

RESEARCH

Open Access



Classification of long-term clinical course of Parkinson's disease using clustering algorithms on social support registry database

Dougho Park^{1,2}, Su Yun Lee³, Jong Hun Kim^{4*} and Hyoung Seop Kim^{5*}

*Correspondence:

Jong Hun Kim

jh7521@naver.com

Hyoung Seop Kim
rehappydoc@gmail.com

¹Medical Research Institute, Pohang Stroke and Spine Hospital, Pohang, Republic of Korea

²Department of Medical Science and Engineering, School of Convergence Science and Technology, Pohang University of Science and Technology, Pohang, Republic of Korea

³Department of Neurology, Pohang Stroke and Spine Hospital, Pohang, Republic of Korea

⁴Department of Neurology, National Health Insurance Service Ilsan Hospital, 100 Ilsan-ro, Goyang 10444, Republic of Korea

⁵Department of Physical Medicine and Rehabilitation, National Health Insurance Service Ilsan Hospital, 100 Ilsan-ro, Goyang 10444, Republic of Korea

Abstract

Although Parkinson's disease (PD) has a heterogeneous disease course, it remains challenging to establish subtypes. We described and clustered the natural course of Parkinson's disease (PD) with respect to functional disability and mortality. This retrospective cohort study utilized the Korean National Health Insurance Service database, which contains the social support registry database for patients with PD. We extracted patients newly diagnosed with PD in 2009 and followed them up until the end of 2018. Functional disability was assessed based on the long-term care insurance (LTCI) and National Disability Registry data. Further, we measured all-cause mortality during the observation period. We included 2944 eligible patients. The surviving patients were followed up for 113.7 ± 3.3 months. Among the patients who died, patients with and without disability registration were followed up for 61.4 ± 30.1 and 43.2 ± 32.0 months, respectively. The cumulative survival rate was highest in cluster 1 and decreased from Cluster 1 to Cluster 6. In the multivariable Cox regression analysis, the defined clusters were significantly associated with increased long-term mortality (adjusted hazard ratio [aHR], 3.440; 95% confidence interval [CI], 3.233–3.659; $p < 0.001$). Further, age (aHR, 1.038; 95% CI, 1.031–1.045; $p < 0.001$), diabetes (aHR, 1.146; 95% CI, 1.037–1.267; $p = 0.007$), and chronic kidney disease (aHR, 1.382; 95% CI, 1.080–1.768; $p = 0.010$) were identified as independent risk factors for increased risk of long-term mortality. Contrastingly, the female gender (aHR, 0.753; 95% CI, 0.681–0.833; $p < 0.001$) and a higher LTCI grade (aHR, 0.995; 95% CI, 0.992–0.997; $p < 0.001$) were associated with a significantly decreased long-term mortality risk. We identified six clinical course clusters for PD using longitudinal data regarding the social support registry and mortality. Our results suggest that PD progression is heterogeneous in terms of disability and mortality.

Keywords Cluster analysis, Disability evaluation, Mortality, Parkinson disease

Introduction

Parkinson's disease (PD), which is the second most common neurodegenerative disease after dementia, is caused by damage to dopamine-secreting neuronal cells in the substantia nigra [1]. Dopamine acts on the striatum and is mainly responsible for regulating

body motor function [2]. Accordingly, dopamine deficiency causes motor symptoms such as tremors, gait disturbance, muscle spasms, and slow gait; further, it causes non-motor symptoms such as autonomic nervous system disorders, depression, and cognitive decline [3].

Symptoms in patients with early-stage PD can be sufficiently improved by a small dose of drug treatment [4]. However, as the disease progresses, the required drug dose and administration frequency are increased [5]; moreover, new symptoms, including dyskinesia and freezing of gait, appear and functional disabilities worsen [6, 7]. Additionally, non-motor symptoms such as sensory change and autonomic nervous system dysregulation significantly contribute to functional limitations [8]. Although PD is not fatal enough to shorten life expectancy [9], complications such as falls, swallowing disorders, and gait restrictions that with disease progression are strongly related to long-term prognosis and mortality [10, 11]. Taken together, PD requires long-term management plans given the symptom changes with disease progression.

PD has heterogeneous disease courses, and thus can be classified into multiple subtypes [12, 13]. Previous studies have defined PD subtypes in terms of symptom domains and progression speed, with the primary subtypes being the mild-motor predominant, intermediate, and diffuse malignant subtypes [14, 15]. Armstrong et al. [16] reported that the mild motor predominant type was the most common and showed slow progression, while the diffuse malignant type was observed in 9–16% of patients. Macleod et al. [9] reported that the average post-diagnosis survival period was 6.9–14.3 years, which considerably varied across patients. Mestre et al. [17] indicated the challenges of PD subtyping and the need to elucidate PD heterogeneity.

To understand the natural clinical course of PD, we explored the long-term outcomes (functional disabilities and mortality) of PD as well as the relationship of demographic features, comorbidities, and functional disabilities with long-term mortality in patients with PD using South Korea's National Health Insurance Service (NHIS) database.

Materials and methods

Data source and patient inclusion

This retrospective, longitudinal cohort study was conducted using customized cohort data from the Korean NHIS database (NHIS-2020-1-160) [18]. This study was reviewed and approved by the Institutional Review Board (NHIS-2023-02-002), which waived the requirement for informed consent given the retrospective study design and anonymity of the NHIS data. This study was conducted in compliance with the principles of the Declaration of Helsinki.

The initial sample comprised 31,167 patients with the International Classification of Diseases (ICD)-10th code G20 as the primary diagnosis. The study cohort comprised patients prescribed related drugs along with the G20 diagnosis code at medical institutions of general hospital level or higher. The drugs used for PD included levodopa, dopamine agonists (ropinirole, pramipexole, etc.), entacapone, amantadine, selegiline, rasagiline, and anticholinergics (trihexyphenidyl HCl, benzotropine mesylate, and procyclidine). In South Korea, patients diagnosed with PD using the G20 code can be registered in the system of 'rare and intractable diseases'. Subsequently, they can receive support from the government for a significant portion of medical expenses related to the

diagnosis. The criteria for registration of PD as a 'rare and intractable disease' are presented in Supplementary Document 1.

