Cardiovasc Surg 2007; 134: 1128–1135.

Austin PC. The performance of different propensity score methods for estimating marginal odds ratios. 2007; 26: 3078–3094.

Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. Med Decis Mak 2009; 29: 661–677.

Austin PC. The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. Stat Med 2010; 29: 2137–2148.

Austin PC. An introduction to propensity score methods for reducing the effects of

confounding in observational studies. Multivariate Behav Res 2011; 46: 399–424.

Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharm Stat 2011; 10: 150–161.

Austin PC. A tutorial and case study in propensity score analysis: An application to estimating the effect of in-hospital smoking cessation counseling on mortality. Multivariate Behav Res 2011; 46: 119–151.

Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. Stat Med 2013; 32: 2837–2849.

Austin PC. A comparison of 12 algorithms for matching on the propensity score. Stat Med 2014; 33: 1057–1069.

Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Stat Med 2016; 35: 5642–5655.

Austin PC, Schuster T. The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: A simulation study. Stat Methods Med Res 2014; 25: 2214–2237.

Austin PC, Stuart E a. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med 2015; 34: 3661–3679.

Baldassari LE, Fox RJ. Therapeutic Advances and Challenges in the Treatment of Progressive Multiple Sclerosis. Drugs 2018; 78: 1549–1566.

Barbin L, Rousseau C, Casey R, Edan G, Lebrun-frenay C. Comparative efficacy of fingolimod vs natalizumab: A French multicenter observational study. Neurology 2016; 87: 1961.

Barett J, Zuppa A. Basic principles of Clinical Pharmacology relevant to pharmacoepidemiology studies. In: Pharmacoepidemiology. 2012. p. 23–38

Le Borgne F. Conception d’un outil simple d’utilisation pour réaliser des analyses statistiques ajustées valorisant les données de cohortes observationnelles de pathologies chroniques.

Doctorat de biostatistique, direction de Yohann Foucher et Magali Giral, Nantes, Fr. 2016

Le Borgne F, Giraudeau B, Querard AH, Giral M, Foucher Y. Comparisons of the performance of different statistical tests for time-to-event analysis with confounding factors: Practical illustrations in kidney transplantation. Stat Med 2016; 35: 1103–1116.

Brassat D. Chapitre2. Génétique. In: Sclérose en plaques. 2017. p. 23–30

Braune S, Lang M, Bergmann a. Second line use of Fingolimod is as effective as Natalizumab in a German out-patient RRMS-cohort. J Neurol 2013; 260: 2981–2985.

Brookhart M alan, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol 2006; 163: 1149–1156.

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; 83: 1022–1024.

Buttmann M. Where mitoxantrone for multiple sclerosis is still valuable in 2018. Eur J Neurol 2018; 25: 1400–1401.

Butzkueven H, Chapman J, Cristiano E, Grand'Maison F, Hoffmann M, Izquierdo G, et al. MSBase: An international, online registry and platform for collaborative outcomes research in multiple sclerosis. Mult Scler 2006; 12: 769–774.

Calocer F, Dejardin O, Droulon K, Launoy G, Defer G. Socio-economic status influences access to second-line disease modifying treatment in Relapsing Remitting Multiple Sclerosis patients. PLoS One 2018; 13: e0191646.

Chartier N, Epstein J, Soudant M, Dahan C, Michaud M, Pittion-Vouyovitch S, et al. Clinical follow-up of 411 patients with relapsing and progressive multiple sclerosis 10 years after discontinuing mitoxantrone treatment: a real-life cohort study. Eur J Neurol 2018; 25: 1439–1445.

Chevret S. Pourquoi et comment randomiser un essai thérapeutique ? Rev Mal Respir 2003; Vol 20: 753–756.

Clavelou P, Vuadens P, Casez O. Traitements symptomatiques. In: Sclérose en plaques. 2017. p. 267–275

Cocco E, Sardu C, Spinicci G, Musu L, Massa R, Frau J, et al. Influence of treatments in multiple sclerosis disability: A cohort study. Mult Scler J 2015; 21: 433–441.

Cohen J a., Barkhof F, Comi G, Hartung H-P, Khatri BO, Montalban X, et al. Oral Fingolimod or Intramuscular Interferon for Relapsing Multiple Sclerosis. N Engl J Med 2010; 362: 402–415.

Cohen JA, Barkhof F, Comi G, Izquierdo G, Khatri B, Montalban X, et al. Fingolimod versus intramuscular interferon in patient subgroups from TRANSFORMS. J Neurol 2013; 260: 2023–2032.

Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. Comput Methods Programs Biomed 2004: 45–49.

Cole SR, Hernàn MA. Practice of Epidemiology Constructing Inverse Probability Weights for Marginal Structural Models. Am J Epidemiol 2008; 168: 656–664.

Comi G, Filippi M, Barkhof F, Durelli L, Edan G, Fernández O, et al. Effect of early interferon treatment on conversion to definite multiple sclerosis: A randomised study. Lancet 2001; 357: 1576–1582.

Comi G, Filippi M, Wolinsky JS, European/Canadian Glatiramer Acetate Study group. European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer acetate on magnetic resonance imaging–measured disease activity and burden in patients with relapsing multiple sclerosis. Ann Neurol 2001; 49: 290–297.

Comi G, Martinelli V, Rodegher M, Moiola L, Bajenaru O, Carra a., et al. Effect of glatiramer acetate on conversion to clinically definite multiple sclerosis in patients with clinically isolated syndrome (PreCISe study): a randomised, double-blind, placebo-controlled trial. Lancet 2009; 374: 1503–1511.

Comi G, Martinelli V, Rodegher M, Moiola L, Leocani L, Bajenaru O, et al. Effects of early treatment with glatiramer acetate in patients with clinically isolated syndrome. Mult Scler 2013; 19: 1074–83.

Comi G, De Stefano N, Freedman M, Barkhof F, Polman C, Uitdehaag B, et al. Comparison of two dosing frequencies of subcutaneous interferon beta-1a in patients with a first clinical demyelinating event suggestive of multiple sclerosis (REFLEX): A phase 3 randomised

controlled trial. Lancet Neurol 2012; 11: 33–41.

Commenges D, Jacqumin-Gadda H. Chapitre 4 : Analyse des données de survie. In: Modèles biostatistiques pour l'épidémiologie. 2015. p. 103–162

Concato J, Lawler E V., Lew RA, Gaziano JM, Aslan M, Huang GD. Observational methods in comparative effectiveness research. Am J Med 2010; 123: e16–e23.

Concato J, Shah N, Horwitz R. Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs. N Engl J Med 2000; 342: 1887–1892.

Confavreux C, Compston D, Hommes O, McDonald W, Thompson A. EDMUS, a European database for multiple sclerosis. J Neurol Neurosurg Psychiatry 1992; 55: 671–676.

Confavreux C, O'Connor P, Comi G, Freedman MS, Miller AE, Olsson TP, et al. Oral teriflunomide for patients with relapsing multiple sclerosis (TOWER): A randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Neurol 2014; 13: 247–256.

Confavreux C, Vukusic S. Natural history of multiple sclerosis: A unifying concept. Brain 2006; 129: 606–616.

Confavreux C, Vukusic S, Adeleine P. Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. Brain 2003; 126: 770–782.

Cotton F, Edan G. L'Observatoire français de la sclérose en plaques : vers un consensus IRM national. La Lett du Neurol 2015; 135: 175–176.

Cox DR. Regression Models and Life-Tables. J R Stat Soc 1972; 34: 187–220.

D'Amico E, Ziemssen T, Cottone S. Induction therapy for the management of early relapsing forms of multiple sclerosis. A critical opinion. Expert Opin Pharmacother 2017; 18: 1553–1556.

Debouverie M, Laforest L, Van Ganse E, Guillemin F, Godet E, Jager A, et al. Earlier disability of the patients followed in Multiple Sclerosis centers compared to outpatients. Mult Scler 2009; 15: 251–257.

Debouverie M, Pittion-Vouyovitch S, Louis S, Guillemin F. Natural history of multiple sclerosis in a population-based cohort. Eur J Neurol 2008; 15: 916–921.

Derache N, Dufay A, Lebarbey C. Organization of care for multiple sclerosis in France. Rev Neurol (Paris) 2018; 174: 475–479.

Drulovic J, Kostic J, Mesaros S, Dujmovic Basuroski I, Stojsavljevic N, Kisic-Tepavcevic D, et al. Interferon-beta and disability progression in relapsing-remitting multiple sclerosis. Clin Neurol Neurosurg 2013; 115: S65–S69.

Edan G. Chapitre 7. Evénements cliniques et évolution pronostique. In: Sclérose en plaques. 2017. p. 97–110

Edan G, Comi G, Le Page E, Leray E, Rocca MA, Filippi M. Mitoxantrone prior to interferon beta-1b in aggressive relapsing multiple sclerosis: A 3-year randomised trial. J Neurol Neurosurg Psychiatry 2011; 82: 1344–1350.

Edan G, Le Page E. Induction Therapy for Patients with Multiple Sclerosis : Why ? When ? How ? 2013: 403–409.

Edan G, Rolland Y, Sabouraud O, Miller D, Moseley I, Clanet M, et al. Therapeutic effect of mitoxantrone combined with methylprednisolone in multiple sclerosis: A randomised multicentre study of active disease using MRI and clinical criteria. J Neurol Neurosurg Psychiatry 1997; 62: 112–118.

