Roux, J., N. Le Meur, O. Grimaud & E. Leray (2016)

Care pathways of patients affected with multiple sclerosis in France from 2007 to 2013 using administrative databases and state sequence analysis

in G. Ritschard & M. Studer (eds), Proceedings of the International Conference on Sequence Analysis and Related Methods, Lausanne, June 8-10, 2016, pp 191-206.

Care pathways of patients affected with multiple sclerosis in France from 2007 to 2013 using administrative databases and state sequence analysis

Jonathan Roux, Nolwenn Le Meur, Olivier Grimaud and Emmanuelle Leray

Abstract In France there is a lack of accurate and up-to-date data on care pathways and care consumption of patients affected with multiple sclerosis (MS). The aim of this study was to describe care consumption of MS people in France and build a typology of their care pathways over the 2007-2013 period. To answer this issue, sequence analysis and agglomerative hierarchical clustering were used on a random sample of 1,000 patients, issued from French health-care databases, split according to the data available. The typologies were then described according to individual characteristics and care consumption. Two similar partitions, using indel costs fixed to 0.9 and transition-based substitution costs, were obtained : a five-cluster and a six-cluster typology for respectively complete and incomplete care pathways (i.e. less than seven years). Low care consumption was associated with older and less treated patients. Patients having a medium-low care consumption were younger and treated, whereas those having a high consumption seems to be older and have more comorbidities than the others. This pioneer study, using an innovative method in health field, gives a first overview of the care consumption of MS-affected people in France using objective and quantitative information. To go further, same work will be performed on the whole French population affected with MS and will include biological and medical imaging exams, and specialists' care to complete care pathways.

Keywords : State sequence analysis, Care pathways, Multiple sclerosis, Administrative databases

Jonathan Roux[1,2],Nolwenn Le Meur[1], Olivier Grimaud[1], Emmanuelle Leray[1,2]
[1] EHESP French School of Public Health, Sorbonne Paris Cité, Rennes, France
[2] CIC-P 1414, CHU de Rennes, Rennes, France

1 Introduction

The increase of life expectancy and medical progress are leading to a steady growth of the number of people affected with a chronic disease [Organisation Mondiale de la Santé (OMS), 2002]. In France, 15 million people are affected with at least one chronic disease, namely slightly more than 20% of the whole population [Ministère de la Santé, de la Jeunesse, des Sports et de la Vie associative, 2007]. They require large amounts of medical care, which lead to new constraints of long-term care for healthcare systems [Organisation Mondiale de la Santé (OMS), 2002]. In France, the Long Disease Duration (LDD) status permits people to have a 100% coverage of their healthcare. Thirty illnesses are apart of the list of LDD, including multiple sclerosis (MS).

MS is a chronic neurological disease affecting 2.3 million people around the world with a women:men sex-ratio varying from 2:1 to 2.5:1 [Browne et al., 2014, Trojano et al., 2012, MS International Federation, 2015]. In France, the estimated prevalence is 151.2 cases per 100 000 inhabitants on 31st December 2012, namely about 100 000 people [Foulon et al., 2015]. MS mostly starts at young adulthood, between 20 and 40 years old, evolves for several decades and frequently leads to a disability, the median duration before the need for a walking aid being from 20 to 25 years [Confavreux et al., 2000, Confavreux et al., 2003]. Moreover, because of a life expectancy reduced of about 7 years [Leray et al., 2015] and availability of more and more DMTs (Disease Modifying Therapies), MS cost becomes bigger and bigger over time for each patient and thus for the society and was equal to 1 billion euros in France for the year 2013 [Direction de la stratégie, des études et des statistiques (DSES), 2015, Lefeuvre et al., 2016].

Care consumption of MS patients is expected to be high due to disease length, disease activity and disability progression. However, there is a lack of accurate and up-to-date data on this topic in France. Indeed, French studies on care consumption mainly focus on MS costs from the point of view of the French national Health Insurance system [Fromont et al., 2014, Lefeuvre et al., 2016] and not on the type, amount and chronology of care. In addition, there are no clearly definite care's recommendations for patients affected with MS at the moment in France.

In order to fill this gap and have an overview of the care consumption of people with MS in France, the first step is to measure the consecutive consultations and treatments constituting the care pathway of each patient. Therefore, this study aimed at describing the overall care consumption of people affected with MS, analysing their care pathways and creating a typology over the 2007-2013 period in France.

2 Methods

2.1 Study population

Data was issued from a French permanent sample of health insurance care users, named *Echantillon Généraliste des Bénéficiaires (EGB)*. It is a random sample of $1/97^{th}$ of the French national Health Insurance system, gathering the out-hospital reimbursed care consumption (consultations and home visits to general practitioners (GP), specialists, drugs deliveries, nurses, physiotherapists, ...). This sample is dynamic, which means that new people can enter the sample anytime. Moreover, we used data from the French Hospital Discharge Database (*Programme de Médicalisation des Systèmes d'Information (PMSI)*), gathering in-hospital care consumption (admissions including day hospital, but except outpatient visits).

Patients affected with MS were identified if at least one of the following criteria was fulfilled in the period stretching from 1^{st} January 2007 to 31^{st} December 2012 : either MS LDD status, either at least one DMT specific of MS dispensed (interferon-β, glatiramer acetate, fingolimod or natalizumab), or at least one hospital admission with an ICD-10 (International Classification of Diseases, 10^{th} version) diagnosis code "G35" [World Health Organization, 2015]. Out-hospital care consumption and hospital admissions for these patients were then extracted from 2007 up to end 2013. As a whole, 1,003 patients were identified, of whom 3 were excluded due to a total lack of care consumption (MS-related and not) during the study period.

