Ryan Marinelli
Critical Path Institute

August 09th 2021

Introduction

The purpose of this analysis was to observe the differential effects Alzheimer's Disease (AD) has on the various strata within a given population. In particular, this research endeavour segments the data based on the age of onset of the disease. This is operationalized because… and is based on whether a patient is 75 or older at disease onset, and with associated variation in the Mini-Mental State Exam (MMSE) score to measure mental acuity. The Mini-Mental State Exam (MMSE) is used to evaluate cognitive function among the elderly; it includes tests of orientation, attention, memory, language and visual-spatial skills. The MMSE score is a measure of mental acuity, with a maximum score of 30 points. A score of 20 to 24 suggests mild dementia, 13 to 20 suggests moderate dementia, and less than 12 indicates severe dementia.The data used in this study comes largely from encounter data, which is data that is generated when a patient interacts with a healthcare provider. The most prevalent claim is for general examination with the average number of general examinations per patient is roughly seven. For strokes, there were 1170 encounters with an average of 1.06 encounters per patient. Likewise, for myocardial infarction (MI), the average number of encounters per patient was 1.06. It appears as though encounters that are considered emergencies have this average, or something similar. However, wellness encounters are far more prevalent. When applying encounter data in tandem with other patient specific information, it is possible to derive additional insights on the initial diagnosis of individual patients, the data can be filtered to discern the prevalence of AD across the populations of interest. The analysis was longitudinal in nature and the variable of interest was the age at the time of diagnosis. Given the characteristics of longitudinal data, time-series methods utilizing various multivariate models will be relied upon in the analysis.

Descriptive Analysis

In terms of the prevalence, there were 658 patients with an additional 153 patients with early onset AD.When observing the distribution of AD more generally, the distribution of the data is right-skewed. This is significant in that it poses challenges to using linear models. For instance, OLS will no longer be BLUE under the Gauss Markov assumptions.

Thus a greater trade-off must be made in this study: having a parsimonious model that may be understood or using a less efficient estimator that may introduce bias.

Figure 1.

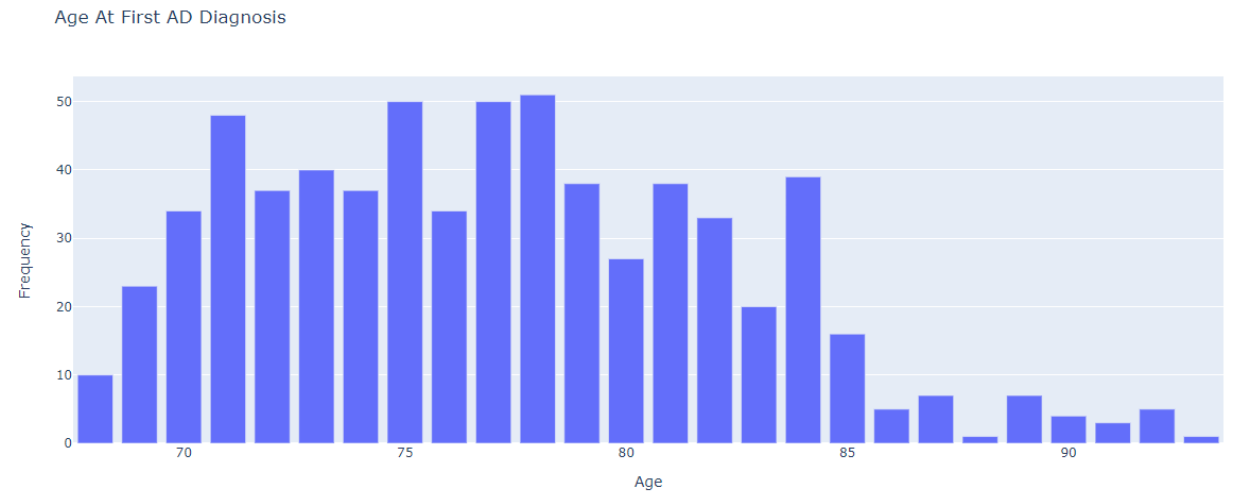


Figure 2 describes that the mean of the distribution appears to be about 78 and the median value being about 75. One facet of the distribution that is noteworthy is the characteristic of the tail. The expectation would be a monotonic property while decreasing with a certain smoothness. This is best reflected in the standard deviation, which was 5.26. The increase in dispersion may be due to the values located at the tail.

Figure 2.