European Agency for the Evaluation of Medicinal Products. Points To Consider on Switching Between Superiority and Non-Inferiority Discussion in the Efficacy Working Party (Ewp) Transmission To the Cpmp Release for Consultation Re-Submission To the Efficacy Working Party. 2000: 0–10.

Fassin D, Grandjean H, Kaminski M, Lang T, Leclerc A. Les inégalités sociales de santé. Recherches. 2000

Fernández Ó. Is there a change of paradigm towards more effective treatment early in the course of apparent high-risk MS? Mult Scler Relat Disord 2017; 17: 75–83.

Filippini G, C DG, Clerico M, Beiki O, Mattoscio M, Piazza F, et al. Treatmentwith disease-modifying drugs for peoplewith a first clinical attack suggestive ofmultiple sclerosis (Review) Filippini. Cochrane Database Syst Rev 2017: CD012200.

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: 1185–1192.

Fox RJ, Miller DH, Phillips T, Hutchinson M, Havrdova E, Kita M, et al. Placebo-Controlled Phase 3 Study of Oral BG-12 or Glatiramer in Multiple Sclerosis. N Engl J Med 2012; 367: 1087–1097.

Freedman MS, Comi G, De Stefano N, Barkhof F, Polman CH, Uitdehaag BMJ, et al. Moving toward earlier treatment of multiple sclerosis: Findings from a decade of clinical trials and implications for clinical practice. Mult Scler Relat Disord 2014; 3: 147–155.

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.

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: 1108–1115.

Fujihara K. BENEFIT 8-year results provide further support for the long-term value of early treatment of multiple sclerosis. J Neurol Neurosurg Psychiatry 2014; 85: 1179–1179.

Gajofatto A, Bianchi MR, Deotto L, Benedetti MD. Are natalizumab and fingolimod analogous second-line options for the treatment of relapsing-remitting multiple sclerosis? A clinical practice observational study. Eur Neurol 2014; 72: 173–180.

Gayat E, Resche-Rigon M, Mary JY, Porcher R. Propensity score applied to survival data analysis through proportional hazards models: A Monte Carlo study. Pharm Stat 2012; 11: 222–229.

Glaser A, Stahmann A, Meissner T, Flachenecker P, Horáková D, Zaratin P, et al. Multiple sclerosis registries in Europe – An updated mapping survey. Mult Scler Relat Disord 2019; 27: 171–178.

Gold R, Kappos L, Arnold D, Bar-O A, Giovannoni G, Selmaj K, et al. Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. N Engl J Med 2012; 367: 1098–1107.

Goulden R, Ibrahim T, Wolfson C. Is high socioeconomic status a risk factor for multiple sclerosis? A systematic review. Eur J Neurol 2015; 22: 899–911.

Gout O, Lebrun-Frenay C, Labauge P, Le Page GE, Clavelou P, Allouche S. Prior suggestive

symptoms in one-third of patients consulting for a ‘first’ demyelinating event. J Neurol Neurosurg Psychiatry 2011; 82: 323–325.

Granziera C, Fischi-Gomez E, Maarouf A, Pelletier J. Chapitre 15: Recherche de marqueurs d’imagerie par résonance magnétique. In: Sclérose en plaques. 2017. p. 213–226

Grossmann C, Alper J. Observational Studies in a Learning Health System: Workshop Summary. In: National Academies Press (US), editor(s). Generalizing Randomized Clinical Trial Results to Broader. Washington: National Academies Press (US); 2013

Gupta SK. Intention-to-treat concept : A review. Persepctives Clin Res 2011; 2: 109–112.

Guyatt G, Cairns J, Churchill D, Cook D, Haynes B, Hirsh J, et al. A New Approach to Teaching the Practice of Medicine. JAMA 1992; 268: 2420–2425.

Hajage D. Utilisation du score de propension et du score pronostique en pharmacoepidemiologie. Doctorat de biostatistique, direction de Florence Tubach et Yann De Rycke, Paris, France. 2017

Hannan EL. Randomized Clinical Trials and Observational Studies. Guidelines for Assessing Respective Strengths and Limitations. JACC Cardiovasc Interv 2008; 1: 211–217.

Hartung H-P, Morrissey SP, Gonsette R, König N, Kwiecinski H, Guseo A, et al. Mitoxantrone in progressive multiple sclerosis: A placebo-controlled, double-blind, randomised, multicentre trial. Lancet 2002; 360: 2018–2025.

HAS. Natalizumab. 2017

HAS. Fingolimod. 2019

Hautecoeur P. Chapitre 1. Historique de la sclérose en plaques. In: La sclérose en plaques. Historique, épidémiologie et pathogénie. 2019. p. 3–30

Hernàn MA, Robins J. Chapter3: Observational studies. In: Causal Inference. 2018. p. 25–40

Higue-van Steenbrugghe F. Prise en charge globale de la sclérose en plaques. 2016

Hill B. Causation and Association. Procedings R Soc Med 1965; 58: 295–300.

Houtchens MK, Kolb CM. Multiple sclerosis and pregnancy: Therapeutic considerations. J Neurol 2013; 260: 1202–1214.

Imbens GW. Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. Rev Econ Stat 2004; 86: 4–29.

Ioannidis JPA. Exposure-wide epidemiology: Revisiting Bradford Hill. Stat Med 2016; 35: 1749–1762.

Irwin JO. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. J Hyg (Lond) 1949; 47: 188–189.

Izquierdo G, Damas F, Páramo MD, Ruiz-Peña JL, Navarro G. The real-world effectiveness and safety of fingolimod in relapsing-remitting multiple sclerosis patients: An observational study. PLoS One 2017; 12: 1–12.

Jacobs L, Beck R, Simon J, Philip Kinkel R, Brownscheidle C, Murray T, et al. Intramuscular interferon Beta-1atherapy initiated during a first demyelinating event in multiple sclerosis. J Med 2000; 343: 898–904.

Jacobs L, Cookfair D, Rudick R, Herndon R, Richert J, Salazar A, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. Ann Neurol 1996; 39: 285–294.

Johnson K, Brooks B, Cohen J, Ford C, Goldstein J, Lisak R, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. Neurology 1995; 45: 1268–1276.

Kalincik T. Multiple sclerosis relapses: Epidemiology, outcomes and management. A systematic review. Neuroepidemiology 2015; 44: 199–214.

Kalincik T, Horakova D, Spelman T, Jokubaitis V, Trojano M, Lugaresi A, et al. Switch to natalizumab versus fingolimod in active relapsing-remitting multiple sclerosis. Ann Neurol 2015; 77: 425–435.

Kaminski M, Lang T, Leclerc A. Inégaux face à la santé. INSERM; 2008

Kaplan EL, Meier P. Nonparametric estimation from incomplete samples. J ASA 1958; 73: 457–481.

Kappos L, Edan G, Freedman M, Montalban X, Hartung HP, Hemmer B, et al. The 11-year long-term follow-up study from the randomized BENEFIT CIS trial. Neurology 2016; 87:

978–987.

Kappos L, Polman CH, Freedman MS, Edan G, Hartung HP, Miller DH, et al. Treatment with interferon beta-1b delays conversion to clinically definite and McDonald MS in patients with clinically isolated syndromes. Neurology 2006; 67: 1242–1249.

Kappos L, Radue EW, O'Connor P, Polman C, Hohfeld R, Calabresi P a., et al. A Placebo-Controlled Trial of Oral Fingolimod in Relapsing Multiple Sclerosis Ludwig. New Engl J 2010; 362: 387–401.

Kappos L, Traboulsee a, Constantinescu C, Forrestal F, Jongen P, Pollard J, et al. Long-term subcutaneous inter·feron beta-la therapy in patients. Neurology 2006; 67: 944–953.

Karim ME, Gustafson P, Petkau J, Zhao Y, Shirani A, Kingwell E, et al. Marginal structural cox models for estimating the association between β-interferon exposure and disease progression in a multiple sclerosis cohort. Am J Epidemiol 2014; 180: 160–171.

Kavaliunas A, Manouchehrinia A, Stawiarz L, Ramanujam R, Agholme J, Hedström AK, et al. Importance of early treatment initiation in the clinical course of multiple sclerosis. Mult Scler J 2016

Kerbrat A, Hamonic S, Leray E, Tron I, Edan G, Yaouanq J. Ten-year prognosis in multiple sclerosis: A better outcome in relapsing-remitting patients but not in primary progressive patients. Eur J Neurol 2015; 22: 507-e35.

Khatri B, Barkhof F, Comi G, Hartung HP, Kappos L, Montalban X, et al. Comparison of fingolimod with interferon beta-1a in relapsing-remitting multiple sclerosis: A randomised extension of the TRANSFORMS study. Lancet Neurol 2011; 10: 520–529.

Kieseier BC. The mechanism of action of interferon-β in relapsing multiple sclerosis. CNS Drugs 2011; 25: 491–502.

Kleinschmidt-DeMasters, B.K. Tyler KL. Progressive Multifocal Leukoencephalopathy Complicating Treatment with Natalizumab and Interferon Beta-1a for Multiple Sclerosis. N Engl J Med 2005; 353: 369–374.

Koch-Henriksen N, Magyari M, Sellebjerg F, Soelberg Sørensen P. A comparison of multiple sclerosis clinical disease activity between patients treated with natalizumab and fingolimod. Mult Scler J 2017; 23: 234–241.

Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983; 33: 1444–1452.

Lachin JM. Statistical properties of randomization in clinical trials. Control Clin Trials 1988; 9: 289–311.

Laffaldano P, Lucisano G, Butzkueven H, Hillert J, Hyde R, Koch-Henriksen N, et al. Big multiple sclerosis data network: marginal structural models confirm the long-term impact of DMTs on disability progression. In: ECTRIMS. Stockolm; 2019

Landfeldt E, Castelo-Branco A, Svedbom A, Löfroth E, Kavaliunas A, Hillert J. The long-term impact of early treatment of multiple sclerosis on the risk of disability pension. J Neurol 2018: 701–707.

Langer-Gould A, Atlas SW, Green AJ, Bollen AW, Pelletier D. Progressive Multifocal Leukoencephalopathy in a Patient Treated with Natalizumab. N Engl J Med 2005; 353: 375–381.

Langer-Gould A, Popat RA, Huang SM, Cobb K, Fontoura P, Gould MK, et al. Clinical and demographic predictors of long-term disability in patients with relapsing-remitting multiple sclerosis: A systematic review. Arch Neurol 2006; 63: 1686–1691.

Leber P. Slowing the progression of Alzheimer disease: methodologic issues. Alzheimer Dis Assoc Disord 1997: 10–21.

Leone MA, Bonissoni S, Collimedaglia L, Tesser F, Calzoni S, Stecco A, et al. Factors predicting incomplete recovery from relapses in multiple sclerosis: A prospective study. Mult Scler 2008; 14: 485–493.

Leray E, Vukusic S, Debouverie M, Clanet M, Brochet B, Sèze J De, 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.

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: 1900–1913.

Lhermitte F, Alperovitch A, Lyon-Caen O. Enquête sur la sclérose en plaques lancée par la télévision française. Rev Neurol 1988; 144: 135–136.

Lin DY, Wei LJ, Ying Z. Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika 1993; 80: 557–572.

Lu B. Propensity score matching with time-dependent covariates. Biometrics 2005; 61: 721–728.

Lubetzki C, Stankoff B. Chapitre 4. Myélinisation, démyélinisation, neurodégénérescence. In: Sclérose en plaques. 2017. p. 47–62

Lublin FD, Reingold SC, Cohen JA, Cutter GR, Thompson AJ, Wolinsky JS, et al. Defining the clinical course of multiple sclerosis: The 2013 revisions. Neurology 2014; 83: 278–286.

Magyari M, Koch-Henriksen N, Sørensen PS. The Danish multiple sclerosis treatment register. Clin Epidemiol 2016; 8: 549–552.

Mahalanobis PC. On the Generalized Distance in Statistics. roceedings Natl Inst Sci India 1936; 2: 49–55.

Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 1966; 50: 163–170.

Marko NF, Weil RJ. The Role of Observational Investigations in Comparative Effectiveness Research. PLoS Med 2010; 13

Marriott JJ, Miyasaki JM, Gronseth G, O'Connor PW. Evidence Report: The efficacy and safety of mitoxantrone (Novantrone) in the treatment of multiple sclerosis. Neurology 2010; 74: 1463–1470.

Mcdonald WI, Compston A, Edan G, Goodkin D, Hartung H, Lublin FD, et al. Recommended Diagnostic Criteria for Multiple Sclerosis : Guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. Ann Neurol 2001: 121–127.

McDonald WI, Miller DH, Barnes D. The pathological evolution of multiple sclerosis. Neuropathol Appl Neurobiol 1992; 18: 319–334.

McKay KA, Jahanfar S, Duggan T, Tkachuk S, Tremlett H. Factors associated with onset, relapses or progression in multiple sclerosis: A systematic review. Neurotoxicology 2017; 61: 189–212.

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: 1093–1100.

Merkel B, Butzkueven H, Traboulsee AL, Havrdova E, Kalincik T. Timing of high-ef fi cacy therapy in relapsing-remitting multiple sclerosis : A systematic review. Autoimmun Rev 2017; 16: 658–665.

Miller DH, Chard DT, Ciccarelli O. Clinically isolated syndromes. Lancet Neurol 2012; 11: 157–169.

Montalban X, Gold R, Thompson AJ, Otero-Romero S, Amato MP, Chandraratna D, et al. ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. Mult Scler 2018; 24: 96–120.

Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. Am J Epidemiol 2011; 174: 1213–1222.

Myhr KM, Vedeler C, Grønning M, Nyland HI, Riise T, Midgard R. Disability and prognosis in multiple sclerosis: Demographic and clinical variables important for the ability to walk and awarding of disability pension. Mult Scler 2001; 7: 59–65.

Nayan M, Hamilton RJ, Juurlink DN, Finelli A, Kulkarni GS, Austin PC. Critical appraisal of the application of propensity score methods in the urology literature. BJU Int 2017; 120: 873–880.

O'Connor P, Wolinsky JS, Confavreux C, Comi G, Kappos L, Olsson TP, et al. Randomized Trial of Oral Teriflunomide for Relapsing Multiple Sclerosis. N Engl J Med 2011; 365: 1293–1303.

Ogdie A, Langan SM, Parkinson J, Dattani H, Kostev K, Gelfand JM. Medical Record Databases. In: Pharmacoepidemiology. 2012. p. 224–243

Le Page E, Leray E, Edan G. Long-term safety profile of mitoxantrone in a French cohort of 802 multiple sclerosis patients: A 5-year prospective study. Mult Scler J 2011; 17: 867–875.

Pantazou V, Du Pasquier R, Leray E. Chapitre3. Facteurs de risque environnementaux. In: Sclérose en plaques. 2017. p. 31–46

Papeix C. Chapitre 8. Etablissement du diagnostic. In: Sclérose en plaques. 2017. p. 111–120

Pearl J. Causality. Cambridge University Press; 2009

Pirracchio R, Carone M, Rigon MR, Caruana E, Mebazaa A, Chevret S. Propensity score estimators for the average treatment effect and the average treatment effect on the treated may yield very different estimates. Stat Methods Med Res 2013; 25: 1938–1954.

Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, et al. A Randomized, Placebo-Controlled Trial of Natalizumab for Relapsing Multiple Sclerosis. J new Engl Med 2006; 354: 899–910.

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: 292–302.

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: 840–846.

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: 227–231.

PRISMS. Randomised double-blind placebo-controlled study of interferon β-1a in relapsing/remitting multiple sclerosis. Lancet 1998; 352: 1498–1504.

Putzki N, Yaldizli Ö, Bühler R, Schwegler G, Curtius D, Tettenborn B. Natalizumab reduces clinical and MRI activity in multiple sclerosis patients with high disease activity: Results from a multicenter study in Switzerland. Eur Neurol 2010; 63: 101–106.

Putzki N, Yaldizli O, Mäurer M, Cursiefen S, Kuckert S, Klawe C, et al. Efficacy of natalizumab in second line therapy of relapsing-remitting multiple sclerosis: Results from a multi-center study in German speaking countries. Eur J Neurol 2010; 17: 31–37.

Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000; 11: 550–560.

Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational

Studies for Causal Effects Effects. Biometrika 1983; 70: 41–55.

Rothwell PM. External validity of randomised controlled trials: 'To whom do the results of this trial apply?' Lancet 2005: 82–93.

Roux J. Parcours de soins des patients atteints de sclérose en plaques à partir des données médico-administratives en France. Doctorat d'Epidémiologie, direction de Emmanuelle Leray et Olivier Grimaud, Rennes 1, France. 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 ? 2016

Rudick R a., Stuart WH, Calabresi P a., Confavreux C, Galetta SL, Radue E-W, et al. Natalizumab plus Interferon Beta-1a for Relapsing Multiple Sclerosis. N Engl J Med 2006; 354: 911–923.

Runmarker B, Andersen O. Prognostic factors in a multiple sclerosis incidence cohort with twenty-five years of follow-up. Brain 1993; 116: 117–134.

Scalfari A, Neuhaus A, Daumer M, Ebers GC, Muraro PA. Age and disability accumulation in multiple sclerosis. Neurology 2011; 77: 1246–1252.

Schoenfeld. Partial Residuals for The Proportional Hazards Regression Model. Biometrika 1982; 69: 239–241.

Sedgwick P. What is per protocol analysis? BMJ 2013; 346: 1–2.

Shirani A, Zhao Y, Karim ME, Evans C, Kingwell E, Van Der Kop M, et al. Association between Use of Interferon Beta and Progression of Disability in Patients. Jama 2012; 308: 247–256.

Shuster JJ. Median follow-up in clinical trials. J Clin Oncol 1991; 9: 191–2.

Signori A, Gallo F, Bovis F, Di Tullio N, Maietta I, Sormani MP. Long-term impact of interferon or Glatiramer acetate in multiple sclerosis: A systematic review and meta-analysis. Mult Scler Relat Disord 2016; 6: 57–63.

Sormani MP, Bruzzi P. Can we measure long-term treatment effects in multiple sclerosis? Nat Rev Neurol 2014; 11: 176–182.

Strom B. What is Pharmacoepidemiology? In: Pharmacepidemiology. 2012. p. 3–22

Strom BL. Methodologic Challenges to Studying Patient Safety and Comparative Effectiveness. Med Care 2007; 45: S13–S15.

