Neurol Ther (2022) 11:385–396
https://doi.org/10.1007/s40120-022-00324-0

ORIGINAL RESEARCH

A Principal Component Analysis Approach to Estimate the Disability Status for Patients with Multiple Sclerosis Using Japanese Claims Data

Izumi Kawachi · Hiromichi Otaka · Kosuke Iwasaki · Tomomi Takeshima · Kengo Ueda

Received: November 30, 2021 / Accepted: January 12, 2022 / Published online: January 22, 2022

ABSTRACT

Introduction: Claims databases are preferred for research on multiple sclerosis (MS) as this condition is characterized by low prevalence and long disease course. However, Japanese claims databases contain no information on disease severity or disability status of MS. Here, we aimed to explore the possibility of utilizing a principal component analysis (PCA) to estimate MS severity using a Japanese claims database.
Methods: An MS severity score was developed using a PCA. Factors related to functional systems for Expanded Disability Status Scale (EDSS) and higher disease severity (74 diagnoses, 68 drug prescriptions, and 77 procedures) were extracted from the claims database (April 2008–August 2018). The score (PC1 score) was developed for each patient-year—each year from the first diagnosis (excluding the year of the first diagnosis), based on the first principal component of the included factors. Finally, the patient-years were classified into quartiles based on the PC1 score, and demographic information and medical status were analyzed.
Results: The database contained 7067 patients with MS. The highest score group had a higher mean age (55.4 ± 0.2 [mean ± standard error] years), lower percentage of women (64.4 ± 0.7%), and longer mean disease duration from first diagnosis (8.1 ± 0.1 years) than the lowest score group (43.3 ± 0.2 years, 68.4 ± 0.8%, and 6.0 ± 0.1 years, respectively). In addition, the PC1 score of each patient positively correlated with disease duration from diagnosis.
Conclusion: We developed a PC1 score to indicate MS severity using information from a Japanese claims database. Since changes in demographic features we observed are consistent with findings of previous research, this score might represent MS severity to some extent. Further research is necessary to validate this score with clinical measurement of disability such as the EDSS.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40120-022-00324-0.

I. Kawachi
Department of Neurology, Brain Research Institute, Niigata University, 1-757 Asahimachidori, Chuo-ku, Niigata, Japan

I. Kawachi
Comprehensive Medical Education Center, Niigata University School of Medicine, 1-757 Asahimachidori, Chuo-ku, Niigata, Japan

H. Otaka (✉) · K. Ueda
Neuroscience Medical Franchise, Medical Division, Novartis Pharma K.K., 1-23-1 Toranomon, Minato-ku, Tokyo, Japan
e-mail: hiromichi.otaka@novartis.com

K. Iwasaki · T. Takeshima
Milliman Inc., 1-6-2 Kojimachi, Chiyoda-ku, Tokyo, Japan

Keywords: Claims database; Disability status; Multiple sclerosis; Principal component analysis; Real-world evidence; Secondary progressive multiple sclerosis; Unsupervised machine learning

Key Summary Points

Why carry out this study?

Claims databases are preferred for research on multiple sclerosis (MS) because of the low prevalence rate and long and heterogeneous disease course of this condition; however, Japanese claims databases contain no information on disease severity or disability status.

To explore the possibility of utilizing principal component analysis (PCA) to estimate the disease severity of MS, we developed a score through PCA using information from a Japanese claims database. We also analyzed the demographic information and medical status of patients with MS based on the score.

What was learned from the study?

The score-associated changes in demographic features are consistent with those reported by previous research. Our results are also in line with those of previous studies that have reported that patients with higher disease severity experience lower costs for MS treatment but higher costs for other healthcare treatments as well as higher frequency of hospital visits and hospitalization.

Considering the consistency of these results with those of previous studies, the score developed by PCA may represent MS severity to some extent; PCA may be exploited as a method to estimate MS severity using claims databases.

Further research is necessary to validate our score with established measurements of disease severity or disability burden, such as the Expanded Disability Status Scale.

INTRODUCTION

Claims databases have been widely used for research on multiple sclerosis (MS) to assess the current healthcare conditions and evaluate the effects of interventions. The utilization of claims databases allows researchers to include a large number of patients throughout the country and observe them over a long period of time. This is especially preferred with respect to research on MS, as the prevalence rate of this condition is low and patients usually experience a long disease course. However, Japanese claims databases do not include measures to evaluate disease severity or disability status, such as the Expanded Disability Status Scale (EDSS) score [1]. Although a national registry dedicated to intractable diseases such as MS is compiled by the Ministry of Health, Labour, and Welfare, the registry only consists of information concerning medical expense subsidies from applicants with intractable diseases, but does not include all patients diagnosed with MS. Moreover, only certain people (such as staff members in national and local governments and researchers of research project on intractable diseases) are allowed access to the registry. Consequently, research investigating the nature and efficacy of MS treatments according to disease severity or disability status has been insufficiently performed in Japan. Although several studies on Japanese patients with MS using claims databases have been reported to date [2–5], these only provide information on the MS population as a whole; none contains analyses grouped based on disease severity or disability status. However, if it were possible to estimate these parameters from claims databases, we could take full advantage of the data, which offer several benefits including a large sample size and long-term follow-up periods.

