Parkinson’s Disease Analysis Report

Isabella Grasso
Nipuni De Silva
Paul Dougall

Introduction

Parkinson's Disease (PD) is an incurable neurodegenerative disease that affects roughly 7-10 million people world wide [1]. It is also categorized as a movement disorder, characterized by tremors, rigidity, and bradykinesia. The difficulty of diagnosing PD is currently very high due to the fact that motor symptoms do not appear until late in the disease process and there are no other symptoms as characteristic as the motor symptoms [2]. Currently, PD diagnoses are clinical as there is no diagnostic test for PD and for that reason the symptoms that accompany PD play an important role in diagnosing the disease. There have been connections made between PD diagnosis and dopaminergic cell death, where enzymatic and non-enzymatic oxidation of dopamine generates reactive oxygen species, which induces apoptotic cell death in dopamine neurons, as well as the accumulation of alpha synuclein (SNCA) which is a protein that is abundant in the brain, heart, muscle, gut, and brint and is concentrated near the tips of nerve cells [4]. This is critical in the evolution of PD because both are involved in controlling the start and stop of voluntary and involuntary movements. Large accumulations of SNCA are called “lewy bodies” which are currently believed to be connected to PD. However, 20% of people over 70 have lewy bodies in their brains who do not suffer from motor or memory dysfunction[1]. There are also non-motor symptoms associated with PD such as cognitive impairment, dementia, constipation, fatigue, restless leg syndrome, sleep behavior disorder, and hallucinations. Comorbid conditions include bone fractures, dementia, diabetes, stroke, hypertension, anemia, and depression.

Unfortunately PD is a very complex disease and is thus difficult to both predict and understand[5]. The biological link between PD and various comorbidities remains unknown. For example, PD and diabetes share similar dysregulated pathways such as inflammation, mitochondrial dysfunction, and impaired insulin signaling, however, we do not understand why they share these pathways. Similarly, there seems to be a direct association between depression and subsequent development of PD and vice versa. Scientists hypothesize that this connection has to do with low serotonin activity in patients with PD as well as patients with

depression. Depression is one of the most frequent non-motor symptoms, with 35% of PD patients eventually being diagnosed. Oddly, there is decreased risk of PD among almost all cancer types except melanoma.

Current models use multivariate logistic regression analysis and can predict PD with less than five years of notice. They utilize >50 years, anxiety & depression, fatigue, apathy, insomnia, balance impairment, dizziness, hypotension, anosmia, hypersalivation, constipation, urinary dysfunction, erectile dysfunction, and memory problems. Some also included smoking status and alcohol consumption. In this study we focus on preliminary data exploration and visualization to determine relevant variables and comorbid conditions for further modeling[2].

Methods

Study Population & Data Extraction

The data utilized for these analyses was obtained from the National Medical Care Survey (NAMCS) from the years of 2007 to 2009. NAMCS is a national survey conducted in the United States sampling both office-based and community health center-based physicians and data from their patients. Office based physicians were selected from the American Medical Association (AMA) and the American Osteopathic Association (AOA) who were classified as “office based, patient care.” These physicians had to meet the additional criteria of: being principally engaged in patient care activities, being non federally employed; and not specializing in anesthesiology, pathology, or radiology. Community health center-based physicians were chosen utilizing guidelines designed by the Health Resources and Service Administration’s Bureau Primary Health Care and the Indian Health Service.

A multisample probability design was used to determine which physicians and entries to include and is described in detail in the NAMCS micro-data file documentation. NAMCS additionally included weighting methods as well as imputed data. However, our analysis focused on the unweighted and unimputed data to get a general understanding of survey without incorrectly interpreting the weighting and imputation methods not yet fully understood. The data set was also put through a meticulous process conducted by SRA International, In., Durham, North Carolina in which entries were reviewed, checked for completeness, and recoded when inconsistent or ambiguous.

The data set contained 440 variables specifying information regarding the medications, procedures, demographics, vital signs, and medical diagnoses of the patient, in addition to data pertaining to the physician, the facility, and the visit itself. The International Classification of

Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) system was used for diagnoses and medical procedure coding, while the National Center for Health Statistics (NCHS) and the Multum Lexicon Drug Database classification systems were both utilized whenever possible.

Demographic Variables and Preprocessing

Parkinson's Disease: Parkinson's presence or absence was determined by primary, secondary, and tertiary diagnosis. It was manipulated into a single column binary yes/no.