Strom BL, Schinnar R, Hennesy S. Chapitre 32: Comaprative effectivennes research. In: Pharmacoepidemiology. 2012. p. 561–579

Sweeney V, Sadovnick A, Brandejs V. Prevalence of multiple sclerosis in British Columbia. Can J Neurol Sci 1986; 13: 47–51.

The IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. Neurology 1993; 43: 655–661.

Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: revision of the McDonald criteria 2017. Lancet Neurol 2018; 89: 1344–1354.

Tintoré M. Rationale for early intervention with immunomodulatory treatments. J Neurol 2008; 255: 37–43.

Tintoré M. New options for early treatment of multiple sclerosis. J Neurol Sci 2009; 277: 9–11.

Tintoré M. The Barcelona cohort. In: The Barcelona cohort. Berlin, Germany: ECTRIMS; 2018

Tintore M, Rovira À, Río J, Otero-Romero S, Arrambide G, Tur C, et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. Brain 2015; 138: 1863–1874.

Tremlett H, Paty D, Devonshire V. Disability progression in multiple sclerosis is slower than previously reported. Neurology 2006; 66: 172–177.

Trojano M, Bergamaschi R, Amato MP, Comi G, Ghezzi A, Lepore V, et al. The Italian multiple sclerosis register. Neurol Sci 2019; 40: 907–907.

Trojano M, Liguori M, Bosco Zimatore G, Bugarini R, Avolio C, Paolicelli D, et al. Age-related disability in multiple sclerosis. Ann Neurol 2002; 51: 475–480.

Trojano M, Pellegrini F, Fuiani A, Paolicelli D, Zipoli V, Zimatore GB, et al. New natural history of interferon-??-treated relapsing multiple sclerosis. Ann Neurol 2007; 61: 300–306.

Trojano M, Pellegrini F, Paolicelli D, Fuiani a, Zimatore GB, Tortorella C, et al. Real-life impact of early interferon beta therapy in relapsing multiple sclerosis. Ann Neurol 2009; 66: 513–20.

Vargas DL, Tyor WR. Update on disease-modifying therapies for multiple sclerosis. J Investig Med 2017; 65: 883–891.

Vaughan AS, Kelley CF, Luisi N, Del Rio C, Sullivan PS, Rosenberg ES. An application of propensity score weighting to quantify the causal effect of rectal sexually transmitted infections on incident HIV among men who have sex with men Data analysis, statistics and modelling. BMC Med Res Methodol 2015; 15: 1–9.

Vermersch P, Berger T, Gold R, Lukas C, Rovira A, Meesen B, et al. The clinical perspective: How to personalise treatment in MS and how may biomarkers including imaging contribute to this? Mult Scler 2016; 22: 18–33.

Vermersch P, Schluep M. Chapitre 17. Traitements de fond et leurs complications. In: Sclérose en plaques. 2017. p. 245–266

Veugelers PJ, Fisk JD, Brown MG, Stadnyk K, Sketris IS, Murray TJ, et al. Disease progression among multiple sclerosis patients before and during a disease-modifying drug program: a longitudinal population-based evaluation. Mult Scler 2009; 15: 1286–1294.

Voskuhl R, Momtazee C. Pregnancy: Effect on Multiple Sclerosis, Treatment Considerations, and Breastfeeding. Neurotherapeutics 2017; 14: 974–984.

Vukusic S, Casey R, Rollot F, Brochet B, Pelletier J, Laplaud DA, et al. Observatoire Français de la Sclérose en Plaques (OFSEP): A unique multimodal nationwide MS registry in France. Mult Scler J 2018: 1–5.

Vukusic S, Confavreux C. Natural history of multiple sclerosis: Risk factors and prognostic indicators. Curr Opin Neurol 2007; 20: 269–274.

Witte J, Didelez V. Covariate selection strategies for causal inference: Classification and comparison. Biometrical J 2018: 1–20.

Wolinsky JS, Narayana PA, O'Connor P, Coyle PK, Ford C, Johnson K, et al. Glatiramer acetate in primary progressive multiple sclerosis: Results of a multinational, multicenter, double-blind, placebo-controlled trial. Ann Neurol 2007; 61: 14–24.

Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. Stat Med 2005; 24: 3089–3110.

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: 321–328.

Yoshida K, Solomon D, Kim S. Active-comparator design and new-user design in observational studies. 2015; 11: 437–441.

Zhao H, Tsiatis A. Efficient Estimation of the Distribution. Biometrics 1999: 1101–1107.

Zhao L, Claggett B, Tian L, Uno H, Pfeffer MA, Solomon SD, et al. On the restricted mean survival time curve in survival analysis. Biometrics 2016; 72: 215–221.

Annexes

A. Systèmes fonctionnels et échelle du score EDSS 248

B. Prise en charge des patients traités par natalizumab vis-à-vis du risque de LEMP 252

C. Exemple de l'interface du logiciel EDMUS pour un patient 254

D. Caractéristiques de 19 cohortes spécifiques SEP en disponible dans le monde..... 256

E. Sharmin S et al, Communication affichée, ECTRIMS, 2019 258

F. Evaluation de l'équilibre de l'étude de l'efficacité de l'initiation précoce du traitement, critère de jugement EDSS 6..... 260

G. Représentation de 1000 valeurs de Hazard Ratio obtenus après un appariement aléatoire 262

H. Exemple de graphique de conception d'une étude par REPEAT 260

A. Systèmes fonctionnels et échelle du score EDSS

Table A1 : 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
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
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

Score	Critères
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
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

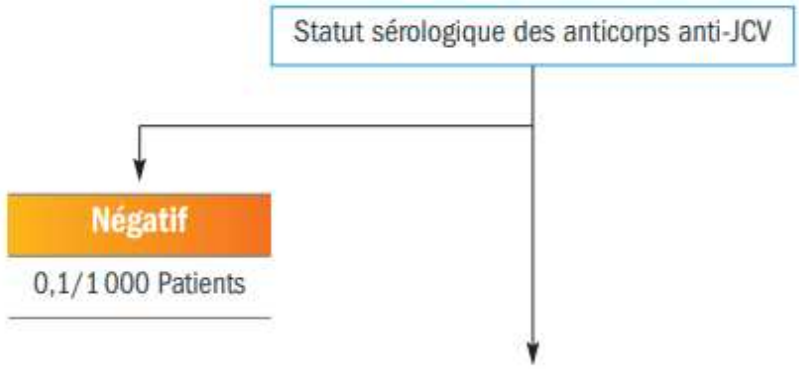
Table A2 : Echelle du score EDSS

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+)
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

B. Prise en charge des patients traités par natalizumab vis-à-vis du risque de LEMP

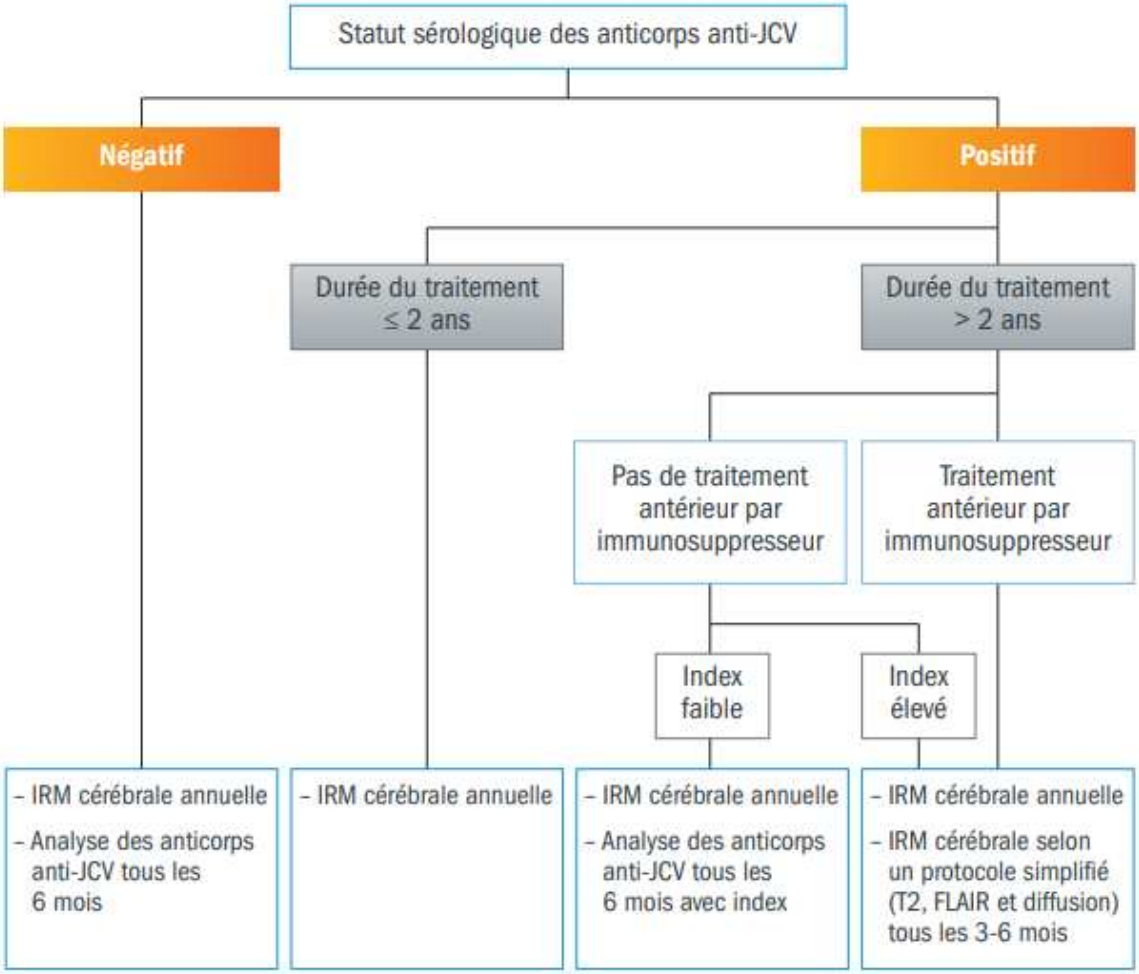
Source : Guide de prescription Biogen France, disponible : www.biogen-france.fr , *consulté le 24/09/2019.*