Unsupervised machine learning, such as principal component analysis (PCA), is considered a possible method to estimate disease severity from data that do not contain any severity information. PCA is commonly used to reduce the number of dimensions by transforming a set of correlated variables into a set of uncorrelated variables, the principal

component, so as to compress the datasets [6, 7]. This method was successfully used to develop a severity score of chronic obstructive pulmonary disease (COPD) by inputting a group of variables available in claims data that reflected aspects of COPD severity [8]. Therefore, we speculated that this method may enable us to estimate the disease severity of patients with MS as well.

Consequently, this study aimed to explore the possibility of utilizing PCA to estimate the disease severity of patients with MS using a Japanese claims database. Although claims databases do not contain details concerning disease severity, they include some closely related information, such as diagnoses associated with neurological dysfunction and prescriptions for symptomatic therapeutic drugs. Utilizing PCA, we developed a score to predict disease severity using this information. We also analyzed the characteristics and treatment status of patients with MS in the database based on this score in an attempt to assess the possibility of this approach.

METHODS

Study Design

This was a claims-based study to develop a score to estimate MS disease severity using a PCA based on patients' claims data. We analyzed treatment pattern, medical resource utilization, and healthcare costs based on the score and assessed the appropriateness of the score by confirming the consistency of the results with those reported by previous research.

Data Source and Settings

The data source was the claims database (April 2008–August 2018) provided by Medical Data Vision Co., Ltd. (MDV). This database contains clinical data on anonymized patients from acute care hospitals that have adopted the Diagnosis Procedure Combination fixed-payment reimbursement system (referred to as DPC hospitals hereafter) [9]. As of June 2018, the database consisted of data from approximately 20 million patients from 329 DPC hospitals (10% of all DPC hospitals).

The study period was between 2009 and 2018. The observation period of each patient corresponded to the period between the first and last record of any medical intervention for that patient. The first diagnosis of each disease was recorded as the "FromDate" in this database, regardless of the observation period. In this study, the earliest "FromDate" for MS was defined as the date of the first diagnosis of MS, which may lie outside the study period.

Patient Definition

We defined patients with MS using the following inclusion and exclusion criteria.

Inclusion Criteria

Patients who had at least one claim for an MS diagnosis, defined as G35 by the International Classification of Diseases, 10th revision (ICD-10) code [10], and met any of the following criteria:

(1) At least one hospitalization claim for an MS diagnosis;
(2) At least one outpatient claim for an MS diagnosis and at least one claim for disease-modifying therapies (DMTs) (defined by a generic name);
(3) At least one outpatient claim for an MS diagnosis and first diagnosis received before the observation period;
(4) At least three outpatient claims for an MS diagnosis.

Exclusion Criterion

At least one claim for a diagnosis of neuromyelitis optica spectrum disorder defined as G36 by the ICD-10 code.

Development of the Score by Principal Component Analysis

We defined MS disease severity as the PC1 score, the first principal component, derived by PCA of the data related to higher disease severity. In

general, a PCA produces n principal components, with the first having the minimum possible distance from all the data and explaining the greatest variance. Therefore, among the factors that reflect disease severity, the first principal component was considered to be associated with a higher disease severity.

We selected factors to apply to the analysis as follows. Diagnosis and drug prescription codes related to the functional systems of the EDSS were selected based on the Japanese MS treatment guidelines [11] and the advice of a medical expert (author). Additional claims such as physical rehabilitation fees, which are supposed to be associated with disease severity, were also included through discussion with the medical expert based on clinical experience. For each patient, the presence of claims with diagnoses (defined by the ICD-10 code), prescription of drugs (defined by generic name), and medical procedures (defined by the procedure code) were observed each year. The conceptual scheme representing the development of the score is shown in Fig. 1, and the selected factors are listed in Supplementary Material Fig. S1.

We developed a score for each year from the first diagnosis (excluding the year of the first diagnosis) of each patient during the study period based on the claims data; that is, each year of each patient was considered a datum for the analysis.