Sex: Sex is self reported binary male/female. Rows with missing data were omitted.

Age: Age was manipulated into a categorical variable with thresholding every 10 years.

Race: Race is self reported; White, Black, Asian, Native Hawaiian/Pacific Island, American Indian/Alaskan Native. Native Hawaiian/Pacific Island and American Indian/Alaskan Native were omitted due to lack of data.

Major Reason For Visit: Defined as New, pre/post surgery, chronic (routine), or chronic (flare up).

Median Household Income in Person's Zip Code: $0 - $32,793, $32,793 - $40,626, $40,627 - $52,387, $52,387 and above.

Cancer: Cancer is a binary (yes/no) variable indicating the presence or absence of the chronic condition.

Chronic Obstructive Pulmonary Disorder (COPD): COPD is a binary (yes/no) variable indicating the presence or absence of the chronic condition.

Asthma: Asthma is a binary (yes/no) variable indicating the presence or absence of the chronic condition.

Arthritis: Arthritis is a binary (yes/no) variable indicating the presence or absence of the chronic condition.

Diabetes: Diabetes is a binary (yes/no) variable indicating the presence or absence of the chronic condition.

Depression: Depression is a binary (yes/no) variable indicating the presence or absence of the chronic condition.

Congestive Heart Failure (CHF): CHF is a binary (yes/no) variable indicating the presence or absence of the chronic condition.

Hypertension: Hypertension is a binary (yes/no) variable indicating the presence or absence of the chronic condition.

Chronic Renal Failure (CRF): CRF is a binary (yes/no) variable indicating the presence or absence of the chronic condition.

Body Mass Index (BMI): BMI is a continuous variable stating the patients BMI as calculated by the physician.

Tobacco Exposure: Tobacco Exposure is a binary (yes/no) variable indicating the presence or absence of exposure to tobacco.

Total Number of Chronic Conditions: A continuous variable stating the number of chronic conditions a patient is currently diagnosed with.

Geographic Region: A categorical variable stating the region of the country in which the patient resides determined by their Zip Code; Northeast, Midwest, South, West

Results

Year	Total Participants	With Parkinson's Disease	%
2007	32,778	154	0.47%
2008	28,741	92	0.32%
2009	32,281	95	0.29%
Total	93,800	341	0.36%

Table 1: Distribution of PD by year.

Variable	No PD (%)	PD (%)
Age		
20-40	100	0
40-60	99.86	0.14
60-80	99.16	0.84
80+	98.7	1.3
Sex		
Male	99.5	0.49

Female	99.73	0.26
Major Reason for Visit		
Preventative	99.87	0.12
Pre/post surgery	99.99	0.01
New	99.82	0.17
Chronic (Routine)	99.30	0.70
Chronic (Flare up)	99.51	0.49
Comorbid Conditions		
Hypertension	99.53	0.46
Diabetes	99.55	0.45
Depression	99.51	0.49
Arthritis	99.54	0.46

Table 2. Percentage with and without PD by demographic variable and comorbid conditions.