Among the initial cohort, we extracted 3227 patients who were first diagnosed with PD in 2009. Subsequently, we excluded patients with previously registered disabilities due to brain lesions, patients with missing values, and patients aged <40 years. Finally, we included 2944 patients with newly diagnosed PD in 2009, who were followed up for approximately 10 years (Fig. 1).

Variable definitions

Regarding sociodemographic variables, the sex and age of the patients were confirmed. Patient insurance was classified into medical aid and health insurance service types; moreover, health insurance services were further classified as self-employed and employee-insured types. The contribution of the self-employed insured type is

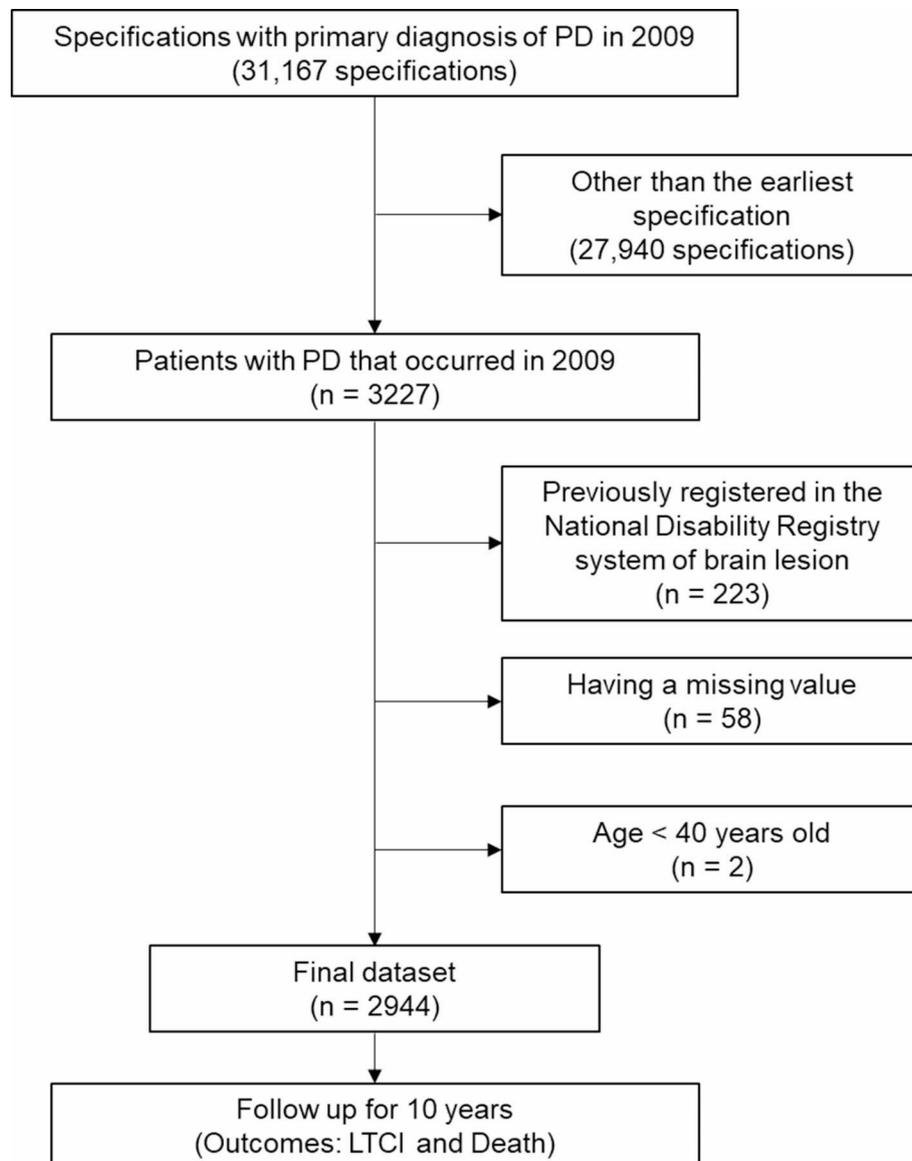


Fig. 1 Flowchart of patient inclusion. Abbreviations: LTCL, long-term care insurance; PD, Parkinson's disease

calculated as the contribution score \times value per score (Korean won). The contribution score is determined by considering the subscriber's income, property, economic activity participation rate, and sex and age of household members. Contrastingly, the contribution of the employee insured type is calculated as the monthly wage \times contribution rate, with the subscriber paying 50% and the employer paying 50% of the insurance premium [19]. Accordingly, we used the national health insurance premium level, which is an indicator of household income level, as a proxy for socioeconomic status and classified patients into four quartile groups. Residential areas were categorized as capital, metropolitan, city, and county. Comorbidities included hypertension (I10–I15), diabetes (E10–E14), dyslipidemia (E78), ischemic heart disease (I25), atrial flutter/fibrillation (I48), chronic kidney disease (N18), cerebral stroke (I60–64), and neoplasm (C00–D49). Finally, we confirmed all-cause mortality from the date of PD diagnosis until the end of 2018.

Social support registry data and group definition

Two social support registry databases for patients with PD were used as indicators of functional disability. Disability-registered patients were considered as those approved in either of these two social support registries. We used the grade at the time of the first registration for patients who underwent multiple reevaluations.

First, the long-term care insurance (LTCI) of South Korea provides nursing services for individuals with limitations in daily activities due to geriatric diseases such as stroke, dementia, and PD, as well as normal elderly individuals aged ≥ 65 years with limitations in daily activities. It is provided in the form of home-based, institution-based, and special cash benefits. The review of long-term care is based on a doctor's opinion after examining the individual's condition, with the Rating Committee subsequently deciding whether to approve LTCI and the approval grade. The doctor's note form required for the LTCI application and its grade definition are provided in Supplementary Document 2 and Table S1, respectively [20, 21].