Figure B1 - Algorithme d'estimation du risque de LEMP

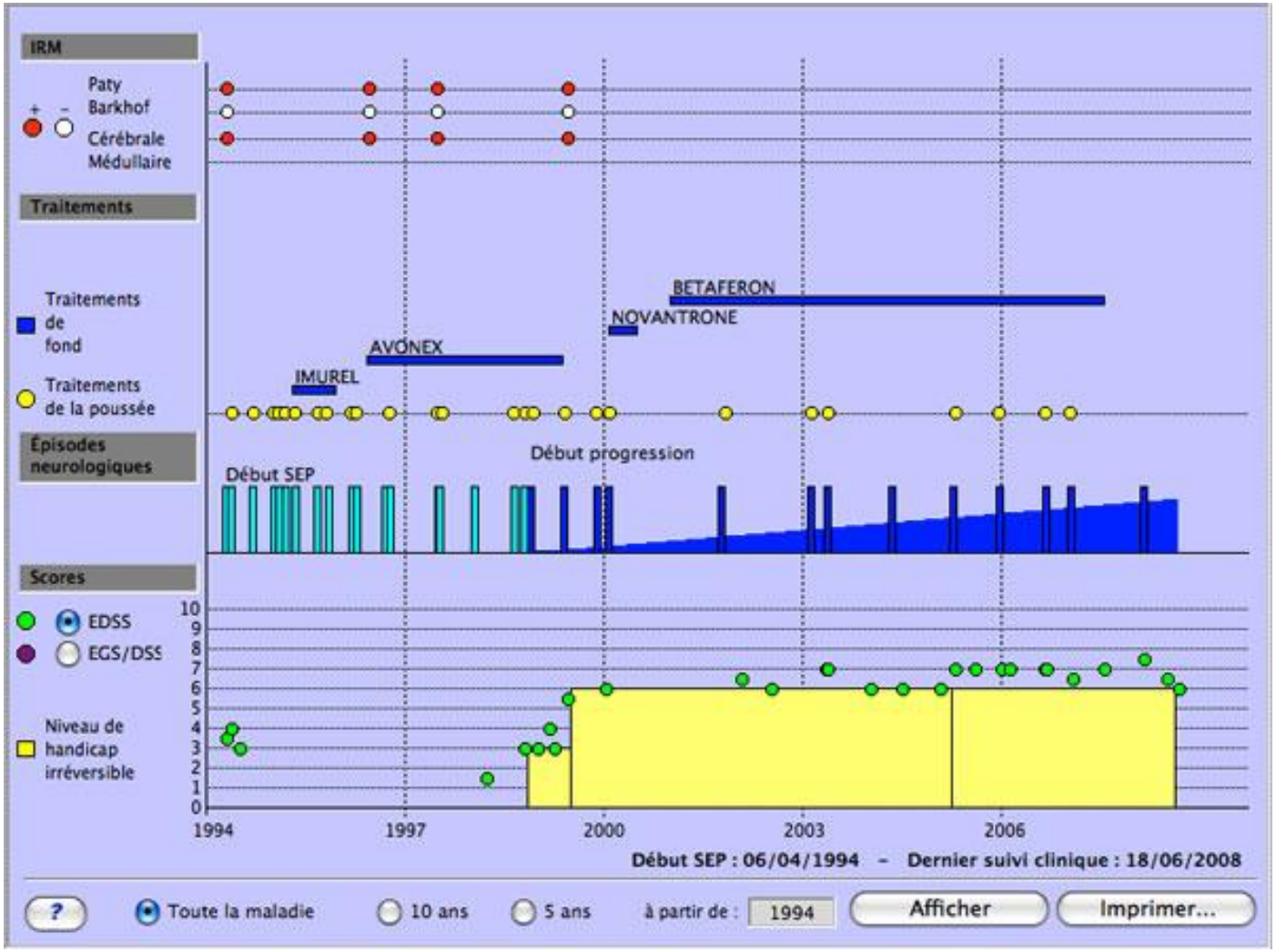


Positif					
Durée d'exposition au natalizumab	Estimation du risque de LEMP pour 1 000 patients				
	Patient sans traitement antérieur par immunosuppresseur				Patient ayant eu un traitement antérieur par immunosuppresseur
	Absence de valeur d'index	Index ≤ 0,9	0,9 < Index ≤ 1,5	Index > 1,5	
1-12 mois	0,1	0,1	0,1	0,2	0,3
13-24 mois	0,6	0,1	0,3	0,9	0,4
25-36 mois	2	0,2	0,8	3	4
37-48 mois	4	0,4	2	7	8
49-60 mois	5	0,5	2	8	8
61-72 mois	6	0,6	3	10	6

Figure B2 - Recommandations pour le suivi du patient



C. Exemple de l'interface du logiciel EDMUS pour un patient



source : https://www.edmus.org/fr/soft/features_use04.html

D. Caractéristiques de 19 cohortes spécifiques SEP en disponible dans le monde

Country	Name of MS registry	Institution	Type of organisation	Start in year	No of patients 2017[b] (No of patients 2014)[c]	Website
Australia[a]	MSBase	MSBase Foundation Ltd	Academic /Research institution, Independent not for profit	2004	49.800	msbase.org
Belgium	BELTRIMS	Belgian Study Group for MS	–	2012	1.000	–
Croatia	Association of multiple sclerosis societies of Croatia AMSSC	Association of MS Societies of Croatia	Patient organisation	2006	2.817 (2.477)	sdmsh.hr/baza; mail.sdmsh.hr/baza2/
Czech Republic	ReMuS	IMPULS	non-profit endowment fund	2013	10.999	nfimpuls.cz
Denmark	The Danish Multiple Sclerosis Registry	The Danish Multiple Sclerosis Center, Rigshospitalet, Copenhagen	Health care organisation	1956	25.000 (16.000)	rigshospitalet.dk/english/departments/neuroscience-centre/department-of-neurology/research/the-danish-multiple-sclerosis-registry
Finland	Finnish MS Register	Finnish Hospital Districts	Health care organisation	2014	8.746	neurorekisteri.fi
France	OFSEP	EDMUS Foundation, Université Claude Bernard, Hospices Civils de Lyon	Academic /Research institution	1976 (2011)[d]	56.400 (40.000)	ofsep.org
Germany						
-GMSR	MS-Register der DMSG (Bundesverband e.V.)	MS Forschungs- und Projektentwicklungs-gGmbH	Non-Profit Organisation	2001 (2014)[e]	48.000 (30.000)	www.msregister.de
-MSDS3D	MSDS study group	Center of Clinical Neuroscience Dresden for the MSDS3D Reasearch Group	Association of research centres	1998	>7.000	msz.uniklinikum-dresden.de/ehealth;
-DMSKW	DMSKW	Kerstin Hellwig	Academic /Research institution	2006	1.500	ms-und-kinderwunsch.de
Greece	The Greek MS Society	The Greek MS Society	Patient organisation	2011	5.323 (3.500)	info@gmss.gr
Italy	Italian MS Register	Fondazione Italiana Sclerosi Multipla (FISM) /University of Bari (UNIBA)	Patient organisation, Academic /Research institution	2001 (2016)[f]	44.894 (20.000)	registroitalianosm.it/
-Liguria	Liguria Regional MS Register[g]			2012	929	registrosclerosimultipla.it
Norway	Norwegian MS Registry & Biobank	Haukeland University Hospital	Academic /Research institution	(1998) 2001 [h]	8.500 (5100)	ms-kompetansesenter.no
Poland	Polish MS Registry (RejSM)	Konskie and AGH University of Science and Technology, Krakow	Academic /Research institution	2011	8.845	rejsm.pl
Serbia	MS Society of Serbia	Clinic of Neurology, Faculty of Medicine, University of Belgrade and at the Clinical Centre of Serbia	Patient organisation, Academic /Research institution	1996	2.250[i](3.500)	–
Spain (Catalonia)	Epidemcat (MS Registry of Catalonia)	Department of Heath of the Government of Catalonia	Academic /Research institution	2008	1.520 (616)	epidemcat.cat/
Sweden	Swedish Neuro Registries – MS	Karolinska University Hospital, Stockholm	Health care organisation	1997	18.700 (12.900)	neuroreg.se
Switzerland	Swiss MS Cohort Study[j]	University Hospital of Basel	Academic /Research institution	2012	1.200 (270)	smsc.rodano.ch
United Kingdom	UK MS Register	Swansea University	Academic /Research institution	2009	16.000 (8.300)	ukmsregister.org

a) Home to MSBase.