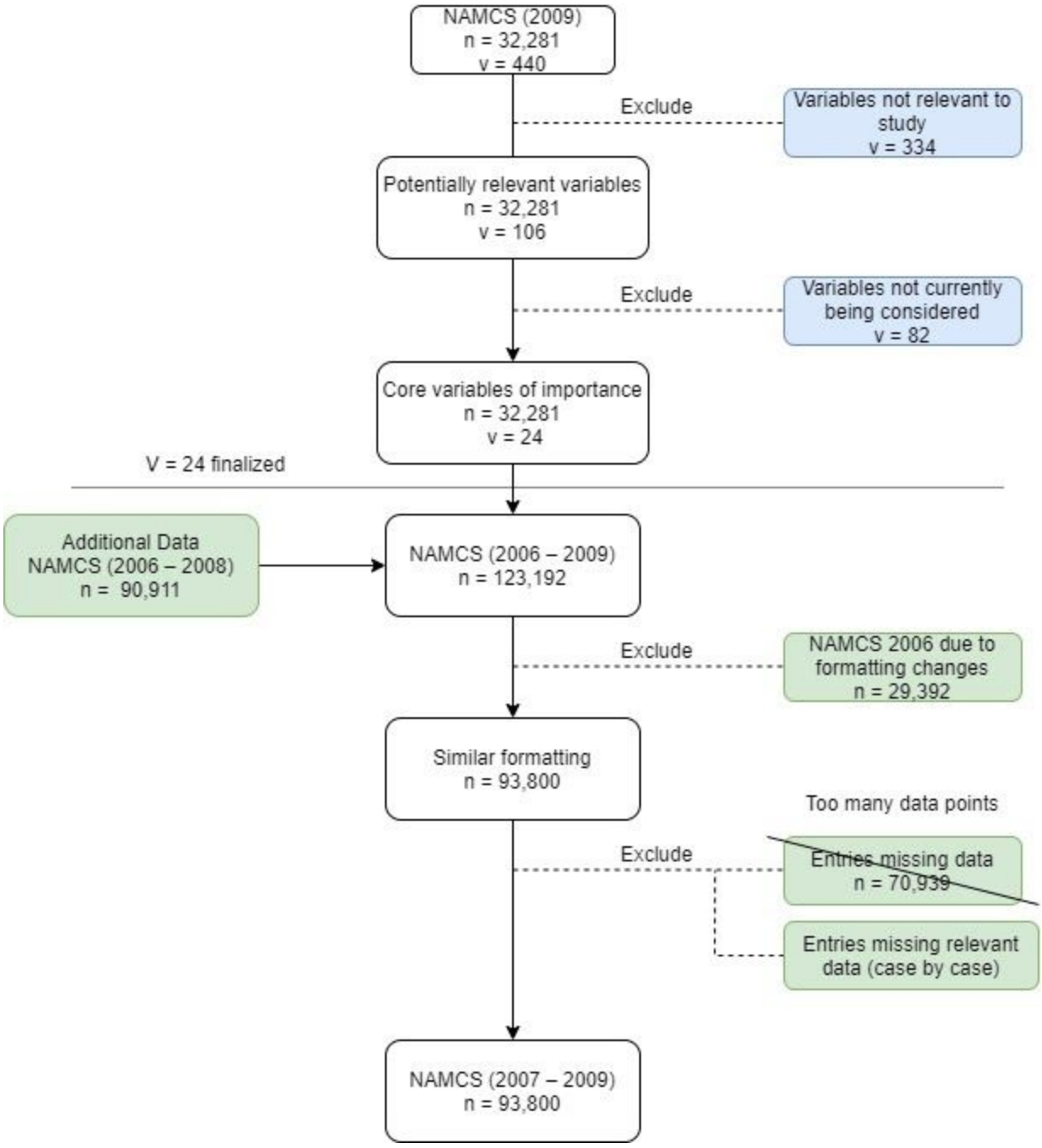

Figure 1. Flow diagram of the selection of study population from NAMCS data (2007-2009)

Statistics

To determine the relationships between PD diagnosis and the various categorical variables, the odds ratio of PD was computed for each variable. The odds ratio is a statistic which measures the strength of the association between two events [6].