According to the National Disability Registry of South Korea, [21] PD can be approved for brain lesion disability after > 1 year of diligent and continuous treatment as well as sufficient medical records indicating major symptoms or dopaminergic neuronal loss confirmed by single photon emission computed tomography or N-(3-[18 F] fluoropropyl)-2 β -carbomethoxy-3 β -(4-iodophenyl)nortropane positron emission tomography. The diagnosis of disability mainly reflects the degree of overall functional impairment based on the degree and extent of paralysis, balance disorder, ataxia symptoms, and ability to perform activities of daily living. The doctor's note form required for application to the National Disability Registry of brain lesions and its grade definitions are provided in Supplementary Document 3 and Table S2, respectively.

Patients were classified according to functional disability and death as follows: survived and no disability registered (S-NDR group), survived and disability registered (S-DR group), death and disability registered (D-DR group), and death and no disability registered (D-NDR group).

Statistical analysis and clustering method

All statistical analyses and clustering were performed using the R software version 4.0.3 (R Core Team, R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was set at $P < 0.05$.

We used the 'NbClust' R software package to determine the optimal number of clusters [22]. The Hubert index, which is a graphical method for determining the optimal number of clusters, was confirmed; accordingly, we determined the optimal number of clusters to be six (Figure S1). Next, we constructed a divisive hierarchical tree for auto-clustering and analyzed the baseline characteristics of the six groups (Clusters 1–6).

Continuous and categorical variables are expressed as mean \pm standard deviation and frequency (proportion), respectively, with between-group comparisons using analysis of variance with Tukey's comparisons and the chi-square test, respectively. A multivariable Cox proportional hazards model was established to determine risk factors for long-term mortality. Multicollinearity between variables was defined as $\text{sqrt}(\text{variation inflation factor}) > 2$. The Cox regression analysis treated variables with six or more categories as continuous variables. We ran the time-dependent Cox regression for the feature of LTCI grade, which was determined after diagnosing PD.

Results

Classification based on functional disability and death

There were 478 and 722 patients in the S-NDR and S-DR groups, respectively, as well as 1313 and 431 patients in the D-DR and D-NDR groups, respectively. Table 1 shows their baseline characteristics. Figure 2 shows cumulative changes in the distribution of patients according to death and disability registration.

Patients who survived were younger and had a higher proportion of females than those who died within the observation period. The death rate was high in both the medical aid and fourth-quartile premium level groups. The rates of hypertension, diabetes, ischemic heart disease, chronic kidney disease, and stroke were high in the deceased patient groups. In the D-DR group, both LTCI and National Disability Registry grades were significantly lower (more severe disability) than in the other groups. Surviving patients were followed up for 113.7 ± 3.3 months. The D-DR and D-NDR groups were followed up for 61.4 ± 30.1 and 43.2 ± 32.0 months, respectively. The S-DR group registered later in both the LTCI and National Disability Registry than the D-DR group; furthermore, the approval rate was higher for LTCI than for the National Disability Registry.

Classification based on the auto-clustering

Table 2 shows the baseline characteristics of each cluster. Table S3 presents the distribution of patients among the four manually classified groups and auto-classified clusters. All surviving patients were allocated to Clusters 1 and 2. Contrastingly, all patients in Clusters 3–6 died during the observation period. The cumulative survival rate was the highest in Cluster 1 and decreased with time from Clusters 1 to 6 (Fig. 3).

Cluster 1 had the lowest average age and a high proportion of women. Additionally, it had the lowest rates of diabetes and stroke and the patients were mainly approved for mild disability in LTCI without being registered in the National Disability Registry. Cluster 6 showed the shortest survival period (4.2 ± 3.0 months) and comprised all patients in the D-NDR group. Contrastingly, disability-registered patients belonged to Clusters

Table 1 Baseline characteristics according to the manually classified outcome groups