Figure D1 : Tableau proposé par Glaser et al, en 2019 résumant les caractéristiques de 19 cohortes spécifiques SEP disponibles dans le monde (Glaser *et al.*, 2019)

Figure D1 : Tableau proposé par Glaser et al, en 2019 résumant les caractéristiques de 19 cohortes spécifiques SEP disponibles dans le monde (Glaser *et al.*, 2019)

E. Sharmin S et al, Communication affichée, ECTRIMS, 2019

Comparative effectiveness of natalizumab and fingolimod in different subgroups of patients with relapsing-remitting multiple sclerosis

Sifat Sharmin[1]; Mathilde Lefort[2]; Johanna Andersen[3]; Dana Horakova[4]; Eva Kubala Havrdova[4]; Raed Alroughani[5]; Guillermo Izquierdo[6]; Sara Eichau[6]; Serkan Ozakbas[7]; Helmut Butzkueven[8]; Francesco Patti[9]; Marco Onofrj[10]; Alessandra Lugaresi[11]; Murat Terzi[12]; Pierre Grammond[13]; Francois Grand'Maison[14]; Bassem Yamout[15]; Alexandre Prat[16]; Marc Girard[16]; Pierre Duquette[16]; Cavit Boz[17]; Maria Trojano[18]; Pamela McCombe[19]; Mark Slee[20]; Jeannette Lechner-Scott[21]; Recai Turkoglu[22]; Patrizia Sola[23]; Diana Ferraro[23]; Franco Granella[24]; Vahid Shaygannejad[25]; Julie Prevost[26]; Olga Skibina[27]; Claudio Solaro[28]; Rana Karabudak[29]; Bart Van Wijmeersch[30]; Tunde Csepany[31]; Daniele Spitaleri[32]; Steve Vucic[33]; Emmanuelle Leray[2]; Melinda Magyari[34]; Sandra Vukusic[35]; Tomas Kalincik[1] **on behalf of MSBase, Danish MS Registry, and OFSEP.**

[1]CORe, Department of Medicine, University of Melbourne, Melbourne, Australia; Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia; [2]EHESP School of Public Health, Department of Quantitative Methods in Public Health (MeTIS), Rennes, France; [3]Department of Neurology, Danish Multiple Sclerosis Center and the Danish Multiple Sclerosis Registry, Rigshospitalet and University of Copenhagen, Copenhagen, Denmark; [4]Department of Neurology and Center of Clinical Neuroscience, First Faculty of Medicine, Charles University in Prague and General University Hospital, Prague, Czech Republic; [5]Division of Neurology, Department of Medicine, Amiri Hospital, Sharq, Kuwait; [6]Hospital Universitario Virgen Macarena, Sevilla, Spain; [7]Dokuz Eylul University, Konak/Izmir, Turkey; [8]Central Clinical School, Monash University, Melbourne, Australia; Department of Neurology, The Alfred Hospital, Melbourne, Australia; Department of Neurology, Box Hill Hospital, Monash University, Melbourne, Australia; [9]GF Ingrassia Department, University of Catania, Catania, Italy; Policlinico G Rodolico; [10]Department of Neuroscience, Imaging, and Clinical Sciences, University G. d'Annunzio, Chieti, Italy; [11]Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy; IRCCS Istituto delle Scienze Neurologiche di Bologna; [12]Medical Faculty, 19 Mayis University, Samsun, Turkey; [13]CISSS Chaudiere-Appalache, Levis, Canada; [14]Neuro Rive-Sud, Quebec, Canada; [15]Nehme and Therese Tohme Multiple Sclerosis Center, American University of Beirut Medical Center, Beirut, Lebanon; [16]Hopital Notre Dame, Montreal, Canada; CHUM and Universite de Montreal, Montreal, Canada; [17]KTU Medical Faculty Farabi Hospital, Trabzon, Turkey; [18]Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari, Bari, Italy; [19]University of Queensland, Brisbane, Australia; Royal Brisbane and Women's Hospital; [20]Flinders University, Adelaide, Australia; [21]School of Medicine and Public Health, University Newcastle, Newcastle, Australia; Department of Neurology, John Hunter Hospital, Hunter New England Health, Newcastle, Australia; [22]Haydarpasa Numune Training and Research Hospital, Istanbul, Turkey; [23]Department of Neuroscience, Azienda Ospedaliera Universitaria, Modena, Italy; [24]Department of Medicine and Surgery, University of Parma, Parma, Italy; Department of Emergency and General Medicine, Parma University Hospital, Parma, Italy; [25]Isfahan University of Medical Sciences, Isfahan, Iran; [26]CSSS Saint-Jerome, Saint-Jerome, Canada; [27]Monash University, Melbourne, Australia; [28]Department of Neurology, ASL3 Genovese, Genova, Italy; Department of Rehabilitation, ML Novarese Hospital Moncrivello; [29]Hacettepe University, Ankara, Turkey; [30]Rehabilitation and MS-Centre Overpelt and Hasselt University, Hasselt, Belgium; [31]Department of Neurology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary; [32]Azienda Ospedaliera di Rilievo Nazionale San Giuseppe Moscati Avellino, Avellino, Italy; [33]Westmead Hospital, Sydney, Australia; [34]Department of Neurology, Danish Multiple Sclerosis Center and the Danish Multiple Sclerosis Registry, Rigshospitalet and University of Copenhagen, Copenhagen, Denmark; [35]Foundeation Eugene Devic EDMUS, Department of Neurology of Pierre Wertheimer, University Hospital of Lyon, Lyon, France

Introduction

- Natalizumab (NAT) is associated with superior control of disease activity than fingolimod (FTY) in patients with active multiple sclerosis (MS). [1-6]
- Relative effectiveness of these two treatments in different subgroups of patients remains to be explored.
- This study assessed the treatment benefit of NAT and FTY in subgroups of patients with relapsing-remitting MS (RRMS), stratified based on patient's demographic and clinical characteristics.

Methods

- 5148 RRMS patients, started NAT or FTY on/after January 1, 2011 and continued for a minimum of 3 months were identified in 3 observational databases: French Observatorie of Multiple Sclerosis, Danish Multiple Sclerosis Treatment Register, & the global MSBase registry.
- Exclusion criteria: participation in randomized controlled trials involving studied agents or treated with endoxan, cyclophosphamide, lemtrada, fingolimod, cladribine, mitoxantrone, natalizumab, rituximab, ocrelizumab, or alemtuzumab before baseline.
- Patients were censored at discontinuation or change of treatment or end of follow-up, whichever occurred first.
- Patients were stratified based on sex, baseline age (18-65 years, ≤median, >median), disease duration from symptom onset to start of treatment (≤median, >median), baseline expanded disability status scale (EDSS) score (<4, 4-5.5, <6, ≥6), disease activity in the year prior to study treatment initiation (progression without relapse, relapse with or without progression, none), and baseline hyperintense T2 lesion or gadolinium-enhancing T1 lesion status (0 new/active lesion, at least 1 new/active lesion).

Models

- Study outcomes were
  - number of relapses (analysed with a negative binomial regression model)
  - 6-month confirmed disability progression and regression events (Andersen-Gill model),

  recorded on study therapy.
- Proportional hazards assumption was assessed with Schoenfeld residuals and violation of the assumption was corrected by introducing an interaction term with a time variable.
- The models were weighted by Inverse Probability of Treatment Weights, this representing a propensity score for treatment allocation.
- Propensity score was generated using a multivariable generalized linear mixed model with treatment allocation as the outcome variable. Independent variables included sex, age and disease duration at baseline, number of relapses in the year preceding baseline, EDSS score at baseline, evidence of disease activity (relapse, progression of disability, or both), and number of previous disease modifying therapies (DMTs). Country was modelled as a random effect

Results

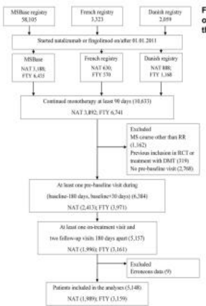


Figure 1: Flow diagram of patients included in the analyses.

Table: Baseline characteristics of patient subgroups.

	Sample size	Sex (% female)	Age, years	Disease duration [b], years	EDSS score [b]
Sex					
Male	1450	-	37 (9.97)[a]	6.23 (2.79-11.78)	2.5 (1.5-3.5)
Female	3698	-	38 (10.09)[a]	7.24 (3.27-12.86)	2.5 (1.5-3.5)
Age, years					
18-65	5055	72	-	7.05 (3.22-12.51)	2.5 (1.5-3.5)
≤38	2616	70	-	4.98 (2.25-8.81)	2 (1.5-3)
>38	2532	74	-	9.92 (4.89-16.19)	3 (2-4)
Disease duration, years					
≤7	2587	70	33 (27-41)[b]	-	2 (1.5-3)
>7	2561	74	41 (8.88)[a]	-	3 (2-4)
EDSS score					
<4	3940	72	36 (30-43)[b]	6.06 (2.69-11.02)	-
4-5.5	932	72	41 (9.70)[a]	10.12 (5.60-16.43)	-
<6	4872	72	37 (30-44)[b]	6.75 (3.03-12.21)	-
≥6	276	75	44 (10.69)[a]	11.04 (6.22-17.92)	-
MRI lesion					
No new lesion	177	75	39 (32.48-47.16)[b]	8.28 (4.52-13.59)	2.5 (1.5-4)
≥1 new hyperintense T_2 lesion or gadolinium-enhancing T_1 lesion	1357	72	35 (10.06)[a]	5.57 (2.32-10.11)	2 (1.5-3.5)
Pre-baseline MS activity					
Progression without relapse	419	70	39 (9.79)[a]	8.09 (4.34-13.86)	3 (2-4)
Relapse with or without progression	3291	71	37 (10.03)[a]	6.51 (2.73-12.07)	2.5 (1.5-3.5)
None	1438	74	40 (33-46)[b]	7.60 (3.76-13.24)	2 (1.5-3.5)

[a] Mean (standard deviation); [b] Median (quartiles)

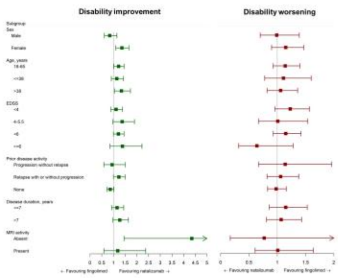


Figure 2: Hazard ratio and 95% confidence interval of 6-month confirmed disability improvement and worsening in different patient subgroups.