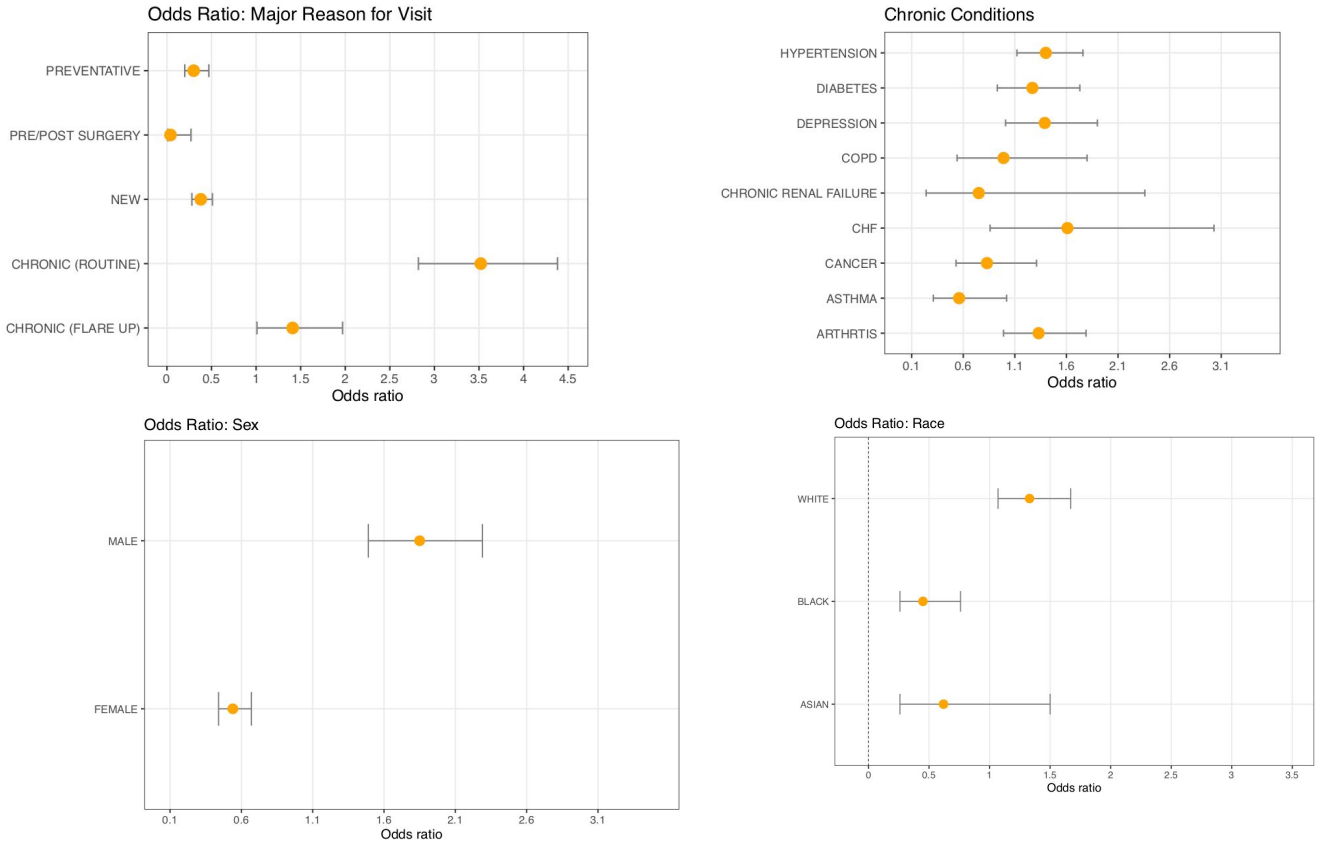

Figure 2: Odds ratio of PD by demographic variables and comorbid conditions. A) major reason for visit, B) selected chronic conditions, C) sex, D) race

Participants with other chronic diseases had a significantly higher likelihood of being diagnosed with PD (Figure 1.A.). Individuals with chronic diseases in for routine visits rather than flare ups had a higher likelihood (3.5:1) of diagnosis as opposed to those in for flare ups (1.5:1). The causal relationship for this phenomena remains unknown. Individuals with hypertension, diabetes, COPD, and arthritis are higher risk (odds ratio >1) of diagnosis with PD (Figure 1.B.) There is not enough data to determine the association between PD incidence and chronic renal failure and congestive heart failure and those with cancer are lower risk (odds ratio <1) of diagnosis with PD. Males are high risk for PD, while females are low risk for PD (Figure 1.C.),

and there was very little data on race other than for black and white participants. Data for American Indian/Native Alaskans and Native Hawaiians/Pacific Islanders was omitted because of lack of data. White participants are at higher risk for diagnosis than for black participants (Figure 1.D.), but the causal relationship remains unknown.