Variables	S-NDR group (n = 478)	S-DR group (n = 722)	D-DR group (n = 1313)	D-NDR group (n = 431)	p-value
Age, years	68.3 ± 8.1	71.1 ± 7.6	75.5 ± 6.9	74.8 ± 6.7	<0.001
Male, n (%)	146 (30.5)	195 (27.0)	556 (42.3)	230 (53.4)	<0.001
Residence areas, n (%)					0.140
Metropolitan Seoul	97 (20.3)	167 (23.1)	253 (19.3)	76 (17.6)	
Metropolitan	78 (16.3)	121 (16.8)	242 (18.4)	91 (21.1)	
City	207 (43.3)	319 (44.2)	602 (45.8)	185 (42.9)	
County	96 (20.1)	115 (15.9)	216 (16.5)	79 (18.3)	
Health insurance types, n (%)					0.003
Self-employed	148 (31.0)	207 (28.7)	367 (28.0)	112 (26.0)	
Employed	283 (59.2)	432 (59.8)	781 (59.5)	238 (55.2)	
Medical-aid	47 (9.8)	83 (11.5)	165 (12.6)	81 (18.8)	
NHIP levels, n (%)					0.002
Medical-aid	47 (9.8)	83 (11.5)	165 (12.6)	81 (18.8)	
1st quartile	61 (12.8)	86 (11.9)	140 (10.7)	42 (9.7)	
2nd quartile	64 (13.4)	92 (12.7)	155 (11.8)	45 (10.4)	
3rd quartile	105 (22.0)	141 (19.5)	221 (16.8)	79 (18.3)	
4th quartile	201 (42.1)	320 (44.3)	632 (48.1)	184 (42.7)	
Comorbidities, n (%)					
Hypertension	346 (72.4)	558 (77.3)	1018 (77.5)	344 (79.8)	0.048
Diabetes	189 (39.5)	310 (42.9)	618 (47.1)	221 (51.3)	0.001
Dyslipidemia	253 (52.9)	376 (52.1)	631 (48.1)	197 (45.7)	0.052
Ischemic heart diseases	36 (7.5)	62 (8.6)	112 (8.5)	58 (13.5)	0.007
Atrial fibrillation	19 (4.0)	30 (4.2)	70 (5.3)	28 (6.5)	0.214
Chronic kidney disease	9 (1.9)	7 (1.0)	44 (3.4)	27 (6.3)	<0.001
Stroke	211 (44.1)	387 (53.6)	733 (55.8)	223 (51.7)	<0.001
Cancer	35 (7.3)	44 (6.1)	107 (8.1)	43 (10.0)	0.105
LTCI grade, n (%)					<0.001
1	0 (0.0)	24 (3.8)	152 (12.1)	0 (0.0)	
2	0 (0.0)	70 (11.0)	244 (19.4)	0 (0.0)	
3	0 (0.0)	334 (52.4)	797 (63.4)	0 (0.0)	
4	0 (0.0)	162 (25.4)	53 (4.2)	0 (0.0)	
5	0 (0.0)	44 (6.9)	11 (0.9)	0 (0.0)	
Out of grade	0 (0.0)	4 (0.6)	1 (0.1)	0 (0.0)	
National Disability Registration grade, n (%)					<0.001
1	0 (0.0)	23 (3.2)	79 (6.0)	0 (0.0)	
2	0 (0.0)	37 (5.1)	108 (8.2)	0 (0.0)	
3	0 (0.0)	69 (9.6)	103 (7.8)	0 (0.0)	
4	0 (0.0)	59 (8.2)	48 (3.7)	0 (0.0)	
5	0 (0.0)	54 (7.5)	32 (2.4)	0 (0.0)	
6	0 (0.0)	25 (3.5)	7 (0.5)	0 (0.0)	
Out of grade	478 (100.0)	455 (63.0)	963 (71.3)	431 (100.0)	
Months to death ^a	113.7 ± 3.2	113.7 ± 3.3	61.4 ± 30.1	43.2 ± 32.0	<0.001
Months to LTCI registration ^a	113.7 ± 3.2	63.7 ± 39.6	25.3 ± 27.2	43.2 ± 32.0	<0.001
Months to National Disability Registration ^a	113.7 ± 3.2	87.5 ± 40.4	49.3 ± 33.1	43.2 ± 32.0	<0.001

Abbreviations: D, death; DR, disability registration; LTCI, long-term care insurance; NDR, no disability registration; NHIP, national health insurance premium; S, survival

^amean observation months

1 to 5. The time from diagnosis to disability registration decreased from Clusters 1 to 5. In Clusters 3 and 4, many patients were approved only for LTCI but not for the National Disability Registry. Patients in Cluster 5 had a severe degree of disability and a higher rate of registration in the National Disability Registry than the other clusters.

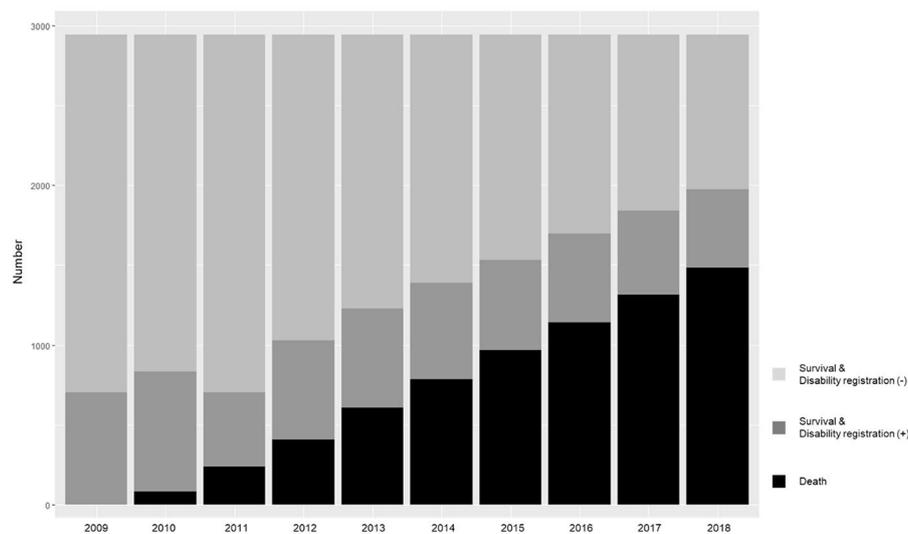


Fig. 2 Annual cumulative changes in target outcomes

Cox-proportional hazards model for long-term mortality

Multivariable Cox-regression revealed that defined clusters were independently associated with long-term mortality after adjusting potential contributing variables; as the defined cluster number increased, the long-term mortality risk significantly increased (adjusted hazard ratio [aHR], 3.440; 95% confidence interval [CI], 3.233–3.659; $p < 0.001$). In addition, age (aHR, 1.038; 95% CI, 1.031–1.045; $p < 0.001$), diabetes (aHR, 1.146; 95% CI, 1.037–1.267; $p = 0.007$), and chronic kidney disease (aHR, 1.382; 95% CI, 1.080–1.768; $p = 0.010$) were identified as independent risk factors for increased long-term mortality in patients with PD. On the other hand, the female gender (aHR, 0.753; 95% CI, 0.681–0.833; $p < 0.001$) and a higher LTCI grade, which indicates lesser disabilities (aHR, 0.995; 95% CI, 0.992–0.997; $p < 0.001$), were associated with a significantly lower long-term mortality risk (Table 3).

Discussion

We classified patients with PD according to their long-term clinical course and described the characteristics of each group. Our auto-clustering analysis revealed six phenotypes of the natural clinical course of PD, which mainly reflected functional disability and mortality. Therefore, our findings may reflect the motor symptom-oriented clinical course of patients with PD. To our knowledge, this is the first study to attempt to classify the natural clinical course of PD based on demographic factors, comorbidities, and social support registry data using the Korean NHIS database.