2.2 Data

Based on data of care consumption, monthly number of consultations or home visits with a GP, a private neurologist and a PM&R (Physical Medicine and Rehabilitation) physician were estimated for each of the 1,000 patients. Only MS-related hospital admissions were considered (main or related diagnosis equal to "G35", except DMT injection) and a parameter was created to count the monthly corresponding number for each individual. Nine DMT used for MS delivered out of the hospital (four interferon-β, glatiramer acetate, fingolimod, mycophenolate mofetil, methotrexate, azathioprine) and one in-hospital treatment (natalizumab) were also explored if applicable. The total duration of each treatment was estimated as the product of the number of boxes and the number of corresponding days. None of the two databases permitted to access consultations with public neurologists. Therefore we assumed that patients receiving a DMT, compulsorily prescribed by a neurologist, without having a visit with a private one, has one visit with a public neurologist (outpatient visit) at the treatment initiation and one every year as long as the treatment was ongoing.

A monthly then annual composite variable characterizing the individual care consumption was created by adding monthly (respectively annual) numbers of consultations with GPs, private and public neurologists, PM&R physicians and MS-related hospital admissions (except monthly DMT injections). Each yearly variable ranged widely and followed a Poisson distribution. Therefore, to have an overview of the care consumption, it was categorized in five groups according to the quartile distribution : no consumption ("0"), less than the first quartile ("$]0;Q1]$"), between the first quartile and the median included ("$]Q1;Q2]$"), between the median and the third quartile included ("$]Q2;Q3]$") and strictly superior to the third one ("$> Q3$"). The sequence of the seven annual categorized parameters constituted the care pathway of each patient during the 7-year period. As inclusion criteria offered the opportunity to identify prevalent cases of MS, patients were not at the same stage of the disease during the 2007-2013 period and thus the sequences were not left-aligned for people entering the study after 2007.

For patients who died during follow-up, the "death" status was attributed the year following the death's year, since only annual states were considered. Therefore only deaths over the 2007-2012 period were considered. Death was coded as NA state, creating care pathways of different lengths. This solution permitted to take into account the death of people without the creation of a specific group, which may have happened with a devoted state.

2.3 Statistical analysis

Two different typologies of care pathways were created using Optimal Matching (OM) analysis and then agglomerative hierarchical clustering using Ward's criterion. The first one was obtained using patients with complete care pathways (n=648), i.e. sequences beginning in 2007, whereas the second focused on incomplete sequences (n=352), corresponding to people entering the cohort later than 2007 (due to dynamic cohort). This separation was chosen to avoid the creation of a cluster devoted to patients entering the cohort with delay appearing when analysing the 1,000 sequences together.

No theoretical knowledge about costs was available, therefore substitution costs were estimated empirically using observed transition rates. To ensure Needleman-Wunsch algorithm could be applied to compute OM distance, metric properties were verified for substitution costs, especially triangle inequality [Studer and Ritschard, 2015]. Costs are one major parameter permitting to have an optimal matching and thus influence clustering results [Lesnard, 2010]. For our purpose, the aim was to compare care pathways by keeping the temporality of events and studying series of sub-sequences and thus long-term care consumption rather than succession of one-state events. Moreover, we did not want to focus on outstanding events but rather on the whole sequence. Therefore, we chose to fix indel costs to 0.9 in order to

approach the Levenshtein II distance and thus compare sequences according to the longest common subsequence [Lesnard, 2010].

An agglomerative hierarchical clustering analysis using Ward's criterion was then performed on the dissimilarity matrix to create homogeneous groups. Several parameters were used to assess the quality of the partition, namely weighted Average Silhouette Width (ASWw) [Kaufman and Rousseeuw, 1990, Studer, 2013], Hubert's C (HC) [Hubert and Levin, 1976], Hubert's Gamma (HG) [Hubert and Arabie, 1985] and Point Biserial Correlation (PBC) [Hennig and Liao, 2010, Milligan and Cooper, 1985]. The maximisation of ASWw, HG and PBC and the minimisation of HC permitted to set the optimal number of clusters. Once the clustering obtained, each group was characterized according to individual characteristics and care consumption.

All computational and statistical analyses were performed using R (Version 3.2.3) [R Core Team, 2015]. Sequence analysis and clustering were performed using the *TraMineR* library (Version 1.8-11), the *WeightedCluster* library (Version 1.2) and the *cluster* library (Version 2.0.3) [Gabadinho et al., 2011, Studer, 2013, Maechler et al., 2015].

3 Results

3.1 Population characteristics

According to the selection criteria, 1,000 care users were identified in the databases, whose characteristics are presented in Table 1. They were mostly women (71.1%) and had a median year of birth of 1963 (range 1914-1997), i.e. aged about 44 years at study start. Patients with a complete care pathway were significantly older than patients having a shorter one ($p<0.001$). As a whole, 577 patients (57.7%) were under the LDD status for MS at the beginning of follow-up, with a median LDD time of 6.8 years (range 0.0-29.2), and 261 (26.1%) acquired the status on the 2007-2013 period. 278 (27.8%) patients had at least another LDD different from MS. The median follow-up duration was 6.8 years (range 0.0-7.0) out of a maximum of 7 years. As a whole, 71 (7.1%) deaths were observed after a median follow-up duration of 3.9 years (range 0.0-6.7). About half the population (53.9%) received at least one DMT and was treated during 63.0% of the follow-up duration in median.

