

A research on the relationship between Age and the incident rate of the Parkinson's disease

Emery HaoHuo, Dino Yundi Zhou, Nathan Yushu Wu

Beijing National Day School

AP Statistic Final Project

May 26, 2023

1 Background Information

Parkinson's disease, discovered by British doctor James Parkinson, is a neurodegenerative disease that might cause Quiescence tremor, bradykinesia, myotonia, and postural gait disorders. Parkinson disease is caused by degeneration and death of dopamine neurons in the midbrain substantia nigra, and the cause of such pathological change is still unknown. Many factors might contributed to the incident of Parkinson's disease include genetic factors, environmental factors, aging and oxidative stress.[1]Our study today will primarily focus on the effect of aging on the incident rate of Parkinson's disease.

We construct an observational study about the distribution of the patients of Parkinson's disease in different age group, The null hypothesis will be the incident rate is equal between 50-75 age group and 75-100 age group, and the alternative hypothesis will be the interpreted distribution in each age group is not equal

2 Research

With the profound research to Parkinson Diseases, we found some basic factors that may affect the results of the relation of ages and the incidence of the Parkinson. Based on the fundamental research from the WHO, reported that "In addition, men are 1.5 times more likely to develop Parkinson's disease than women". [2] Sex, are the main confounding variables which can disturb this observational study and alter the conclusion, which should must be considered as an important factor. During the Research, we find that incidence of Parkinson's disease increases with age, and the average PD onset age is estimated to be 60 years old, and only about 4 percent of PD patients are diagnosed before the age of 50.

Thus, in the following random sampling method, all of Parkinson patients we randomly selected are above 50 years old to ensure that there is no extreme small outliers which may trigger the problem of underestimating the database. In the process of observational studies, we approximately randomly select 72 Parkinson's patients and look for their distribution of their ages. We find that age 75 is the clear

boundary that can separate and it is appropriate to detect the relationship between ages and the incidence of Parkinson Diseases. In this study, we separate into two main groups, they are separately Age 50-75 groups and Age75-90 groups and we assume that both of them have the same incidence of Parkinson, which is the probability of 0.5 for each group in both male and female groups. Finally, we will use the chi-square of homogeneity to test the result of our experiment.

All the data are collected from Tiantan Hospital, thanks to Yundi's father for helping us collecting the data from the neurology department. Tiantan Hospital, as the most famous hospital in Beijing even in the China, can cover the comprehensive examples so we can use this sample to represent the whole population.

3 Methods and Procedures

Figure 1

Keywords— : F:Female M:Male

3.1 Data Collected

Data Collected: The data for this research project were collected from the neurology department at Tiantan Hospital. The data collection process involved the following steps:

1. Sampling Method: The researchers utilized a random sampling method to select the clinical history of patients with Parkinson's disease. They accessed the hospital's database and randomly selected a subset of patient records from the neurology department.
2. Informed Consent: Prior to data collection, the researchers obtained informed consent from the patients or their legal guardians. The purpose of the study, data collection procedures, and any potential risks or benefits were explained to the participants, ensuring their voluntary participation.
3. Data Extraction: Dino's father, a neurological doctor and collaborator on the project, played a crucial role in the data extraction process. He carefully reviewed the clinical history records of the selected patients and extracted relevant information, including age and gender. The extracted data were anonymized to maintain patient privacy and confidentiality.
4. Data Validity and Reliability: To ensure the validity and reliability of the data, Dino's father, being an experienced neurological doctor, verified the accuracy of the extracted information. He cross-checked the recorded ages and gender to minimize any potential data entry errors.
5. Ethical Considerations: Throughout the data collection process, ethical considerations were followed. Patient confidentiality and privacy were strictly maintained. The researchers adhered to ethical guidelines and regulations, ensuring that the data collection process was conducted in an ethical and responsible manner.

3.2 Inference

Chi-square of independence

Conditions:

Random: We randomly pick 72 patients who caught a Parkinson disease in a row from the medical record of Beijing Tiantan hospital

Sample size: expected value in each age group is above 5

| 序号 | 年龄 | 性别 |
|----|----|----|
| 1 | 54 | M |
| 2 | 70 | M |
| 3 | 68 | M |
| 4 | 74 | F |
| 5 | 61 | M |
| 6 | 77 | F |
| 7 | 79 | M |
| 8 | 63 | M |
| 9 | 52 | M |
| 10 | 80 | F |
| 11 | 81 | M |
| 12 | 68 | F |
| 13 | 75 | F |
| 14 | 67 | M |
| 15 | 57 | M |
| 16 | 64 | M |
| 17 | 74 | F |
| 18 | 81 | F |
| 19 | 77 | F |
| 20 | 76 | M |
| 21 | 59 | F |
| 22 | 62 | M |
| 23 | 70 | M |
| 24 | 75 | M |
| 25 | 76 | M |
| 26 | 71 | F |
| 27 | 80 | M |
| 28 | 59 | M |
| 29 | 69 | F |
| 30 | 75 | F |
| 31 | 82 | M |
| 32 | 77 | M |
| 33 | 59 | F |
| 34 | 51 | M |
| 35 | 64 | M |
| 36 | 69 | M |
| 37 | 72 | F |
| 38 | 74 | M |
| 39 | 80 | F |
| 40 | 75 | M |
| 41 | 61 | M |
| 42 | 60 | M |
| 43 | 71 | F |
| 44 | 70 | F |
| 45 | 74 | M |
| 46 | 64 | F |
| 47 | 55 | M |
| 48 | 71 | M |
| 49 | 82 | M |
| 50 | 59 | M |
| 51 | 68 | F |
| 52 | 77 | F |
| 53 | 74 | M |
| 54 | 72 | M |
| 55 | 75 | M |
| 56 | 68 | M |
| 57 | 65 | M |
| 58 | 70 | F |
| 59 | 54 | M |
| 60 | 68 | F |
| 61 | 62 | F |
| 62 | 78 | F |
| 63 | 75 | M |
| 64 | 74 | M |
| 65 | 78 | F |
| 66 | 69 | M |
| 67 | 65 | M |
| 68 | 70 | M |
| 69 | 72 | F |
| 70 | 80 | F |
| 71 | 81 | M |
| 72 | 79 | M |