We could objectively identify information regarding mortality approximately 10 years after PD diagnosis without loss to follow-up. Additionally, functional disabilities could be identified by combining two types of social support data, which minimized information bias. Furthermore, both the LTCI and National Disability Registry allow quantification of the disability degree through grade definition. Taken together, these characteristics allowed reliable description of the long-term clinical course of PD.

Previous studies have identified subtypes of PD using clustering methods [23], which were primarily performed based on disease progression and symptom domains. Belvisi et al. [24] performed agglomerative hierarchical clustering based on motor and

Table 2 Baseline characteristics according to the auto-clustering groups

Variables	Cluster 1 (n=750)	Cluster 2 (n=979)	Cluster 3 (n=425)	Cluster 4 (n=547)	Cluster 5 (n=187)	Cluster 6 (n=56)	p-value
Age, years	69.3±7.9	72.5±7.3	76.1±6.7	76.6±6.7	75.0±7.0	75.3±8.3	<0.001
Male, n (%)	220 (29.3)	336 (34.3)	171 (40.2)	260 (47.5)	110 (58.8)	30 (53.6)	<0.001
Residence areas, n (%)							0.231
Metropolitan Seoul	163 (21.7)	212 (21.7)	82 (19.3)	96 (17.6)	30 (16.0)	10 (17.9)	
Metropolitan	118 (15.7)	182 (18.6)	74 (17.4)	110 (20.1)	38 (20.3)	10 (17.9)	
City	328 (43.7)	429 (43.8)	190 (44.7)	245 (44.8)	89 (47.6)	32 (57.1)	
County	141 (18.8)	156 (15.9)	79 (18.6)	96 (17.6)	30 (16.0)	4 (7.1)	
Health insurance types, n (%)							<0.001
Self-employed	225 (30.0)	278 (28.4)	120 (28.2)	144 (26.3)	53 (28.3)	14 (25.0)	
Employed	444 (59.2)	603 (61.6)	236 (55.5)	303 (55.4)	116 (62.0)	32 (57.1)	
Medical-aid	81 (10.8)	98 (10.0)	69 (16.2)	100 (18.3)	18 (9.6)	10 (17.9)	
NHIP levels, n (%)							<0.001
Medical-aid	81 (10.8)	98 (10.0)	69 (16.2)	100 (18.3)	18 (9.6)	10 (17.9)	
1st quartile	94 (12.5)	110 (11.2)	44 (10.4)	53 (9.7)	21 (11.2)	7 (12.5)	
2nd quartile	99 (13.2)	121 (12.4)	48 (11.3)	57 (10.4)	24 (12.8)	7 (12.5)	
3rd quartile	164 (21.9)	181 (18.5)	60 (14.1)	90 (16.5)	41 (21.9)	10 (17.9)	
4th quartile	312 (41.6)	469 (47.9)	204 (48.0)	247 (45.2)	83 (44.4)	22 (39.3)	
Comorbidities, n (%)							
Hypertension	556 (74.1)	750 (76.6)	344 (80.9)	432 (79.0)	144 (77.0)	40 (71.4)	0.092
Diabetes	303 (40.4)	432 (44.1)	205 (48.2)	279 (51.0)	91 (48.7)	28 (50.0)	0.003
Dyslipidemia	387 (51.6)	518 (52.9)	191 (44.9)	242 (44.2)	96 (51.3)	23 (41.1)	0.004
Ischemic heart diseases	63 (8.4)	80 (8.2)	38 (8.9)	61 (11.2)	19 (10.2)	7 (12.5)	0.381
Atrial fibrillation	34 (4.5)	42 (4.3)	16 (3.8)	39 (7.1)	12 (6.4)	4 (7.1)	0.097
Chronic kidney disease	13 (1.7)	11 (1.1)	14 (3.3)	34 (6.2)	9 (4.8)	6 (10.7)	<0.001
Stroke	355 (47.3)	528 (53.9)	227 (53.4)	300 (54.8)	111 (59.4)	33 (58.9)	0.012
Cancer	50 (6.7)	74 (7.6)	27 (6.4)	50 (9.1)	22 (11.8)	6 (10.7)	0.118
LTCI grade, n (%)							<0.001
1	5 (2.0)	61 (7.3)	27 (8.5)	37 (11.3)	46 (27.7)	0 (0.0)	
2	17 (6.9)	120 (14.4)	53 (16.6)	89 (27.2)	35 (21.1)	0 (0.0)	
3	73 (29.4)	553 (66.1)	219 (68.7)	201 (61.5)	85 (51.2)	0 (0.0)	
4	118 (47.6)	80 (9.6)	17 (5.3)	0 (0.0)	0 (0.0)	0 (0.0)	
5	32 (12.9)	20 (2.4)	3 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	
Others	3 (1.2)	2 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
National Disability Registration grade, n (%)							<0.001
1	3 (0.4)	71 (7.3)	1 (0.2)	0 (0.0)	27 (14.4)	0 (0.0)	
2	4 (0.5)	97 (9.9)	0 (0.0)	0 (0.0)	44 (23.5)	0 (0.0)	
3	11 (1.5)	116 (11.8)	7 (1.6)	0 (0.0)	38 (20.3)	0 (0.0)	
4	14 (1.9)	81 (8.3)	1 (0.2)	0 (0.0)	11 (5.9)	0 (0.0)	
5	23 (3.1)	57 (5.8)	1 (0.2)	0 (0.0)	5 (2.7)	0 (0.0)	
6	5 (0.7)	27 (2.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Out of grade	690 (90.2)	530 (54.1)	415 (97.6)	547 (100)	62 (33.2)	56 (100)	
Months to death ^a	113.7±3.2	99.8±19.8	72.0±10.7	31.1±13.8	25.3±14.7	4.2±3.0	<0.001
Months to LTCI registration ^a	106.6±12.9	45.9±36.6	38.4±28.8	15.7±14.7	11.1±10.2	4.2±3.0	<0.001
Months to National Disability Registration ^a	111.9±8.0	69.1±42.6	70.9±13.2	31.1±13.8	9.8±7.9	4.2±3.0	<0.001