We calculated the PC1 score for each patient in each year from their first diagnosis (excluding the year of the first diagnosis). We examined the yearly change in the PC1 score using a random-effects model, expressed as follows:

$$\begin{aligned} \text{Score}(i,\, t) &\sim \beta_0 + \beta_1 \times\, t + \gamma(i) + \varepsilon(i,\, t) \\ \varepsilon(i,\, t) &\sim \text{Normal}\left(0,\, \sigma^2\right), \end{aligned}$$

where Score(i, t) represents the PC1 score for each patient (i) in the year from the first diagnosis (t), $\gamma(i)$ indicates the random effect, β_0 is the intercept, and $\varepsilon(i, t)$ represents the random error. We calculated the distribution of the coefficient of years from the first diagnosis (β_1). Patients included in this model had been observed for > 7 years following their first diagnosis, excluding the year of the first diagnosis.

Treatment Status and Healthcare Costs by the Score

Patient-years (excluding the year of the first diagnosis) were classified into quartile groups using the PC1 score. Treatment pattern, healthcare resource utilization, and healthcare costs were calculated for each group. Treatment pattern represents the frequency of treatments for MS (Supplementary Material Table S1) per patient per month (PPPM); healthcare resource utilization represents the frequency of clinical tests (Supplementary Material Table S2), hospitalizations, hospital visits, and relapses (PPPM); healthcare costs represent total costs and their breakdown for MS treatment, clinical tests, and others. Relapse was estimated based on the type of treatment (Supplementary Material Table S1), except for oral administration of prednisolone. The standard error was calculated for each value using the number of patient-years as the sample size.

We used SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) and Microsoft Excel 2016 (Microsoft, Redmond, WA, USA) for the analyses.

Ethical Approval

This study was approved by the Clinical Research Promotion Network Japan (CR-IRB-0094). The database includes data collected for secondary use and was provided after anonymization; therefore, informed consent was not required according to ethical guidelines in Japan [12]. Permission to access and use the data for this study was obtained from Medical Data Vision Co., Ltd.

RESULTS

Development of the Score

A total of 7067 patients with MS were extracted from the database according to the inclusion and exclusion criteria.

We obtained the eigenvector for each variable of the first principal component, as shown

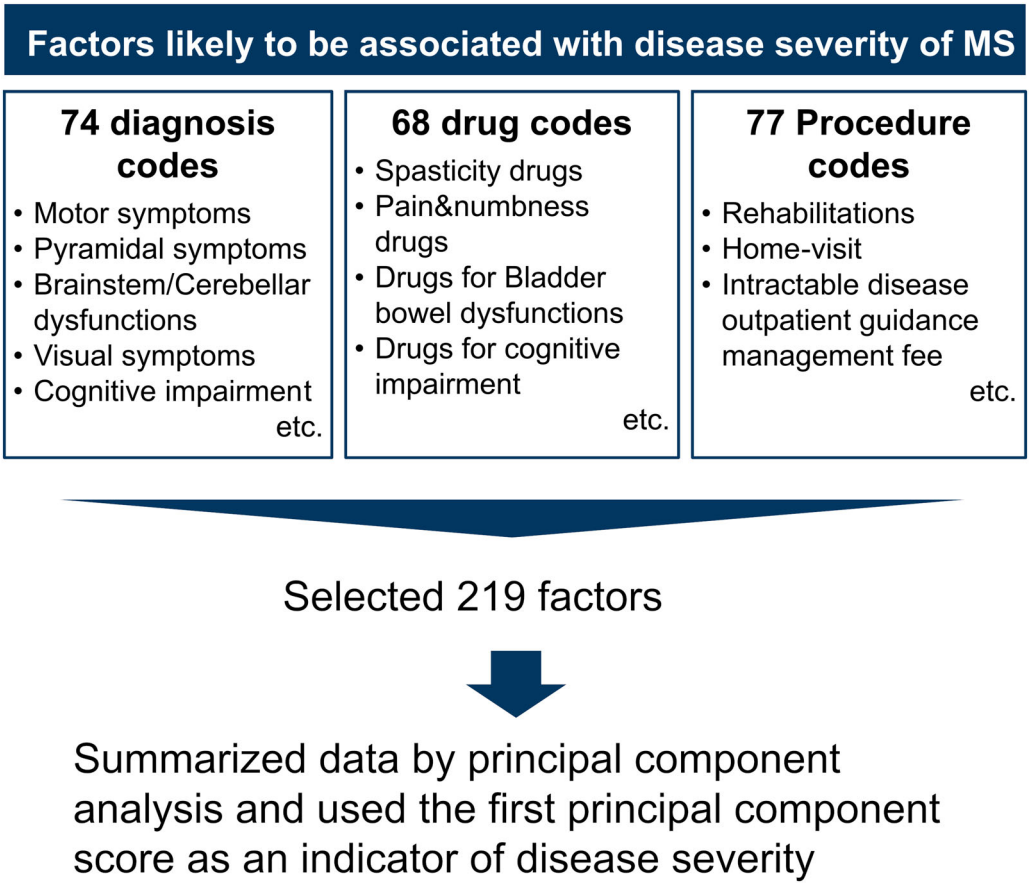

Fig. 1 Conceptual scheme illustrating the development of the disability score

in Supplementary Material Fig. S1. The scree plot of each principal component and explained variance are described in Supplementary Material Fig. S2. The first principal component explained 3.2% of the total variance.