Independence: the sample number is less than 10 percent of the total patient number

Null hypothesis: The age has no relationship with the incidence of the Parkinson diseases. Alternate Hypothesis: The age has relationship with the incidence of the Parkinson Disease.

4 Analysis of data

| Age | Male | Female |
|-----------|-------------|-------------|
| Age 50-75 | E:22.5 O:31 | E:13.5 O:15 |
| Age 75+ | E:22.5 O:14 | E:13.5 O:12 |

Keywords— E :Expected value O:Observed Counts

$$\sum \frac{(31 - 22.5)^2}{22.5} + \frac{(14 - 22.5)^2}{22.5} + \frac{(15 - 13.5)^2}{13.5} + \frac{(12 - 13.5)^2}{13.5} \quad (1)$$

X-square=1.30003

p-value:0.254

Alpha=0.05

Degree of freedom=1

Within the chi-square test, we finally make the conclusion that we fail to reject the null hypothesis, since the p-value is bigger than the significance level= 0.05. There is not enough evidence can support the null hypothesis. This can be shown by the table from the upper side, we can see that the although people below 50 years old only have 4 percent of possibility of having incidence of getting Parkinson. However, through the observational studies, the distribution of incidence of Parkinson is hard to guess and there is not enough evidence to support that the age has the relationship with the incidence of getting Parkinson. Male has larger difference but female has lower so we fail to reject the null hypothesis.

5 Problems in the project/suggestion for future

During the data collection and analysis phase of the project, several difficulties were encountered. Firstly, one of the challenges was the availability of comprehensive and accurate clinical histories of Parkinson's disease patients. Due to privacy concerns and limited access to patient records, it was difficult to gather a large and diverse sample size. This limitation might have affected the generalizability of the findings to the entire Chinese population.

Another difficulty arose in the data analysis process. The provided dataset lacked information on the diagnostic criteria used to confirm Parkinson's disease cases. Without clear diagnostic criteria, there is a possibility of misclassification or inclusion of individuals with other similar conditions, which could introduce bias into the results. In future studies, it would be essential to ensure strict adherence to standardized diagnostic protocols to enhance the accuracy of the findings.

Furthermore, the dataset had an unequal distribution of genders, with a higher representation of males. This gender imbalance might introduce confounding effects when analyzing the relationship between age and Parkinson's disease incidence. A more balanced gender representation would be desirable to obtain a clearer understanding of the age-specific incidence rates in both males and females.

Considering the limitations of the study, there are several recommendations for further work in this area. Firstly, it would be beneficial to expand the study's scope beyond a single hospital or region to include multiple healthcare facilities across China. This would help increase the sample size and improve the representativeness of the findings.

Additionally, future research should incorporate more comprehensive data collection, including genetic factors and potential environmental influences. By examining a broader range of variables, it would be possible to better understand the complex interplay between age, genetics, and environmental factors in the development of Parkinson's disease.

Furthermore, longitudinal studies would provide valuable insights into the temporal relationship between age and Parkinson's disease incidence. Following individuals over an extended period would enable researchers to observe the disease's progression and identify potential risk factors or protective factors associated with aging.

To enhance the rigor of the study, it would be beneficial to employ a matched control group without Parkinson's disease. This would allow for better comparisons between individuals with and without the condition, reducing the potential for confounding variables.

In summary, while the proposed project aimed to investigate the relationship between age and the incidence rate of Parkinson's disease, several difficulties were encountered during data collection and analysis. Recommendations for future work include expanding the study's scope, improving data collection protocols, addressing gender imbalances, and conducting longitudinal studies. These enhancements would contribute to a more comprehensive understanding of the factors influencing the incidence of Parkinson's disease with advancing age.

6 Conclusion

In our experiment, we don't come up to an association between the age and the incident rate of the Parkinson's disease due to confounding variable like gender. However, Most cases of PD occur over the age of 60, suggesting that aging is related to the disease. Some of the research suggest that dopaminergic neurons in the substantia nigra gradually decrease with age in normal adults, which might be the cause of the association of ageing and the incident rate.

References

- [1] Parkinson's Disease on JSTOR — jstor.org. <http://www.jstor.org/stable/29524136>. [Accessed 26-May-2023].
- [2] The second most common neurodegenerative disease: Parkinson's disease - ACROBiosystems — acrobiosystems.com. <https://www.acrobiosystems.com/A1560-The-second-most-common-neurodegenerative-disease> [Accessed 26-May-2023].