Abbreviations: LTCI, long-term care insurance; NHIP, National Health Insurance Premium

^amean observation months

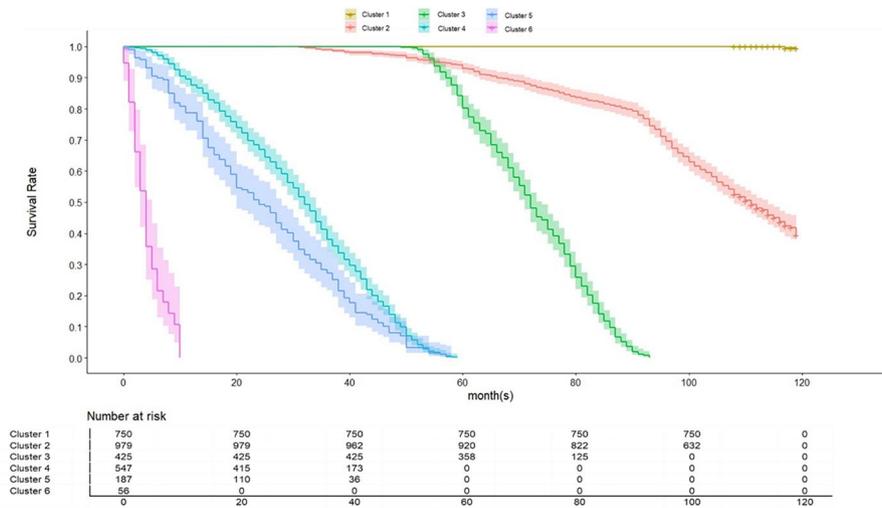


Fig. 3 Cumulative survival rate according to the auto-clustering groups

Table 3 Cox-proportional hazards model for long-term mortality after diagnosis of Parkinson's disease

Variables	Adjusted HR	95% confidence interval	p-value
Clusters (per a cluster)	3.440	3.233–3.659	<0.001
LTCI grade (per a grade) ^a	0.995	0.992–0.997	<0.001
Age	1.038	1.031–1.045	<0.001
Female	0.753	0.681–0.833	<0.001
Residence areas			
Metropolitan Seoul	1.000		
Metropolitan	1.120	0.959–1.307	0.152
City	1.005	0.881–1.146	0.944
County	1.060	0.903–1.244	0.475
Health insurance types ^b			
Self-employed	1.000		
Employed	1.012	0.904–1.132	0.838
NHIP levels			
Medical-aid	1.000		
1st quartile	0.896	0.699–1.148	0.386
2nd quartile	0.953	0.743–1.222	0.706
3rd quartile	0.898	0.710–1.136	0.371
4th quartile	0.926	0.758–1.131	0.451
Comorbidities			
Hypertension	0.969	0.859–1.094	0.610
Diabetes	1.146	1.037–1.267	0.007
Dyslipidemia	0.901	0.868–1.062	0.430
Ischemic heart diseases	1.113	0.947–1.308	0.193
Atrial fibrillation	1.046	0.850–1.287	0.670
Chronic kidney disease	1.382	1.080–1.768	0.010
Stroke	0.979	0.888–1.079	0.666
Cancer	1.152	0.970–1.367	0.106

Abbreviations: HR, hazard ratio; LTCI, long-term care insurance; NHIP, National Health Insurance Premium

^atime-dependent variable

^bexcluding the medical-aid group

neurophysiological features and identified two clinical clusters: mild motor dominant and diffuse malignant types. Additionally, they confirmed that the diffuse malignant type is characterized by increased cortical excitability and decreased plasticity. Lawton et al. [25] performed K-means clustering based on data regarding the motor, cognitive, and non-motor domains for two cohorts of patients with idiopathic PD. The following four clusters were identified: fast motor progression with symmetrical motor disease, intermediate motor progression with mild motor and non-motor disease, intermediate motor progression with severe motor disease, and slow motor progression with unilateral disease. Krishnagopal et al. [26] used the trajectory profile clustering method and described mild, severe, and mixed subtypes of PD. Salmanpour et al. [27] performed a 4-year longitudinal clustering analysis using multiple dimensions of classification algorithms and found that 35% of the patients showed a stable course, while others showed disease escalation.

Based on previous studies, our clusters can be described as follows: Cluster 1 corresponds to the mild motor or non-motor predominant slow progression type; Clusters 2 and 3 correspond to intermediate types; and Clusters 4–6 correspond to diffuse, malignant, and bilateral motor disease subtypes. Patients in Cluster 6 lacked enough time to register their disabilities since their survival time was only about 4 months. Furthermore, in Cluster 6, late diagnosis or other PD-unrelated causes may have played a more critical role in death as indicated by the relatively higher proportion of comorbidities, including ischemic heart disease, atrial fibrillation, chronic kidney disease, and stroke. Moreover, we operationally defined patients with PD into four groups based on their survival and disability registrations. The S-NDR and S-DR groups may correspond to the mild motor or non-motor predominance and slow progression subtypes, while the D-DR and D-NDR groups had a mixture of intermediate and diffuse malignant subtypes.