The patients' demographic characteristics in each quartile based on the PC1 score (Q1–Q4, from lowest to highest, respectively) are shown in Fig. 2. There was a difference in the number of patient-years among quartile groups because patients were allocated according to the PC1 score. Regarding sex, 68.4 ± 0.8% (mean ± standard error) and 64.4 ± 0.7% of patients in the lowest and highest score groups were women, respectively. The average age was 43.3 ± 0.2 and 55.4 ± 0.2 years in the lowest and highest score groups, respectively. The years elapsed from the first diagnosis were approximately 6 for the lowest score group and 8.1 for the highest score group. The distribution of PC1 scores according to quartile groups is shown in Supplementary Material Fig. S3.

We examined the yearly change in the PC1 score from the first diagnosis based on the coefficient of years from the first diagnosis (β_1) using a random-effects model; the distribution is shown in Fig. 3. The mean of β_1 was 0.069 (95% confidence interval [CI] [0.062, 0.076]), meaning that the score increased with a statistical significance that was commensurate to the yearly increase from the first diagnosis for each patient.

Treatment Status and Healthcare Costs Based on the Score

Figure 2 describes the treatment pattern based on the PC1 score by each quartile. DMTs were most frequently prescribed for those in the

Level based on the PC1 score		Q1	Q2	Q3	Q4
Number of patient years		3,390	7,035	5,214	5,213
Number of patient months		31,909	65,064	50,615	53,254
Percentage of women		68.4± 0.8%	69.2± 0.6%	66.6± 0.7%	64.4± 0.7%
Age (year)		43.3± 0.2	48.1± 0.2	51.9± 0.2	55.4± 0.2
Duration from the first diagnosis of MS (year)		6.0± 0.1	5.9± 0.1	7.1± 0.1	8.1± 0.1
Disease-modifying therapy (PPPM)	Dimethyl fumarate	0.036± 0.003	0.011± 0.001	0.019± 0.002	0.038± 0.003
	Interferon (IFN)-ß1a	0.072± 0.005	0.048± 0.003	0.058± 0.003	0.052± 0.003
	IFN-ß1b	0.041± 0.003	0.042± 0.002	0.057± 0.003	0.067± 0.004
	Glatiramer acetate	0.015± 0.002	0.003± 0.001	0.007± 0.001	0.013± 0.002
	Fingolimod	0.128± 0.006	0.043± 0.002	0.081± 0.004	0.103± 0.004
	Natalizumab	0.009± 0.002	0.003± 0.001	0.007± 0.001	0.005± 0.001
Oral steroid (PPPM)	Prednisolone	0.075± 0.005	0.068± 0.003	0.154± 0.005	0.316± 0.008
Immunosuppressant (PPPM)	Azathioprine	0.003± 0.001	0.006± 0.001	0.021± 0.002	0.033± 0.003
	Tacrolimus	0.009± 0.002	0.006± 0.001	0.013± 0.002	0.028± 0.002
	Cyclosporine	0.002± 0.001	0.001± 0	0.006± 0.001	0.03± 0.002
	Mycophenolate mofetil	0± 0	0± 0	0.001± 0	0.006± 0.001
	Cyclophosphamide	0± 0	0± 0	0.001± 0	0.003± 0.001
	Methotrexate	0.002± 0.001	0.002± 0.001	0.008± 0.001	0.026± 0.002
	Mitoxantrone	0± 0	0± 0	0± 0	0± 0
Relapse treatment (PPPM)	Prednisolone (relapse therapy)	0.011± 0.002	0.012± 0.001	0.037± 0.003	0.131± 0.005
	Prednisolone sodium succinate	0± 0	0.001± 0	0.006± 0.001	0.027± 0.002
	Dexamethasone phosphate ester sodium	0± 0	0.003± 0.001	0.002± 0.001	0.019± 0.002
	Methylprednisolone sodium ester succinate	0.039± 0.003	0.024± 0.002	0.057± 0.003	0.118± 0.005
	Plasma purification therapy	0.001± 0	0± 0	0.001± 0	0.007± 0.001

Low High

◀**Fig. 2** Patient demographics and treatment patterns according to the PC1 score. Disability status was predicted using a principal component analysis. Patient-years were classified into four levels by their PC1 score (Q1–Q4: from lowest to highest). Each value for drugs represents the frequency per patient per month, which was calculated as the total number of frequencies in each group divided by patient months in each group. The values are shown with their respective SE. Patient-years were applied as the sample size to calculate the SE. Colors are assigned for each demographic item (percentage of women, age, or duration from the first diagnosis of MS) or for all drugs as per the color scale bar. *MS* multiple sclerosis, *PPPM* per patient per month, *PC1* first principal component, *SE* standard error

lowest score group. On the other hand, in the highest score group, prednisolone was most frequently prescribed at 0.316 ± 0.008.