Table 1 Population characteristics according to the length of care pathways (N=1000)

	Complete pathway n=648	**Incomplete pathway n=352**	**Total N=1000**	**p-value**[a]
Number of women (%)	473 (73.0%)	238 (67.6%)	711 (71.1%)	0.086
Birth Year *	1961 (1914-1991)	1966 (1924-1997)	1963 (1914-1997)	<0.001
Number of LDD for MS [b] (%)	589 (90.9%)	249 (70.7%)	838 (83.8%)	<0.001
Duration since MS-LDD's beginning [b,c]* (year)	12.7 (0.0-34.6)	6.5 (0.2-33.8)	11.6 (0.0-34.6)	<0.001
At least another LDD [b,d] (%)	186 (28.7%)	92 (26.1%)	278 (27.8%)	0.429
Follow-up duration * (year)	6.9 (0.0-7.0)	3.5 (0.0-6.0)	6.8 (0.0-7.0)	<0.001
Number of deaths observed (%)	55 (8.5%)	16 (4.5%)	71 (7.1%)	0.029
At least one DMT prescription [e] (%)	368 (56.8%)	171 (48.6%)	539 (53.9%)	0.015
Part of the follow-up duration under treatment (%) [f]*	61.6 (3.3-100.0)	66.0 (7.6-100.0)	63.0 (3.3-100.0)	0.218
GPs [g]*	6.1 (0.0-50.0)	4.3 (0.0-48.0)	5.6 (0.0-50.0)	<0.001
Neurologists [h]*	6 (0-48)	2 (0-19)	4 (0-48)	<0.001
At least one MS-related hospital admission [i] (%)	307 (47.3%)	166 (47.2%)	473 (47.3%)	1.000

* Median (minimum-maximum). [a] P-value of the comparison of complete and incomplete pathways' groups using either Kruskal-Wallis test, Pearson's chi-squared test or Fisher's exact test if needed. [b] LDD : Long Disease Duration. [c] Calculated at the date of last information (31/12/2013 or date of death). [d] LDD other than MS. [e] Disease Modifying Therapy. [f] For patients having at least one DMT use. [g] Annualized number of consultations and home visits taken together per patient. [h] Total number of consultations, home and imputed outpatient visits taken together per patient over the 2007-2013 period. [i] Total MS-related hospital admissions per patient over the 2007-2013 period (except monthly DMT injections).

3.2 Complete care pathways' clustering

The partition of the 648 complete care pathways led to a typology of five clusters as presented in Figure 1, which conducted to the best quality partition's parameters with ASWw, HC, HG and PBC respectively equal to 0.238, 0.102, 0.765 and 0.551.

Among the five clusters, two are dominating, i.e. clusters 1.5 and 1.4, with respectively 188 (29.0%) and 180 (27.8%) patients. These two groups corresponded to people having an overall high care consumption according to the mean duration in each state (Table 2). The group 1.3 characterized by a medium consumption is formed of 128 (19.8%) patients. At the opposite, people categorized in group 1.2 (n=124 (19.1%)) had a low care consumption and the 28 (4.3%) in the cluster 1.1 had no consumption during almost half the follow-up period.

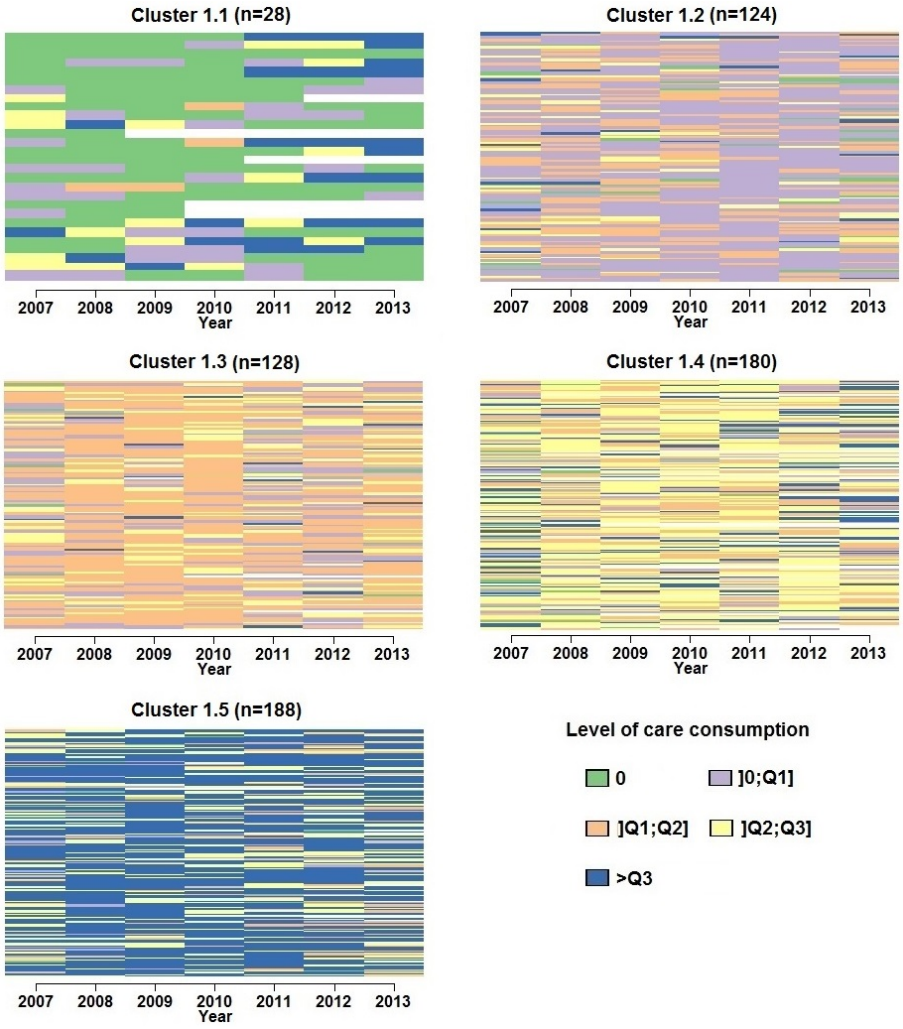

Fig. 1 Index plots of the groups obtained after the clustering procedure with indel costs fixed to 0.9 and transition-based substitution costs for complete care pathways (N=648)