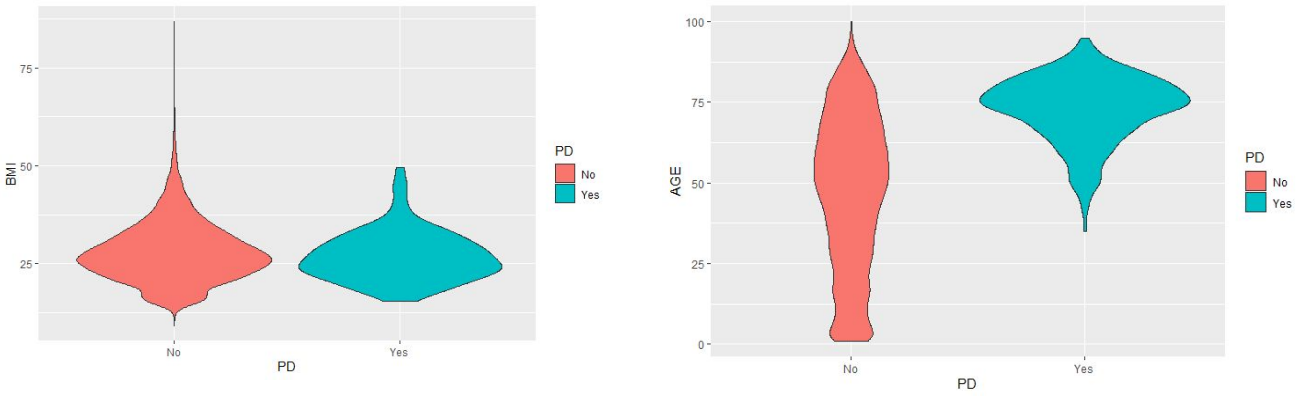

Figure 3: Violin plots comparing the distribution of: A) BMI, B) Age

Patients with PD tend to have a lower BMI than patients without PD (Figure 3.A.), but the causal relationship with this remains unknown. We hypothesize that this is because a high proportion of PD diagnoses are made after the age of 70 (Figure 3.B.), which is associated with lower BMI than middle aged individuals.

Discussion & Limitations

Analysis from the NAMCS data so far has mostly revealed the role that biological sex, race, hypertension, and depression are associated with higher risk of PD. Hypertension is a condition in which the pressure of blood against the artery walls is too strong. We have hypothesized the failure to effectively provide blood to the body without complications (hypertension) might be related to the various motor conditions that are associated with PD. However, as of right now this needs to be further analyzed and researched. The relationship that depression has with PD have not yet been formulated. Outside of these results, it seems that, from these preliminary findings, that there are no other significant relationships between the other categorical variables and PD.

There are very few patients with PD; less than one percent of the participants in this study were diagnosed. The severe imbalance of the data will provide difficulty in modeling. Furthermore, the NAMCS constantly goes through changes and updates in formatting. Ideally we would be able to combine at least ten years of data, but there were significant changes in the data collected, formatting, and the range of quantities in each variable. To expand beyond the scope of 2007-2009 may not be possible without losing predictive power.

Future Work

Incorporating the proper population weighting is the next step before modeling. Then preliminary modeling is required for feature selection. We also need to begin with two-way interaction terms to start gaining deeper understanding of the relationships with PD. We will implement a logistic regression with forward selection and backward elimination to determine which features are the best predictors of PD.

The final direction for the project is the creation of a predictive model. The major purpose of this analysis was to further understand PD and help practitioners with early detection. Most PD patients are latent for years before showing symptoms. At this point they progress quickly and there are few treatment opportunities.

References:

[1] Santiago Jose A., B. V. (2017). Biological and Clinical Implications of Comorbidities in Parkinson's Disease. Frontiers in Aging Neuroscience, 9, 394. doi:10.3389/fnagi.2017.00394

https://dx.doi.org/10.3389/fnagi.2017.00394

[2] Schrag, A., Anastasiou, Z., Ambler, G., Noyce, A. and Walters, K. (2019), Predicting diagnosis of Parkinson's disease: A risk algorithm based on primary care presentations. Mov Disord., 34: 480-486. doi:10.1002/mds.27616 https://onlinelibrary.wiley.com/doi/full/10.1002/mds.27616

[3] Enochs, W.S., Sarna, T., Zecca, L. *et al.* The roles of neuromelanin, binding of metal ions, and oxidative cytotoxicity in the pathogenesis of Parkinson's disease: A hypothesis. *J Neural Transm Gen Sect* 7, 83–100 (1994). https://doi.org/10.1007/BF02260963

[4] Jaeschke, Hartmut & Ramachandran, Anup. (2017). Antioxidant Defense Mechanisms. 10.1016/B978-0-12-801238-3.64200-9.

[5] Mayfield Brain & Spine. (n.d.). Parkinson's Disease (PD). Retrieved February 10, 2020, from https://mayfieldclinic.com/pe-pd.htm

[6] Wikipedia contributors. (2020, April 23). Odds ratio. In *Wikipedia, The Free Encyclopedia*. Retrieved 22:25, May 9, 2020, from https://en.wikipedia.org/w/index.php?title=Odds_ratio&oldid=952596389