Figure 4 indicates that the frequency of hospitalizations, relapses, and outpatient visits showed an increasing tendency with the rise in the PC1 score. The costs for MS treatment were higher in the lowest score group, while other healthcare costs were higher in the highest score group. Total healthcare costs were highest for the highest score group.

DISCUSSION

We estimated MS disease severity using PCA based on the information of Japanese claims data. We developed a score (PC1) based on the first principal component as an indicator of disease severity. When dividing patients into quartiles based on the score, patients with a higher score displayed a longer period since diagnosis, older age, and a higher male-to-female ratio.

We used only the first principal component to develop the score. The analysis is often repeated for the second, third, and more components to understand the data features. However, as we aimed to generate a score to predict the severity of the disease, we used only the first principal component, which was considered to mostly represent severity.

We found that the changes in demographic features based on the PC1 score are consistent with those outlines by previous reports [13, 14]. The observation that the male-to-female ratio was higher in the group with the greatest PC1 score is consistent with previous reports showing that male sex is a predictive factor of poor prognosis, and it is associated with a more progressive and severe outcome [15]. We also confirmed that the PC1 score of each patient positively correlated with the period since diagnosis. These results suggest that the PC1 score might represent the disease severity to some extent, as we hypothesized.

The higher healthcare costs (excluding MS treatment costs) and resource utilization (frequency of hospital visits and hospitalizations)

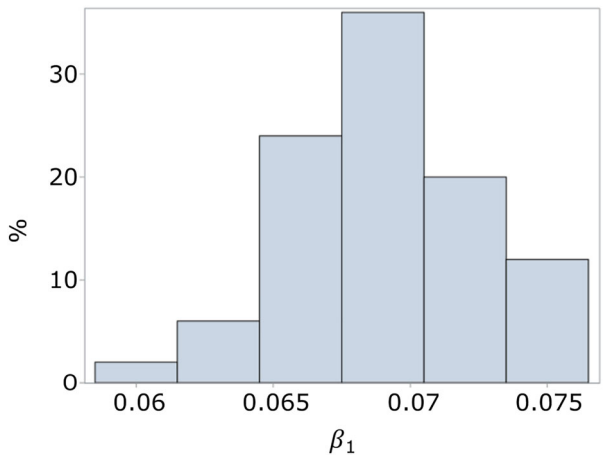

Fig. 3 Distribution of the coefficient of years from first diagnosis (β_1). Mean (95% confidence interval) = 0.069 [0.0618, 0.0756]; standard deviation = 0.004

Level based on the PC1 score		Q1	Q2	Q3	Q4
Number of patient years		3,390	7,035	5,214	5,213
Number of patient months		31,909	65,064	50,615	53,254
Clinical tests (PPPM)	Magnetic resonance imaging	0.164 ± 0.007	0.097 ± 0.004	0.121 ± 0.005	0.155 ± 0.005
	Clinical psychological or neuropsychological test	0 ± 0	0.001 ± 0	0.001 ± 0.001	0.003 ± 0.001
	Neurological test	0.001 ± 0.001	0.002 ± 0	0.001 ± 0.001	0.002 ± 0.001
	Ophthalmological test	0.306 ± 0.009	0.294 ± 0.006	0.313 ± 0.008	0.416 ± 0.009
	Cerebrospinal fluid test	0 ± 0	0 ± 0	0.005 ± 0.001	0.011 ± 0.001
	Visual evoked potential test	0.001 ± 0	0 ± 0	0.001 ± 0	0.002 ± 0.001
	Somatosensory evoked potential test	0 ± 0	0 ± 0	0.001 ± 0	0.005 ± 0.001
Hospitalizations (PPPM)		0.004 ± 0.001	0.005 ± 0.001	0.016 ± 0.002	0.066 ± 0.004
Relapses (PPPM)		0.011 ± 0.002	0.009 ± 0.001	0.019 ± 0.002	0.041 ± 0.003
Hospital visits (PPPM)		0.685 ± 0.014	0.601 ± 0.009	0.971 ± 0.014	1.292 ± 0.016
Healthcare costs (JPY, PPPM)	Total healthcare costs	102,053 ± 1,994	55,275 ± 1,156	84,403 ± 1,435	157,387 ± 2,895
	MS treatment costs	90,816 ± 1,938	43,216 ± 1,054	56,022 ± 1,248	47,740 ± 1,160
	Clinical test costs	2,071 ± 45	1,419 ± 28	1,671 ± 35	2,169 ± 45
	Other healthcare costs	9,165 ± 289	10,640 ± 400	26,711 ± 665	107,478 ± 2,704