Table 2 Mean duration in years (part of total duration) in each state according for complete care pathways' clustering (N=648)

State	”0”	”]0;Q1]”	”]Q1;Q2]”	”]Q2;Q3]”	”>Q3”
Group 1.1	**3.46 (49.4%)**	1.18 (16.9%)	0.14 (2.0%)	0.64 (9.1%)	0.93 (13.3%)
Group 1.2	0.31 (4.4%)	**4.06 (58.0%)**	1.86 (26.6%)	0.55 (7.9%)	0.19 (2.7%)
Group 1.3	0.09 (1.3%)	1.39 (19.9%)	**4.07 (58.1%)**	1.20 (17.1%)	0.16 (2.3%)
Group 1.4	0.13 (1.9%)	0.57 (8.1%)	1.76 (25.1%)	**2.95 (42.1%)**	0.95 (13.6%)
Group 1.5	0.10 (1.4%)	0.24 (3.4%)	0.45 (6.4%)	1.37 (19.6%)	**4.71 (67.3%)**

The predominant state for each cluster is indicated in bold.

3.3 Incomplete care pathways' clustering

According to quality parameters, the 352 incomplete care pathways were partitioned into six clusters presented in Figure 2. Indeed, ASWw, HG and PBC were maximised respectively with 0.238, 0.731 and 0.487, and HC was equal to 0.139.

The different groups had similar consumption to those obtained with complete care pathways, plus a sixth cluster (group 2.6), representing 26.4% of incomplete sequences, concerning patients mostly entering the cohort after 2010 and having a quite low-medium consumption (Table 3). Group 2.5 was formed of 52 patients (14.9%) having an overall high consumption and cluster 2.4 composed of 55 patients (15.6%) having a medium-high consumption. The cluster 2.3 characterized by a medium consumption was formed of 68 patients (19.3%). At the opposite, people categorized in group 2.2 (n=49 (13.9%)) had a low care consumption and the 35 (9.9%) in the cluster 2.1 had no consumption during almost three quarters of the follow-up period.

Table 3 Mean duration in years (part of total duration including NA state) in each state according for incomplete care pathways' clustering (N=352)

State	"0"	"]0;Q1]"	"]Q1;Q2]"	"]Q2;Q3]"	">Q3"
Group 2.1	**3.34 (73.6%)**	1.03 (22.7%)	0.14 (3.1%)	0.00 (0.0%)	0.03 (0.7%)
Group 2.2	0.41 (8.4%)	**3.31 (67.6%)**	0.59 (12.0%)	0.43 (8.8%)	0.06 (1.2%)
Group 2.3	0.29 (6.0%)	1.09 (22.6%)	**2.34 (48.5%)**	0.94 (19.5%)	0.16 (3.3%)
Group 2.4	0.20 (4.3%)	0.44 (9.5%)	0.95 (20.5%)	**2.36 (50.9%)**	0.67 (14.4%)
Group 2.5	0.35 (7.5%)	0.37 (8.0%)	0.35 (7.5%)	0.77 (16.6%)	**2.77 (59.6%)**
Group 2.6	0.22 (8.6%)	**0.82 (32.0%)**	**0.85 (33.2%)**	0.28 (10.9%)	0.30 (11.7%)

The predominant state for each cluster is indicated in bold.

3.4 Characteristics of the final clustering

The different patients' characteristics and care consumption according to the two partitioning are presented in Tables 4 & 5.

As expected, the number of consultations and home visits with GPs and neurologists increased from cluster 1.1 to 1.5 (Table 4). In group 1.1, the follow-up duration was the shortest one, which was probably related to the higher proportion of deaths observed over the study period ($p<0.001$). Moreover, the number of patients having used at least one DMT was lower in this cluster ($p=0.001$). Patients belonging to group 1.2 were significantly younger ($p<0.001$) and were the most treated for a median treatment duration of about three-quarters of follow-up. Furthermore, this cluster was almost composed of half men and half women, compared