Approval for LTCI is based on the assessment score indicating the degree of long-term care required by the applicant. For individuals without dementia, patients with a score of >51 points for the sum of physical function, cognitive function, behavioral change, nursing care, and rehabilitation area are considered eligible. Contrastingly, the cut-off criterion for the National Disability Registry for brain lesions is a modified Barthel index ≤ 96 points. We could not present direct data regarding the Unified Parkinson's Disease Rating Scale (UPDRS) score and Hoehn and Yahr (HY) stages, which are widely used to evaluate PD. However, from the perspective of starting to need help from others, the initiation of disability registration can be considered equivalent to a total UPDRS score of 50–60 and HY stage 2.5–3 [28]. We found that patients with PD in South Korea were more registered in the LTCI than in the National Disability Registry for functional disabilities. This could be attributed to several reasons. First, the National Disability Registry of brain lesions is primarily based on the modified Barthel index and emphasizes the evaluation of activities of daily living. Contrastingly, the LTCI evaluates both functional level and other motor items, including ataxia and tremor. Moreover, LTCI considers non-motor symptoms including cognitive function and problem behavior, which allows better objective evaluation of disabled patients with PD. This is evident from the fact that the time frame for registration with LTCI was shorter than that for all the groups.

We identified age, male gender, greater degree of disabilities, diabetes, and chronic kidney disease as risk factors for long-term mortality after the diagnosis of PD, which is consistent with previous studies [9, 29]. Moreover, socioeconomic status was not

associated with long-term mortality, which is inconsistent with previous reports [30, 31]. As aforementioned, these results could be attributed to the government-provided medical cost support for the diagnosis code of G20 in South Korea. Further studies are warranted to identify the associations between socioeconomic status and long-term mortality of PD in South Korea.

This study had several limitations. First, since this was a retrospective study, we could not present the individual patient's medication history or compliance, which are critical for the long-term clinical course of PD. Further, the NHIS database cannot reflect detailed records of motor and non-motor symptoms at the time of diagnosis, including the UPDRS and HY stages. Second, our findings could not reflect the biological signatures of PD. Although the NHIS database allows analysis of social registry data and objective evaluation of longitudinal events, including death, it cannot accurately reflect detailed information about the disease, imaging data, and biological information regarding the pathogenesis of each patient. Third, we could not confirm whether the death was due to complications directly related to PD. Therefore, future studies should conduct systematic research using combined databases from hospitals and the NHIS to overcome these limitations. Finally, while the dataset used in this study has sufficient clinical and demographic characteristics at the time of PD diagnosis, it lacks variables measured repeatedly after the PD diagnosis. Then, we applied clustering analysis to overcome this limitation of our dataset and ensure statistical robustness. Clustering analysis is unsupervised learning that searches for patterns in data and provides relatively simple and intuitive results, but the results do not consider changes over time [32]. On the other hand, the growth mixture model can identify subpopulations with different patterns within the data based on repeated measurement data [33]. However, relatively complex statistical modeling is applied for the growth mixture model, requiring repeated measurement data from sufficient time points [34]. In the case where the repeated measurement data are unavailable, as in this study, applying the growth mixture model may be limited. We suggest that the future study utilizing the growth mixture modeling with repeated measurement values in the PD cohort will contribute significantly to the robustness and generalizability of the results.

Conclusion

We described six PD clinical course clusters using longitudinal data from the social support registry and mortality data obtained from the NHIS database. We confirmed that PD progression is heterogeneous with respect to disability and mortality. Our findings may inform long-term management strategies for patients with PD.

Abbreviations

aHR	adjusted hazard ratio
CI	confidence interval
DR	disability-registered
ICD	International Classification of Diseases
NDR	no disability registered
LTCI	long-term care insurance
NHIS	National Health Insurance Service
PD	Parkinson's disease

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40537-023-00819-z>.

Supplementary Material 1

Acknowledgements

This study used data from the database compiled by the National Health Insurance Service (research management number: NHIS-2020-1-160).

Authors' contributions

Conceptualization and design: H.S.K.; Data curation: J.H.K. and H.S.K.; Investigation: D.P. and S.Y.L.; Formal analysis: J.H.K. and H.S.K.; Validation: J.H.K. and H.S.K.; Writing—original draft: D.P. and S.Y.L.; Writing—review and editing: D.P. and S.Y.L., J.H.K. and H.S.K.; Visualization: D.P. and H.S.K. All authors have read and agreed to the published version of the manuscript.

Funding

None.

Data availability

The data from this study are not publicly available due to privacy and ethical restrictions of the Korean National Health Insurance data-sharing system. The dataset used in this study can only be accessed by authorized researchers. Dr. Jong Hun Kim and Dr. Hyoung Seop Kim had full access to all of the data analyzed in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Declarations**Competing interests**

The authors declare no competing interests.

Ethics approval and consent to participate

This study was reviewed and approved by the Institutional Review Board of National Health Insurance Service Ilsan Hospital (NHIS-2023-02-002). Informed consent was waived due to the retrospective study design and anonymity of the National Health Insurance Service database.

Consent for publication

N/A.