Fig. 4 Healthcare resource utilization and healthcare costs according to the PC1 score. Principal component analysis was used to predict disability status. Patient-years were classified into four levels by their PC1 score (Q1–Q4: from lowest to highest). Each value represents the frequency per patient per month, which was calculated as the total number of frequencies divided by patient months in each group. The values are shown with their respective SE. Patient-years were applied as sample size to calculate the SE. Colors are assigned for each item group (clinical tests, hospitalization, relapse, hospital visit, or healthcare costs) as indicated in the color scale bar. *MS* multiple sclerosis, *JPY* Japanese yen, *PPPM* per patient per month, *PC1* first principal component, *SE* standard error

observed in patients with the highest PC1 score are in line with the findings of previous reports. In fact, it has been shown that patients with a higher disability status as assessed through the EDSS sustained higher total healthcare costs [16, 17], healthcare costs excluding MS treatment [18], and costs for inpatient care [16]. Although there may be differences in available treatments and healthcare systems among countries, a similar trend in costs and resource utilization was observed with respect to disease severity in previous studies performed in other countries. However, these differences may directly influence the treatment received by patients with MS, which, in turn, could affect the research and clinical applicability of the score in regions outside of Japan. For example, the greater use of prednisolone in patients with

the highest PC1 score observed in this study has instead hardly been observed in other countries. One possible reason for this is that there was no standardized treatment for high-severity secondary progressive MS in the observational period of our study, and some of these patients might have been treated with oral steroids despite the absence of clear evidence for their effectiveness. Another possible reason is that the cohort we analyzed may contain patients with atypical presentations of MS. Such presentations are diagnosed as MS but have features that are similar to those of other demyelinating diseases, such as neuromyelitis optica spectrum disorder and myelin oligodendrocyte glycoprotein antibody-associated disease, which leads to the prescription of oral corticosteroids even for patients diagnosed with MS. A recent multicenter retrospective study in Japan reported that 27% of patients with MS were treated with corticosteroids and/or immunosuppressants, and, among these, individuals with atypical MS were more frequently observed compared to those treated with DMTs [19]. It should be mentioned that we excluded prednisolone coded within 3 months of relapse treatment drug prescription; therefore, its frequent use cannot be explained by the follow-up treatment after pulse steroid therapy. Data obtained in the same country in the same period may be required to assess the validity of the score based on treatment status.

Our results showed an increase in the frequency of relapses with the rise in the PC1 score. A previous study has reported opposite findings, as a reduction in relapse rate was observed with the rise in EDSS levels [20]. One of the possible reasons for this discrepancy could be the definition of relapse we employed in this study. Since no information concerning relapse is present in the database, we defined relapse based on the treatments, including pulse steroid therapy and plasmapheresis. Another reason for this discrepancy was probably due to the temporary rise in disease severity observed after relapses. In fact, a higher PC1 score might be associated with greater severity due to higher disease activity after relapses, alongside accumulated disability. It should be noted that we excluded the data from the year of the first diagnosis when relapse rate is high, with the intent to reduce the effect of claims associated with symptoms that transiently appear because of relapse, and not because of accumulated disability. Regarding the group with low PC1 scores, a lower frequency of relapses and a relatively frequent use of DMT were observed, suggesting that the group might include individuals that better respond to DMT.

We reviewed previous studies that defined MS severity using information obtained from claims data. Although we found some relevant studies, their procedures are different from those that we used and cannot be directly applied to the Japanese claims databases because of the differences in code definitions and settings. Johnson et al. compared the relapse rate among different treatment patterns using claims data from MarketScan Commercial and Medicare Supplemental (USA) [21]. In their study, symptoms corresponding to the Kurtzke Functional System were selected based on the ICD, 9th revision (ICD-9), and the weighted sum of these measures served as an indicator of MS severity, which was used as a propensity score matching factor to evaluate relapse rates. Although the basic idea was similar to our procedure, the method to provide a summarized score of disease severity from the set of claims was different, and we found that many of the ICD-10 codes corresponding to those in the ICD-9 that were used in their study are not available or rarely seen in claims data in Japan (e.g., difficulty in walking). Thus, to appropriately estimate disease severity from claims databases, it is necessary to select claims suited for each database depending on each specific country.