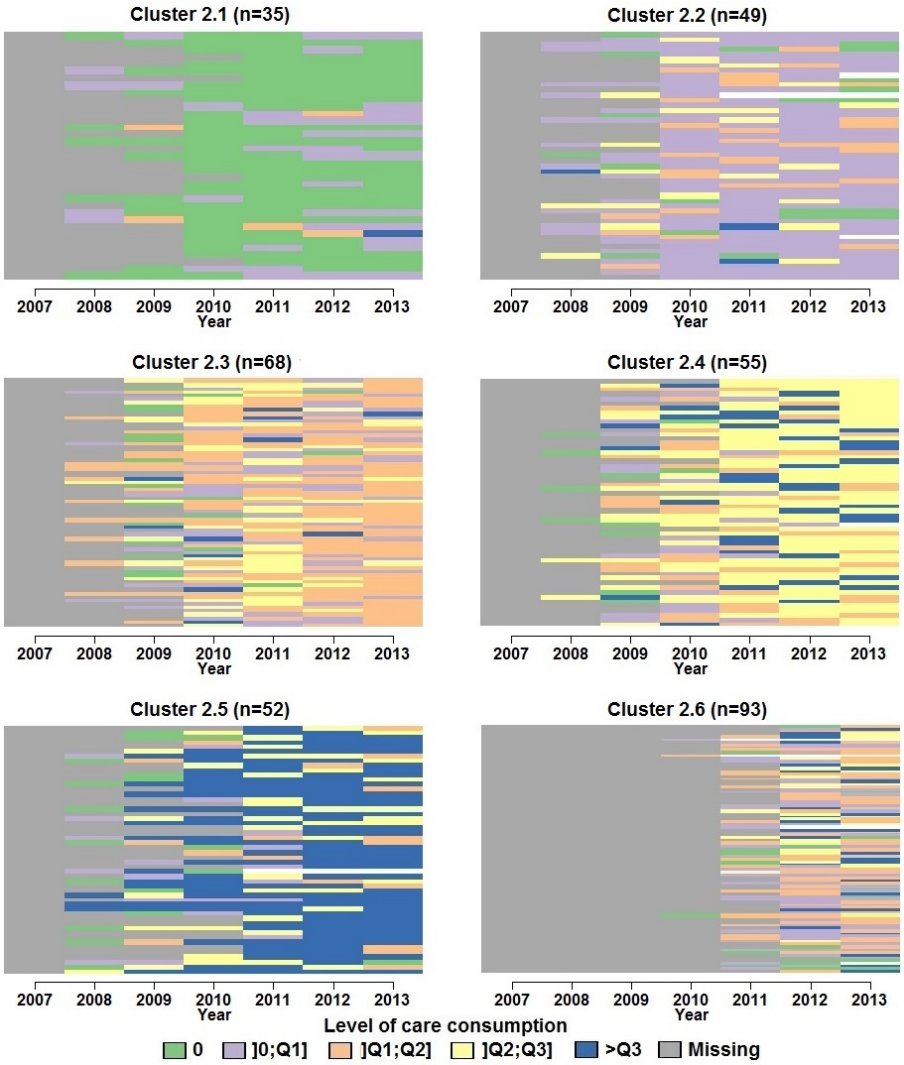

Fig. 2 Index plots of the groups obtained after the clustering procedure with indel costs fixed to 0.9 and transition-based substitution costs for incomplete care pathways (N=352)

to others mostly formed of women. The cluster 2.3 was also composed of patients using a DMT for the most part of follow-up duration. Group 1.4 was the oldest group, which could explain the percentage of deaths in this cluster. In group 1.5, i.e. having the highest consumption, patients seemed to be old with significantly more comorbidities, expressed as other LDD ($p<0.001$). These patients, who went more to hospital, were mostly treated.

As for the precedent typology, the care consumption with GPs and neurologists rose from cluster 2.1 to 2.5 in the second typology (Table 5). Patients in cluster 2.1 were the oldest and had the longest LDD duration compared to other groups, which could explain that they were significantly less treated ($p<0.001$). In groups 2.2 and 2.3, men and women were almost equally represented. Patients from cluster 2.2 were the youngest and those being the most treated. Like in the first typology, patients belonging to the highest consumption cluster 2.5 seemed to have more comorbidities than the others. Patients in cluster 2.6 were among the youngest ones with the most recent LDD status.

Table 4 Description of the complete care pathways' clustering ordered according to care consumption (N=648)

Cluster	**1.1 n=28 (4.3%)**	**1.2 n=124 (19.1%)**	**1.3 n=128 (19.8%)**	**1.4 n=180 (27.8%)**	**1.5 n=188 (29.0%)**	**p-value**[a]
Number of women (%)	23 (82.1%)	72 (58.1%)	92 (71.9%)	142 (78.9%)	144 (76.6%)	<0.001
Birth Year *	1958.5 (1930-1989)	1965.5 (1929-1991)	1960 (1923-1988)	1957.5 (1914-1986)	1959 (1916-1987)	<0.001
Number of LDD for MS [b] (%)	24 (85.7%)	110 (88.7%)	120 (93.8%)	167 (92.8%)	168 (89.4%)	0.373
Duration since MS-LDD's beginning [b,c]* (year)	12.4 (1.0-26.2)	12.6 (1.0-28.6)	13.0 (3.0-34.6)	12.1 (0.6-34.0)	12.8 (0.0-34.3)	0.629
At least another LDD [b,d] (%)	9 (32.1%)	19 (15.3%)	32 (25.0%)	48 (26.7%)	78 (41.5%)	<0.001
Follow-up duration * (year)	6.5 (0.0-7.0)	6.9 (0.0-7.0)	6.9 (0.0-7.0)	7.0 (0.0-7.0)	7.0 (0.0-7.0)	<0.001
Number of deaths observed (%)	6 (21.4%)	4 (3.2%)	4 (3.1%)	26 (14.4%)	15 (8.0%)	<0.001
At least one DMT prescription [e] (%)	6 (21.4%)	78 (62.9%)	74 (57.8%)	95 (52.8%)	115 (61.2%)	0.001
Part of the follow-up under treatment (%) [f]*	14.3 (3.6-59.0)	75.5 (3.3-100.0)	76.5 (4.4-100.0)	60.2 (3.3-100.0)	51.8 (4.4-100.0)	<0.001
GPs [g]*	2.7 (0.0-10.4)	2.6 (0.0-7.3)	4.6 (1.4-7.7)	6.9 (0.0-13.7)	12.7 (4.4-50.0)	<0.001
Neurologists [h]*	0 (0-8)	6 (0-23)	7 (0-35)	5 (0-38)	7 (0-48)	<0.001
At least one MS-related hospital admission [i] (%)	8 (28.6%)	61 (49.2%)	44 (34.3%)	84 (46.7%)	110 (58.5%)	<0.001

* Median (minimum-maximum). [a] P-value of the comparison of the five groups using either Kruskal-Wallis test, Pearson's chi-squared test or Fisher's exact test if needed. [b] LDD : Long Disease Duration. [c] Calculated at the date of last information (31/12/2013 or date of death). [d] LDD other than MS. [e] Disease Modifying Therapy. [f] For patients having at least one DMT use. [g] Annualized number of consultations and home visits taken together per patient. [h] Total number of consultations, home and imputed outpatient visits taken together per patient over the 2007-2013 period. [i] Total MS-related hospital admissions per patient over the 2007-2013 period (except monthly DMT injections).