The chance of recovery from disability is greater on natalizumab than on fingolimod in:

- Females
- Age category 18-65 years, in particular >38 years
- Patients with EDSS score<6
- Patients who experienced relapses 12 months prior baseline
- Patients with no new T1 / T2 lesion identified in baseline MRI

No evidence of statistically significant associations between treatment allocation and disability worsening is observed within any of the subgroups.

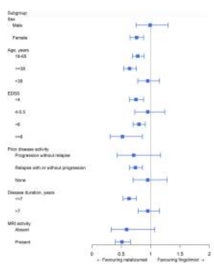


Figure 3: Incidence rate ratio and 95% confidence interval of on-treatment relapse in different patient subgroups.

Significantly lower on-treatment relapse rate on natalizumab than fingolimod, observed in:

- Females
- Age category 18-65 years, in particular <=38 years
- Patients with EDSS score<4, EDSS score<6, and EDSS score>=6
- Patients who experienced relapses 12 months prior baseline
- Patients with MS duration 7 years or less
- Patients with at least 1 new T1 / T2 lesion identified in baseline

Conclusion

- Natalizumab is associated with lower relapse frequency and higher likelihood of improvement in disability than fingolimod.
- The effectiveness of natalizumab tends to be higher particularly in patients with relapses 12 months prior to commencing therapy.
- Furthermore, females, patients of younger age (for the relapses outcome), and lower disability derive more benefit from natalizumab than fingolimod.
- In more progressive MS phenotype, the effectiveness of natalizumab is comparable to that of fingolimod.

References

1. Lorscheider J, Benkert P, Lienert C, et al. *Mult Scler* 2017. DOI: 10.1177/1352458518768433.
2. Prosperini L, Sacca F, Cordioli C, et al. *J Neurol* 2017; 264: 284–294.
3. Barbin L, Rousseau C, Jousset N, et al. *Neurology* 2016; 86: 771-778.
4. Baroncini D, Ghezzi A, Annovazzi PO, et al. *Mult Scler* 2016; 22: 1315–1326.
5. Kalincik T, Horakova D, Spelman T, et al. *Ann Neurol* 2015; 77: 425-435.
6. Gajofatto A, Bianchi MR, Deotto L, et al. *European neurology* 2014;72: 173-180.

Acknowledgements and Disclosures

This study was financially supported by NHMRC grant 1129789 and fellowship 1140766.

MSBase would like to acknowledge financial contributions to support the MSBase Registry from Biogen, Novartis, Merck, Roche, and Sanofi Genzyme. The Observatoire Français de la Sclérose en Plaques (OFSEP) is supported by a grant provided by the French State and handled by the "Agence Nationale de la Recherche", within the framework of the "Investments for the Future" program, under the reference ANR-10-COHO-002, by the Eugène Devic EDMUS Foundation against multiple sclerosis and by the ARSEP Foundation.

The authors have received research support, support to attend conferences, speaker honoraria and fees for participation at advisory boards from Actelion, Almirall, Bayer-Schering, Biogen, Celgene, EMD, Geneuro, Medday, Merck, Myaln, Novartis, Roche, Sanofi-Genzyme, and Teva, outside the submitted work.

F. Evaluation de l'équilibre de l'étude de l'efficacité de l'initiation précoce du traitement, critère de jugement EDSS 6

Figure F1 : Distribution des scores de propension relatifs aux critères de jugement EDSS 6 pour chaque définition de groupes

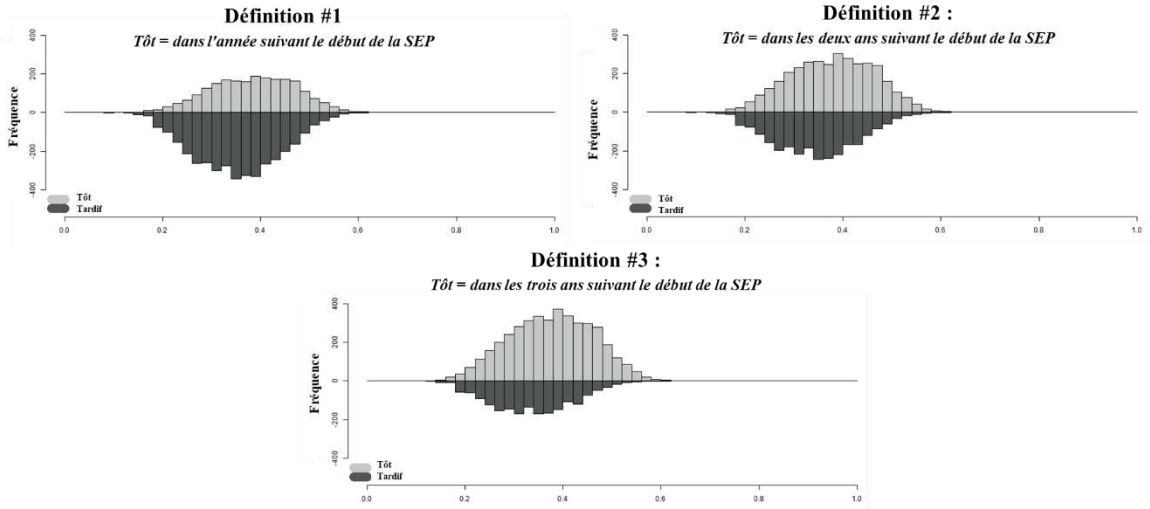
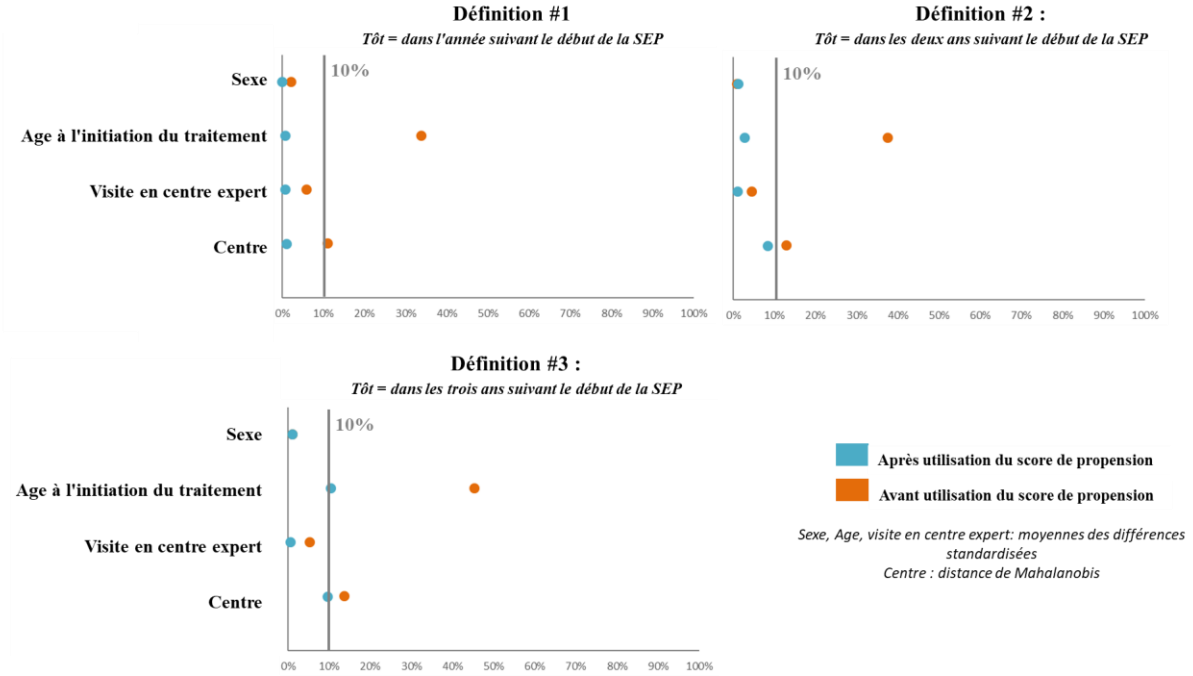


Figure F2 : Différences standardisées entre les groupes selon les trois définitions