There are several limitations to this study. The score was based on the first principal component derived by PCA, and the degree of uniqueness was confirmed by quartiles of the score. Based on the nature of the method used, the numerical value of the score is not a definitive indicator of the level of disease severity. Moreover, we defined relapse based on the records suggesting relapse treatment was being used. Therefore, treatments that we defined as relapse treatments performed for other purposes were also counted as relapses,

and relapses not accompanied by treatment were not counted. In addition, we defined disease duration as the period from the first diagnosis of MS since there is no exact information on its precise onset in the database. In the MDV database, the data source is limited to DPC hospitals, and there are no data on diagnoses and treatments given in other facilities. Furthermore, since the information is based on the records of diagnoses and treatments, any lack of records or inaccuracy in recording may be reflected in the study results. Finally, although we showed a possibility to estimate the severity using PCA, we did not confirm its validity compared with actual clinical measures. Further research using medical records or institutional registries should be performed to compare our score to relevant clinical measures such as the EDSS, Multiple Sclerosis Functional Composite [22], or Barthel Index function scale [23], with the intent to improve and validate the developed score.

CONCLUSIONS

We developed a score to represent MS disease severity using information obtained from a Japanese claims database by the PCA approach. The group with higher scores showed a longer disease duration since diagnosis, older age, and a higher male-to-female ratio. Moreover, patients in this group sustained lower costs for MS treatment but higher costs for other healthcare treatments and experienced a higher frequency of hospital visits and hospitalizations. As these results are mostly consistent with previous research, we considered that the score we developed may represent disease severity to some extent. Consequently, PCA may be exploited as a method to estimate MS severity, thus allowing researchers to leverage claims databases despite the possible lack of information on disease severity. Therefore, we believe this study took the first step toward fully utilizing claims databases for research on MS. Further research is necessary to validate the score developed by this method with actual disease severity or disability scores such as the EDSS.

ACKNOWLEDGEMENTS

Funding. Sponsorship for this study and Rapid Service Fee were funded by Novartis Pharma K.K.

Medical Writing, Editorial, and Other Assistance. Editorial assistance in the preparation of this article was provided by Elsevier Language Editing Service. Support for this assistance was funded by Novartis Pharma K.K.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Author Contributions. Conceptualization: Izumi Kawachi, Hiromichi Otaka, Kengo Ueda; Methodology: Izumi Kawachi, Hiromichi Otaka, Kosuke Iwasaki; Formal analysis and investigation: Hiromichi Otaka, Kosuke Iwasaki, Tomomi Takeshima; Writing—Original draft preparation: Tomomi Takeshima; Writing—Review and editing: Izumi Kawachi, Hiromichi Otaka, Kosuke Iwasaki, Kengo Ueda; Project administration: Hiromichi Otaka; Supervision: Kengo Ueda.

Prior Presentation. Parts of this work were presented at the 8th Joint ACTRIMS-ECTRIMS Meeting (MSVirtual 2020), September 11–13, 2020.

Disclosures. Izumi Kawachi reports funding for research, travel/speaker honoraria or serving on scientific advisory boards from Novartis Pharma K.K, Biogen, Bayer, Mitsubishi Tanabe Pharma, Astellas Pharma Inc., Bristol-Myers Squibb, Takeda Pharmaceutical Co., Japan Blood Product Organization, Chugai Pharmaceutical Co., and Alexion Pharmaceuticals, and research support from a JSPS KAKENHI Grant. Hiromichi Otaka and Kengo Ueda are employees of Novartis Pharma K.K. Kosuke Iwasaki and Tomomi Takeshima are employees of Milliman

Inc, which has received consultancy fees from Novartis Pharma K.K.

Compliance with Ethics Guidelines. This study was approved by the Clinical Research Promotion Network Japan (CR-IRB-0094). The database includes data collected for secondary use and was provided after anonymization; therefore, informed consent was not required according to the ethical guidelines of Japan. Permission to access and use the data for this study was obtained from Medical Data Vision Co., Ltd.

Data Availability. The data that support the findings of this study are available from Medical Data Vision Co., Ltd., but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Medical Data Vision Co., Ltd.

REFERENCES

1. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology. 1983;33:1444–52. https://doi.org/10.1212/wnl.33.11.1444.

2. Kawachi I, Okamoto S, Sakamoto M, et al. Recent transition of medical cost and relapse rate of multiple sclerosis in Japan based on analysis of a health insurance claims database. BMC Neurol. 2019;19:324. https://doi.org/10.1186/s12883-019-1534-9.

3. Ogino M, Kawachi I, Otake K, et al. Current treatment status and medical cost for multiple sclerosis based on analysis of a Japanese claims database. Clin Exp Neuroimmunol. 2016;7:158–67. https://doi.org/10.1111/cen3.12299.

4. Ogino M, Okamoto S, Ohta H, et al. Prevalence, treatments and medical cost of multiple sclerosis in Japan based on analysis of a health insurance claims database. Clin Exp Neuroimmunol. 2017;8:318–26. https://doi.org/10.1111/cen3.12411.