Table 5 Description of the incomplete care pathways' clustering ordered according to care consumption (N=352)

Cluster	**2.1 n=35 (10.0%)**	**2.2 n=49 (13.9%)**	**2.3 n=68 (19.3%)**	**2.4 n=55 (15.6%)**	**2.5 n=52 (14.8%)**	**2.6 n=93 (26.4%)**	**p-value**[a]
Number of women (%)	23 (65.7%)	26 (53.1%)	39 (57.4%)	42 (76.4%)	38 (73.1%)	70 (75.3%)	0.022
Birth Year *	1960 (1930-1992)	1969 (1930-1997)	1966.5 (1931-1988)	1962 (1924-1989)	1964 (1924-1987)	1969 (1926-1994)	0.010
Number of LDD for MS [b] (%)	31 (88.6%)	32 (65.3%)	48 (70.6%)	39 (70.9%)	36 (69.2%)	63 (67.7%)	0.324
Duration since MS-LDD's beginning [b,c]* (year)	8.2 (0.2-33.0)	5.3 (0.3-14.7)	4.2 (1.5-33.2)	4.6 (1.6-30.8)	4.3 (2.0-27.0)	2.1 (1.0-33.8)	<0.001
At least another LDD [b,d] (%)	5 (14.3%)	10 (20.4%)	17 (25.0%)	17 (30.9%)	18 (34.6%)	25 (26.9%)	0.303
Follow-up duration * (year)	3.3 (0.0-5.6)	4.1 (0.0-5.9)	4.2 (0.0-5.9)	4.4 (0.0-5.9)	4.2 (0.0-6.0)	1.9 (0.0-3.0)	<0.001
Number of deaths observed (%)	0 (0.0%)	4 (8.2%)	2 (2.9%)	1 (1.8%)	3 (5.8%)	6 (6.5%)	0.399
At least one DMT prescription [e] (%)	4 (11.4%)	30 (61.2%)	35 (51.5%)	27 (49.1%)	27 (51.9%)	48 (51.6%)	<0.001
Part of the follow-up under treatment (%) [f]*	28.2 (15.4-44.4)	67.6 (15.8-100.0)	73.2 (18.4-98.3)	58.9 (10.2-97.8)	59.9 (7.6-98.1)	70.7 (11.2-95.6)	0.061
GPs [g]*	0.0 (0.0-4.5)	2.2 (0.0-6.3)	4.2 (1.2-7.5)	7.6 (2.4-12.7)	11.5 (2.8-48.0)	3.5 (0.0-24.0)	<0.001
Neurologists [h]*	0 (0-2)	4 (0-12)	4 (0-19)	4 (0-14)	4 (0-15)	2 (0-10)	<0.001
At least one MS-related hospital admission [i] (%)	10 (28.6%)	23 (46.9%)	33 (48.5%)	24 (43.6%)	34 (65.4%)	42 (45.2%)	0.031

* Median (minimum-maximum). [a] P-value of the comparison of the six groups using either Kruskal-Wallis test, Pearson's chi-squared test or Fisher's exact test if needed. [b] LDD : Long Disease Duration. [c] Calculated at the date of last information (31/12/2013 or date of death). [d] LDD other than MS. [e] Disease Modifying Therapy. [f] For patients having at least one DMT use. [g] Annualized number of consultations and home visits taken together per patient. [h] Total number of consultations, home and imputed outpatient visits taken together per patient over the 2007-2013 period. [i] Total MS-related hospital admissions per patient over the 2007-2013 period (except monthly DMT injections).

4 Discussion

The main aim of this pioneer study was to estimate the care consumption of the patients affected with MS over the 2007-2013 period in France. State sequence analysis using optimal matching was used to approach the concept of care pathway and build a typology of care consumption. Sequence analysis is usually employed in social sciences (transitions from education to work [Brzinsky-Fay, 2007, McVicar and Anyadike-Danes, 2002], career patterns [Anyadike-Danes and McVicar, 2005, Biemann and Datta, 2014], etc.) and genetics. Our research question on chronological events and type of data at our disposal offered the opportunity to apply this method in health field.

The study of care pathways reveals five groups of consumption for complete pathways and six for incomplete ones. In the two cases, clusters are mostly driven by a particular state dominating the sequences and show a gradation in care consumption. So, it seems that patients having a low care consumption are older, with a long LDD duration and less treated. Patients having a medium-low care consumption are younger and use DMTs, whereas those having a high consumption seems to be older and have more comorbidities than the others. It appears that the two partitions' clusters are similar on individual characteristics and care consumption. This result are observed despite the split realised on sequences' length and the last cluster of the second typology of people entering the cohort later than 2010, which could need to be studied separately. Therefore, it would be interesting to study more accurately the two partitions gathered together in order to have a five clusters' typology.