G. Représentation de 1000 valeurs de Hazard Ratio obtenus après un appariement aléatoire

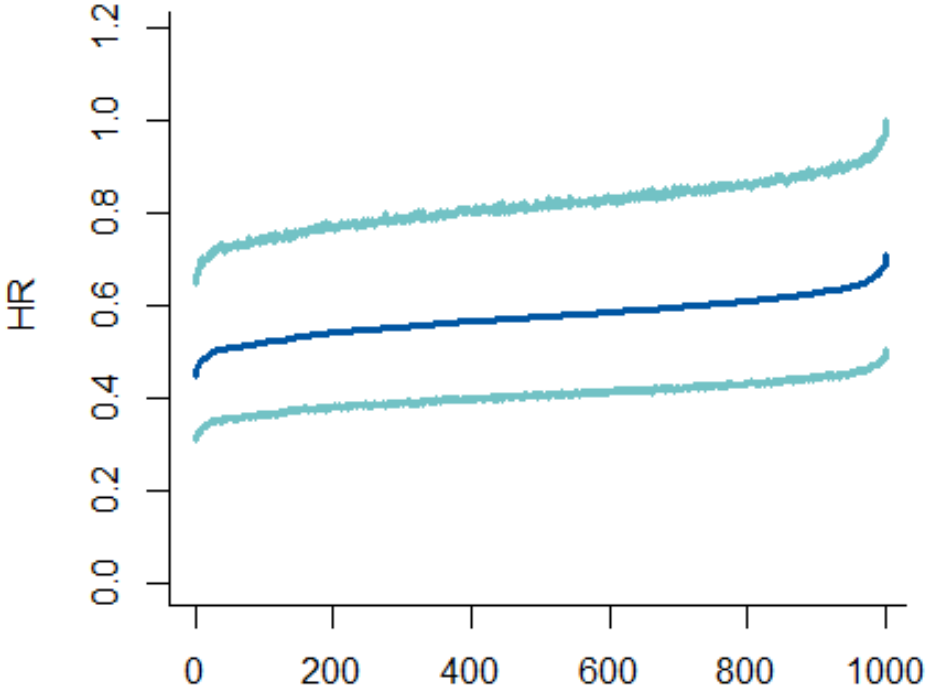
Commentaire :

Etude du délai d'atteinte d'un score EDSS 6 chez des patients atteints de sclérose en plaques rémittente traités par interférons-β ou acétate de glatiramère en comparaison à des patients non traités. Les patients étaient inclus à l'éligibilité aux traitements sur la période 1996-2015. Le modèle du score de propension incluait le sexe, l'âge, la durée de la SEP, le nombre de poussée et la période d'étude. Les HR ont été obtenu par des modèles de Cox.

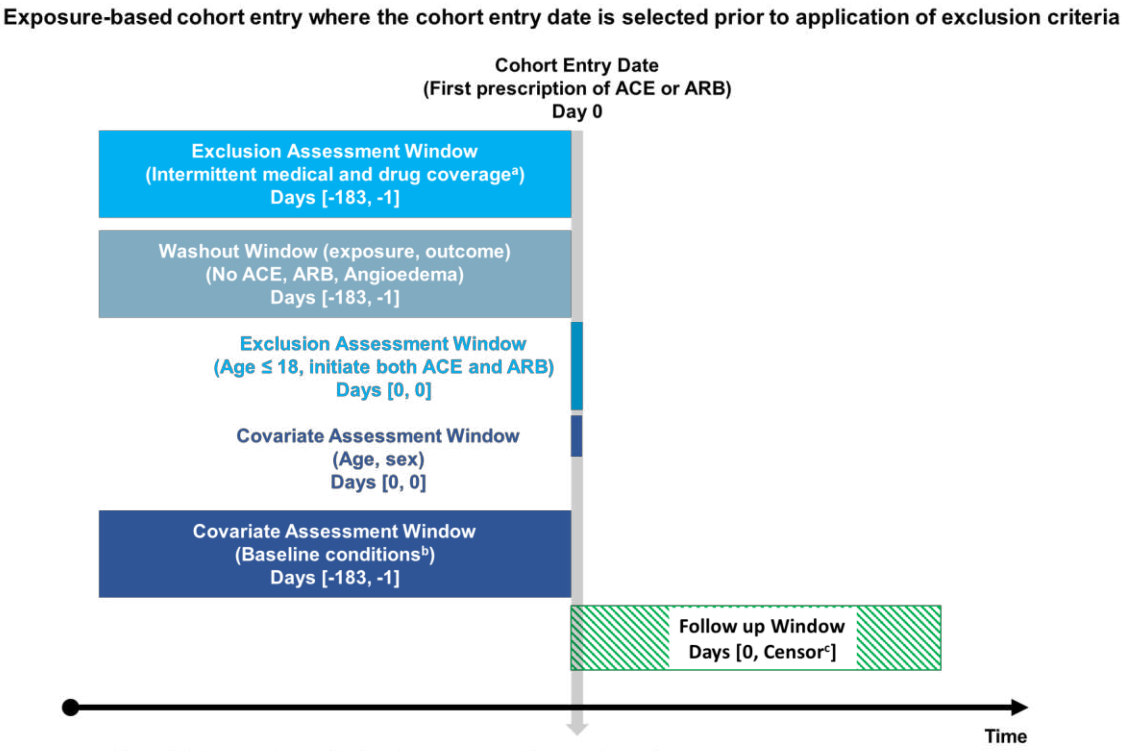
Légende :

Bleu foncé, valeur ponctuelle du hazard ratio

Bleu clair, Intervalle de confiance à 95% borne inférieure et supérieure



H. Exemple de graphique de conception d'une étude par REPEAT



a. Up to 45 day gaps in medical or pharmacy enrollment allowed
b. Baseline conditions included: allergic reactions, diabetes, heart failure, ischemic heart disease, non-steroidal anti-inflammatory drugs
c. Earliest of: outcome of interest (angioedema), switching or discontinuation of study drugs, death, disenrollment, 365 days of follow-up, end of the study period

ACE = angiotensin converting enzyme inhibitor
ARB = angiotensin II receptor blockers

UNIVERSITE BRETAGNE LOIRE / BIOLOGIE SANTE

UNIVERSITÉ DE RENNES 1

Titre : **Etude de l'effet des stratégies thérapeutiques en conditions réelles d'utilisation sur la progression du handicap chez les patients atteints de sclérose en plaques rémittente**

Mots clés : sclérose en plaques, traitement de fond, efficacité, données observationnelles, score de propension

Résumé : La sclérose en plaques (SEP) est une maladie neurologique chronique affectant le jeune adulte. Dans sa forme la plus fréquente, la forme rémittente, elle évolue par poussée. A long-terme, celles-ci peuvent laisser des séquelles et conduire à un handicap irréversible. Depuis le milieu des années 90, le répertoire des traitements de fond spécifiques de la SEP ne cesse d'évoluer faisant émerger différentes stratégies thérapeutiques, définies à partir des caractéristiques des patients et du profil bénéfice/risque des traitements. L'efficacité des traitements de fond en vie réelle a déjà été explorée mais a conduit à des résultats controversés. L'objectif principal de cette thèse est d'évaluer l'efficacité des traitements et des stratégies thérapeutiques en vie réelle sur la progression du handicap. L'objectif secondaire était d'explorer les méthodes de score de propension, utilisées dans le but de contrôler le biais d'indication.

Ce travail a été réalisé à partir des données de l'Observatoire Français de la Sclérose en Plaques (OFSEP), principalement collectées au cours des consultations en centre expert SEP. Ce travail a été organisé autour de trois axes. L'évaluation de l'effet des traitements de première ligne injectables a mené à utiliser une méthode basée sur un score de propension dépendant du temps, conduisant à ne pas montrer de bénéfice à long terme, résultat probablement altéré par de la confusion résiduelle. La comparaison de deux traitements de deuxième ligne, natalizumab et fingolimod, a conduit à confirmer la supériorité du natalizumab sur la progression du handicap et a offert l'opportunité d'explorer l'impact des méthodes utilisant les scores de propension sur les résultats. Enfin, les différentes stratégies thérapeutiques ont été décrites en France, des résultats préliminaires sur leur efficacité ont été proposés.

Title: **Studying effects of real-life therapeutic strategies on disability progression in patients with relapsing-onset multiple sclerosis**

Keywords-: multiple sclerosis, disease-modifying drug, effectiveness, observational data, propensity score

Abstract: Multiple sclerosis (MS) is a chronic neurological disease starting in early adulthood. Relapsing-onset MS is the most common phenotype and defined as occurrence of relapses, a relapse being defined as apparition or worsening of new symptoms. At long term, symptoms get worst and lead to irreversible disability. Since the mid-1990s, the arsenal of specific MS disease-modifying drugs has evolved, leading to different therapeutic strategies based on patient's characteristics and benefit/risk profile of treatment. The main objective of this PhD thesis was to assess the effectiveness of treatments and therapeutic strategies in real-life on disability progression. The secondary objective was to explore the propensity score methods used to control indication bias.

This work was based upon observational data from the "Observatoire Français de la Sclérose en plaques" (OFSEP), mainly collected at consultations in MS expert center. This work was organized around three axes. The evaluation of the effect of first-line injectable treatments led to the use of a method based on a time-dependent propensity score. No benefit was shown on long-term outcome, but this result was probably altered by residual confusion. Comparison of second-line treatments (natalizumab and fingolimod) confirmed the superiority of natalizumab on disability progression and provided the opportunity to explore the impact of methods using propensity scores on results. Finally, the different usual therapeutic strategies were described in France, and preliminary results on their effectiveness were proposed.