5. Ogino M, Shiozawa A, Ota H, Okamoto S, Hiroi S, Kawachi I. Treatment and comorbidities of multiple sclerosis in an employed population in Japan: analysis of health claims data. Neurodegener Dis Manag. 2018;8:97–103. https://doi.org/10.2217/nmt-2017-0047.

6. Sarker IH. Machine learning: algorithms, real-world applications and research directions. SN Comput Sci. 2021;2:160. https://doi.org/10.1007/s42979-021-00592-x.

7. IBM Cloud Education. Unsupervised learning [Internet]. Chicago (USA): IBM Cloud Learn Hub; 2020 [updated 2021 Sept 21; cited 2021 Oct 17]. https://www.ibm.com/cloud/learn/unsupervised-learning.

8. Wu EQ, Birnbaum HG, Cifaldi M, Kang Y, Mallet D, Colice G. Development of a COPD severity score. Curr Med Res Opin. 2006;22:1679–87. https://doi.org/10.1185/030079906X115621.

9. Ishii M. DRG/PPS and DPC/PDPS as prospective payment systems. Jpn Med Assoc J. 2012;55:279–91. https://doi.org/10.1007/s100720070018.

10. Ministry of Health, Labor and Welfare. Statistical classification of diseases and cause of death 2013 [Internet]. Tokyo (JP): Ministry of Health, Labor and Welfare; 2015 [cited 2020 Dec 23]. http://www.mhlw.go.jp/toukei/sippei/. Japanese.

11. Multiple Sclerosis and Neuromyelitis Optica Guideline Development Committee. Guidelines for

multiple sclerosis and neuromyelitis optica 2017]. Tokyo (JP): Igaku-shoin; 2017. Japanese.

12. Ministry of Education, Culture, Sports, Science and Technology; Ministry of Health, Labour and Welfare. Ethical guidelines for medical and health research involving human subjects [Internet]. Tokyo (JP): Ministry of Education, Culture, Sports, Science and Technology; Ministry of Health, Labour and Welfare; 2014 [updated 2017; cited 2020 Dec 23]. https://www.mhlw.go.jp/file/06-Seisakujouhou-10600000-Daijinkanboukouseikagakuka/0000153339.pdf. Japanese.

13. Confavreux C, Vukusic S. Age at disability milestones in multiple sclerosis. Brain. 2006;129:595–605. https://doi.org/10.1093/brain/awh714.

14. Ebers GC. The natural history of multiple sclerosis. Neurol Sci. 2000;21:S815–7. https://doi.org/10.1007/s100720070018.

15. Vukusic S, Confavreux C. Prognostic factors for progression of disability in the secondary progressive phase of multiple sclerosis. J Neurol Sci. 2003;206:135–7. https://doi.org/10.1016/s0022-510x(02)00426-4.

16. Palmer AJ, Colman S, O'Leary B, Taylor BV, Simmons RD. The economic impact of multiple sclerosis in Australia in 2010. Mult Scler. 2013;19:1640–6. https://doi.org/10.1177/1352458513488230.

17. Ruutiainen J, Viita AM, Hahl J, Sundell J, Nissinen H. Burden of illness in multiple sclerosis (DEFENSE) study: the costs and quality-of-life of Finnish patients with multiple sclerosis. J Med Econ. 2016;19:21–33. https://doi.org/10.3111/13696998.2015.1086362.

18. Karampampa K, Gustavsson A, van Munster ET, et al. Treatment experience, burden, and unmet needs (TRIBUNE) in Multiple Sclerosis study: the costs and utilities of MS patients in The Netherlands. J Med Econ. 2013;16:939–50. https://doi.org/10.3111/13696998.2013.807267.

19. Ashida S, Ochi H, Hamatani M, et al. Radiological and laboratory features of multiple sclerosis patients with immunosuppressive therapy: a multicenter retrospective study in Japan. Front Neurol. 2021;12:749406. https://doi.org/10.3389/fneur.2021.749406.

20. Melendez-Torres GJ, Auguste P, Armoiry X, et al. Clinical effectiveness and cost-effectiveness of beta-interferon and glatiramer acetate for treating multiple sclerosis: systematic review and economic evaluation. Health Technol Assess. 2017;21:1–352. https://doi.org/10.3310/hta21520.

21. Johnson BH, Bonafede MM, Watson C. Platform therapy compared with natalizumab for multiple sclerosis: relapse rates and time to relapse among propensity score-matched US patients. CNS Drugs. 2015;29:503–10. https://doi.org/10.1007/s40263-015-0251-1.

22. Tiftikçioğlu BI. Multiple Sclerosis Functional Composite (MSFC): scoring instructions. Noro Psikiyatr Ars. 2018;55(Suppl 1):S46–8. https://doi.org/10.29399/npa.23330.

23. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J. 1965;14:61–5.