Our study has several advantages. Firstly, it is based on a representative random sample of the French population, reducing selection bias. Characteristics of our study population being close to those of MS patients in France [Foulon et al., 2015], it tends to consider that our population is quite representative of MS-affected people in France. Then, the two databases permit to access all reimbursed care consumption of patients, independently from self declaration and thus minimising memory bias. Moreover, it conducts to a care consumption as closely as possible to the true consumption of the study population (except concerning adherence to DMTs).

Concerning the statistical analysis, state sequence analysis is an innovative method in the field of care pathways to our knowledge. Furthermore, the choice to split and then analyse sequences according to their length avoids the creation of a cluster devoted to patients entering the cohort with delay. Similarly, the creation of a state devoted to the "death" status would have conducted to a cluster mainly formed of people dying during the follow-up.

However, some limitations can be mentioned. The first comes from the database itself, since people can get out from the EGB without any indication in the data if they leave one of the covered system, but this should concern only a very little number of people and not affect our results. Furthermore the database does not include outpatient visits with public neurologists, despite we know that MS expert

centres are mainly located in French university hospital. Therefore they were imputed, but the true consumption is certainly greater than the one assigned. Indeed, with the chosen scheme of imputation, we can't take into account adverse events at DMT initiation or follow-up care visits of untreated patients, but only the visits needed for yearly prescriptions. However these two limits can be overcome with the access to the French National Health Insurance Information System (*Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM)*) from which our data were randomly selected.

Concerning statistics, only few individual characteristics were available to describe the partitions, since administrative databases do not contain clinical data. Although a discrepancy analysis and a regression model were considered, they were not presented since the pseudo-R^2 calculated were too small (lower than 0.05). We are currently developing algorithms to estimate relevant clinical parameters such as motor disability, based on the rentals and purchases of mobility aids and on spasticity's drugs deliveries, or adherence to treatments, using parameters described by Hess et al. [Hess et al., 2006]. Comorbidities will also be deeper explored, following works on Canadian databases and recommendations from Marrie et al. [Marrie et al., 2014, Marrie et al., 2016].

Another limit of this study comes from the weakness of the quality of the two partitions obtained, which can not permit to assess a strong structure in the data. However, this frailty can come from the fact that patients were not at the same stage of the disease during the 2007-2013 period and that care consumption may depend on several factors, such as MS-relapses or DMTs for example. Moreover, we worked on calendar years and had no information on MS-onset of patients, which do not permit to left-align sequences. To overcome this problem, we are trying to identify incident MS cases with an algorithm based on the work of Marrie et al. [Marrie et al., 2010] and thus study the impact of MS-onset on care consumption.

Finally, this study permits to describe for the first time the care consumption of MS-affected people in France over the 2007-2013 period with objective and quantitative information. The two typologies based on care pathways and obtained through sequence analysis are close to each other and reveal different ways of consumption among patients. This work, only constituting a preliminary draft, is going to be continued on the whole French population affected with MS, thanks to French National Health Insurance Information System. It will aim at examining more accurately care pathways, by analysing outpatient visits, biological exams, medical imaging exams, paramedical and other specialists care amongst others. In parallel, it will be designed to exploit administrative databases in order to have more relevant parameters explaining care consumption, such as level of disability or adherence to treatments.

References

[Anyadike-Danes and McVicar, 2005] Anyadike-Danes, M. and McVicar, D. (2005). You'll never walk alone: Childhood influences and male career path clusters. *Labour Economics*, 12(4):511 – 530. European Association of Labour Economists 16th Annual Conference, Universidade Nova de Lisboa, Lisbon, 9th – 11th Sepetember, 2004.

[Biemann and Datta, 2014] Biemann, T. and Datta, D. K. (2014). Analyzing sequence data: Optimal matching in management research. *Organizational Research Methods*, 17(1):51–76.

[Browne et al., 2014] Browne, P., Chandraratna, D., Angood, C., Tremlett, H., Baker, C., Taylor, B. V., and Thompson, A. J. (2014). Atlas of multiple sclerosis 2013: A growing global problem with widespread inequity. *Neurology*, 83(11):1022–1024.

[Brzinsky-Fay, 2007] Brzinsky-Fay, C. (2007). Lost in transition? labour market entry sequences of school leavers in europe. *European Sociological Review*, 23(4):409–422.

[Confavreux et al., 2003] Confavreux, C., Vukusic, S., and Adeleine, P. (2003). Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. *Brain*, 126(4):770–782.

[Confavreux et al., 2000] Confavreux, C., Vukusic, S., Moreau, T., and Adeleine, P. (2000). Relapses and progression of disability in multiple sclerosis. *N. Engl. J. Med.*, 343(20):1430–1438.

[Direction de la stratégie, des études et des statistiques (DSES), 2015] Direction de la stratégie, des études et des statistiques (DSES) (2015). Personnes prises en charge pour sclérose en plaques (sep) en 2013. Fiche pathologie, Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS). Update on July 22nd, 2015.

[Foulon et al., 2015] Foulon, S., Weill, A., Maura, G., Dalichampt, M., Debouverie, M., and Moreau, T. (2015). Prévalence de la sclérose en plaques en france en 2012 et mortalité associée en 2013 à partir des données du sniiram-pmsi. *Revue d'Épidémiologie et de Santé Publique*, 63, Supplement 1:S17 – S18. XXVIIIe Congrès national Émois, Nancy, 26 et 27 mars 2015.

[Fromont et al., 2014] Fromont, A., Lehanneur, M.-N., Rollot, F., Weill, A., Clerc, L., Bonithon-Kopp, C., Binquet, C., and Moreau, T. (2014). Coûts de la sclérose en plaques en france. *Revue Neurologique*, 170(6-7):432–439.