Received: 23 March 2023 / Accepted: 30 August 2023

Published online: 09 September 2023

References

1. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. *Lancet Neurol*. 2006;5(6):525–35.
2. Ni A, Ernst C. Evidence that Substantia Nigra Pars Compacta dopaminergic neurons are selectively vulnerable to oxidative stress because they are highly metabolically active. *Front Cell Neurosci*. 2022;16:826193.
3. Pelzer EA, Sturmer S, Feis DL, Melzer C, Schwartz F, Schrage M, et al. Clustering of Parkinson subtypes reveals strong influence of DRD2 polymorphism and gender. *Sci Rep*. 2022;12(1):6038.
4. De Pablo-Fernández E, Lees AJ, Holton JL, Warner TT. Prognosis and neuropathologic correlation of clinical subtypes of Parkinson disease. *JAMA Neurol*. 2019;76(4).
5. Pirtosek Z, Bajenaru O, Kovacs N, Milanov I, Relja M, Skorvanek M. Update on the management of Parkinson's disease for general neurologists. *Parkinsons Dis*. 2020;2020:9131474.
6. Prange S, Danaila T, Laurencin C, Caire C, Metereau E, Merle H, et al. Age and time course of long-term motor and non-motor complications in Parkinson disease. *Neurology*. 2019;92(2):e148–e60.
7. Chou KL, Stacy M, Simuni T, Miyasaki J, Oertel WH, Sethi K, et al. The spectrum of off in Parkinson's disease: what have we learned over 40 years? *Parkinsonism Relat Disord*. 2018;51:9–16.
8. Schapira AHV, Chaudhuri KR, Jenner P. Non-motor features of Parkinson disease. *Nat Rev Neurosci*. 2017;18(7):435–50.
9. Macleod AD, Taylor KSM, Counsell CE. Mortality in Parkinson's disease: a systematic review and meta-analysis. *Mov Disord*. 2014;29(13):1615–22.
10. Moscovich M, Boschetti G, Moro A, Teive HAG, Hassan A, Munhoz RP. Death certificate data and causes of death in patients with parkinsonism. *Parkinsonism Relat Disord*. 2017;41:99–103.
11. Pennington S, Snell K, Lee M, Walker R. The cause of death in idiopathic Parkinson's disease. *Parkinsonism Relat Disord*. 2010;16(7):434–7.
12. Goetz CG, Nutt JG, Stebbins GT. The Unified Dyskinesia Rating Scale: presentation and clinimetric profile. *Mov Disord*. 2008;23(16):2398–403.
13. Ray Chaudhuri K, Rojo JM, Schapira AH, Brooks DJ, Stocchi F, Odin P, et al. A proposal for a comprehensive grading of Parkinson's disease severity combining motor and non-motor assessments: meeting an unmet need. *PLoS ONE*. 2013;8(2):e57221.
14. Fereshtehnejad SM, Postuma RB. Subtypes of Parkinson's disease: what do they tell us about disease progression? *Curr Neurol Neurosci Rep*. 2017;17(4):34.
15. Fereshtehnejad SM, Zeighami Y, Dagher A, Postuma RB. Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression. *Brain*. 2017;140(7):1959–76.
16. Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease: a review. *JAMA*. 2020;323(6):548–60.
17. Mestre TA, Fereshtehnejad S-M, Berg D, Bohnen NI, Dujardin K, Erro R, et al. Parkinson's disease subtypes: critical appraisal and recommendations. *J Parkinson's Disease*. 2021;11(2):395–404.

18. Cheol Seong S, Kim YY, Khang YH, Heon Park J, Kang HJ, Lee H, et al. Data resource profile: the national health information database of the National Health Insurance Service in South Korea. *Int J Epidemiol*. 2017;46(3):799–800.
19. Song SO, Jung CH, Song YD, Park CY, Kwon HS, Cha BS, et al. Background and data configuration process of a nationwide population-based study using the Korean national health insurance system. *Diabetes Metab J*. 2014;38(5):395–403.
20. Ga H. Long-term care system in Korea. *Ann Geriatr Med Res*. 2020;24(3):181–6.
21. Kim H, Kwon S. A decade of public long-term care insurance in South Korea: policy lessons for aging countries. *Health Policy*. 2021;125(1):22–6.
22. Charrad M, Ghazzali N, Boiteau V, Niknafs A, NbClust. An R package for determining the relevant number of clusters in a data set. *J Stat Softw*. 2014;61(6):1–36.
23. Khasawneh MT, Hendricks RM. A systematic review of Parkinson's disease cluster analysis research. *Aging and Disease*. 2021;12(7).
24. Belvisi D, Fabbrini A, De Bartolo MI, Costanzo M, Manzo N, Fabbrini G, et al. The pathophysiological correlates of Parkinson's disease clinical subtypes. *Mov Disord*. 2021;36(2):370–9.
25. Lawton M, Ben-Shlomo Y, May MT, Baig F, Barber TR, Klein JC, et al. Developing and validating Parkinson's disease subtypes and their motor and cognitive progression. *J Neurol Neurosurg Psychiatry*. 2018;89(12):1279–87.
26. Krishnagopal S, Coelln RV, Shulman LM, Girvan M. Identifying and predicting Parkinson's disease subtypes through trajectory clustering via bipartite networks. *PLoS ONE*. 2020;15(6):e0233296.
27. Salmanpour MR, Shamsaei M, Saberi A, Klyuzhin IS, Tang J, Sossi V, et al. Machine learning methods for optimal prediction of motor outcome in Parkinson's disease. *Phys Med*. 2020;69:233–40.
28. Shulman LM, Gruber-Baldini AL, Anderson KE, Fishman PS, Reich SG, Weiner WJ. The clinically important difference on the unified Parkinson's disease rating scale. *Arch Neurol*. 2010;67(1):64–70.
29. Gonzalez MC, Dalen I, Maple-Grodem J, Tysnes OB, Alves G. Parkinson's disease clinical milestones and mortality. *NPJ Parkinsons Dis*. 2022;8(1):58.
30. Yang F, Johansson ALV, Pedersen NL, Fang F, Gatz M, Wirdefeldt K. Socioeconomic status in relation to Parkinson's disease risk and mortality: a population-based prospective study. *Med (Baltim)*. 2016;95(30):e4337.
31. Genc G, Abboud H, Oravivattanukul S, Alsallom F, Thompson NR, Cooper S, et al. Socioeconomic status may impact functional outcome of deep brain stimulation surgery in Parkinson's disease. *Neuromodulation*. 2016;19(1):25–30.
32. Serra-Burriel M, Ames C. Machine learning-based clustering analysis: foundational concepts, methods, and applications. *Acta Neurochir Suppl*. 2022;134:91–100.
33. Ram N, Grimm KJ. Growth mixture modeling: a method for identifying differences in longitudinal change among unobserved groups. *Int J Behav Dev*. 2009;33(6):565–76.
34. Kwon JY, Sawatzky R, Baumbusch J, Lauck S, Ratner PA. Growth mixture models: a case example of the longitudinal analysis of patient-reported outcomes data captured by a clinical registry. *BMC Med Res Methodol*. 2021;21(1):79.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.