[Gabadinho et al., 2011] Gabadinho, A., Ritschard, G., Müller, N. S., and Studer, M. (2011). Analyzing and visualizing state sequences in r with traminer. *Journal of Statistical Software*, 40(4):1–37.

[Hennig and Liao, 2010] Hennig, C. and Liao, T. (2010). Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification. Technical report, Department of Statistical Science, UCL, Department of Sociology, University of Illinois.

[Hess et al., 2006] Hess, L. M., Raebel, M. A., Conner, D. A., and Malone, D. C. (2006). Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. *The Annals of Pharmacotherapy*, 40(7-8):1280–88.

[Hubert and Arabie, 1985] Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classificationcation*, 2:193–218.

[Hubert and Levin, 1976] Hubert, L. and Levin, J. (1976). A general statistical framework for assesing categorical clustering in free recall. *Psychological Bulletin*, 83:1072–1080.

[Kaufman and Rousseeuw, 1990] Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley Series in Probability and Statistics. Wiley, New York.

[Lefeuvre et al., 2016] Lefeuvre, D., Rudant, J., Foulon, S., Weill, A., and Alla, F. (2016). Prise en charge des patients atteints de sclérose en plaques en france en 2013–consommations de soins et montants remboursés par l'assurance maladie. *Revue d'Épidémiologie et de Santé Publique*, 64, Supplement 1:S26 – S27.

[Leray et al., 2015] Leray, E., Vukusic, S., Debouverie, M., Clanet, M., Brochet, B., de Sèze, J., Zéphir, H., Defer, G., Lebrun-Frenay, C., Moreau, T., Clavelou, P., Pelletier, J., Berger, E., Cabre, P., Camdessanché, J. P., Kalson-Ray, S., Confavreux, C., and Edan, G. (2015). Excess mortality in patients with multiple sclerosis starts at 20 years from clinical onset: Data from a large-scale french observational study. *PLoS ONE*, 10(7):e0132033.

[Lesnard, 2010] Lesnard, L. (2010). Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns. *Sociological Methods & Research*, 38(3):389–419.

[Maechler et al., 2015] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2015). cluster: Cluster analysis basics and extensions. Technical report, R package version 2.0.3.

[Marrie et al., 2014] Marrie, R. A., Fisk, J. D., Stadnyk, K. J., Tremlett, H., Wolfson, C., Warren, S., Bhan, V., and Yu, B. N. (2014). Performance of administrative case definitions for comorbidity in multiple sclerosis in manitoba and nova scotia. *Chronic Diseases and Injuries in Canada*, 34(2-3):145–153.

[Marrie et al., 2016] Marrie, R. A., Miller, A., Sormani, M. P., Thompson, A., Waubant, E., Trojano, M., O'Connor, P., Fiest, K., Reider, N., Reingold, S., Cohen, J., and the attendees of the International Workshop on Comorbidity in Multiple Sclerosis, F. (2016). Recommendations for observational studies of comorbidity in multiple sclerosis. *Neurology*.

[Marrie et al., 2010] Marrie, R. A., Yu, N., Blanchard, J., Leung, S., and Elliott, L. (2010). The rising prevalence and changing age distribution of multiple sclerosis in manitoba. *Neurology*, 74(6):465–471.

[McVicar and Anyadike-Danes, 2002] McVicar, D. and Anyadike-Danes, M. (2002). Predicting successful and unsuccessful transitions from school to work by using sequence methods. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(2):317–334.

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

[Ministère de la Santé, de la Jeunesse, des Sports et de la Vie associative, 2007] Ministère de la Santé, de la Jeunesse, des Sports et de la Vie associative (2007). Plan 2007-2011 pour l'amélioration de la qualité de vie des personnes atteintes de maladies chroniques. Technical report, Ministère de la Santé, de la Jeunesse, des Sports et de la Vie associative.

[MS International Federation, 2015] MS International Federation (2015). Atlas of MS. http://www.msif.org/about-us/advocacy/atlas/. Last accessed on Jul 28, 2015.

[Organisation Mondiale de la Santé (OMS), 2002] Organisation Mondiale de la Santé (OMS) (2002). Des soins novateurs pour les affections chroniques : Eléments constitutifs : Rapport mondial. Technical report, OMS.

[R Core Team, 2015] R Core Team (2015). R: A language and environment for statistical computing. Technical report, R Core Team, Vienna, Austria.

[Studer, 2013] Studer, M. (2013). Weightedcluster library manual: A practical guide to creating typologies of trajectories in the social sciences with r. Technical report, LIVES Working Papers 24. DOI: 10.12682/lives.2296-1658.2013.24.

[Studer and Ritschard, 2015] Studer, M. and Ritschard, G. (2015). What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

[Trojano et al., 2012] Trojano, M., Lucchese, G., Graziano, G., Taylor, B. V., Simpson, Jr, S., Lepore, V., Grand'Maison, F., Duquette, P., Izquierdo, G., Grammond, P., Amato, M. P., Bergamaschi, R., Giuliani, G., Boz, C., Hupperts, R., Van Pesch, V., Lechner-Scott, J., Cristiano, E., Fiol, M., Oreja-Guevara, C., Saladino, M. L., Verheul, F., Slee, M., Paolicelli, D.and Tortorella, C., D'Onghia, M., Iaffaldano, P., Direnzo, V., Butzkueven, H., Group, M. S., and the New Zealand MS Prevalence Study Group (2012). Geographical variations in sex ratio trends over time in multiple sclerosis. *PLoS ONE*, 7(10):e48078.

[World Health Organization, 2015] World Health Organization (2015). ICD-10 Version:2016. Last accessed on Jun 22, 2015.