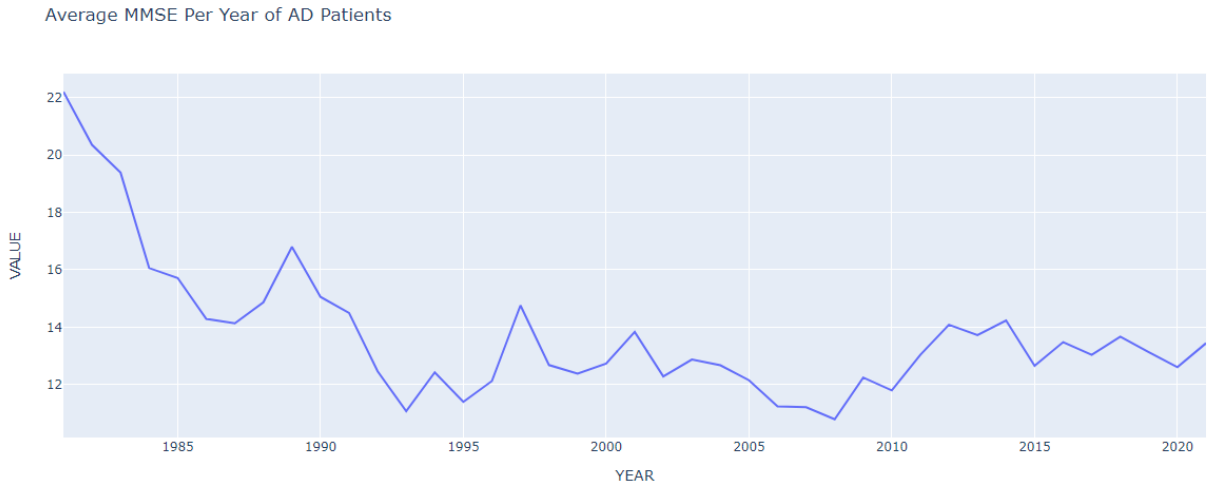
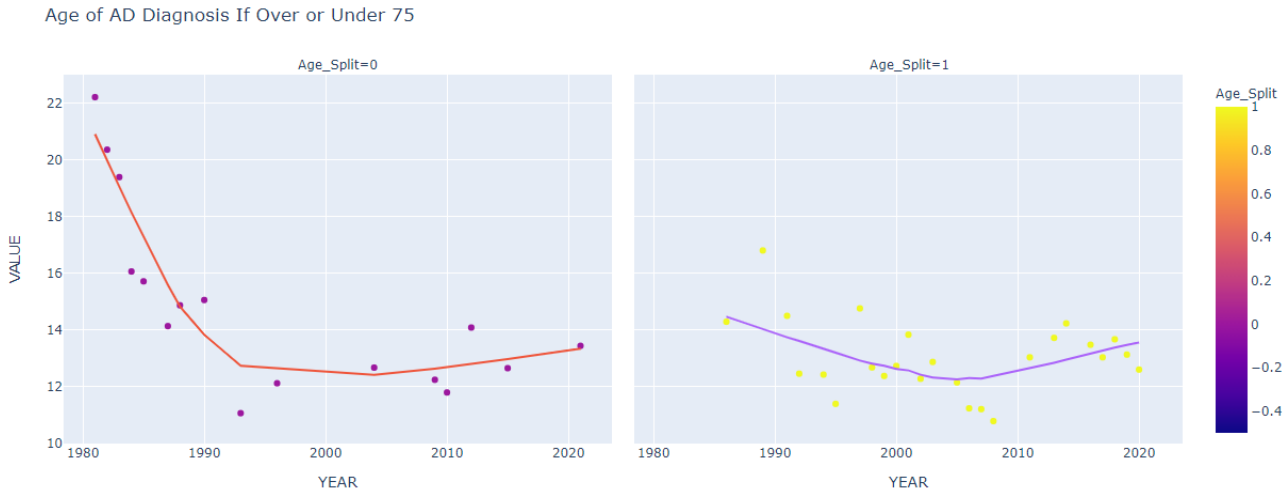


Figure 2 depicts the variation in the average MMSE score per year for patients. The general trend is negative which follows expectations. As one ages, the expectation is that there is a gradual erosion of mental facilities. There are some interesting forces that are being captured in this graph. There was a relatively large decline in the age of diagnosis during the earlier years. The data trends nadir is around the year 2008, after which the MSSE scores begin to increase gradually. What is being reflected is variation in the cohort. As time progresses, those with comorbidities are more likely to succumb to disease. These patients also likely have lower MSSE scores thus lowering the average. As more time passes, the more likely that these patients will be to die and will have lower MSSE further bringing down the average. Also, this effect is likely distributed given the nature of patients with AD. They are not located within the same location, but they are spread amongst various contexts.

Figure 3.



When segmenting this data based on age during the initial diagnosis, this point was further emphasized. There is substantially more curvature in the patients who had an AD diagnosis prior to age 75 than those diagnosed later in life. For those diagnosed at 75 or later, the decline in MMSE score is far more gradual.

Modeling

To further elucidate this variation in demographic groups, several models are applied to investigate. An initial exploratory model is applied with MMSE scores as the dependent variable with years and the independent variable. The beta-coefficient of years is -.011. The

interpretation of this coefficient is that for each year that passes, there is a .011 associated drop in MMSE. An autoregressive model is then fit to observe the variation amongst difference years with lag factors.

$$\widehat{MMSE} = \beta_0 + \beta_1 \widehat{MMSE}_{t-1} + B_2 \widehat{MMSE}_{t-5} + \beta_3 \widehat{MMSE}_{t-10} + \epsilon$$

The table in the below summarizes the results of the lag terms with the associated p-values. The first term is the intercept term followed by the lag years in their respective order.

Beta_Coefficients	P_Value
12.95608342283647	4.3943008107420767e-150
-0.014138717359519994	0.5067860143293506
-0.0026491359103272714	0.9009787903241229
0.015818138979765153	0.45733527250640127

The base MMSE score is about 13. This is the intercept term.The one year lag has a negative association. This suggests that the effects that dampen MMSE scores may linger for prolonged periods. This is substantiated with the negative coefficient for the five-year lag period. While the p-value for this lag is .9, it still is valuable in arguing for the duration of the effects while the effect size itself is a bit more difficult to support. As for the ten-year lag period, it is a positive association. This suggests that the dampening effects tend to disappoint over the period of a decade.

An alternative model this is explored includes age and gender. Gender is encoded in a binary fashion while age is centered. Additionally, a dummy variable is introduced to create a discrete split for those diagnosed before or after age 75. The only consideration in this respect is with multicollinearity; however, it enables for a more parsimonious analysis when attempting to question how that particular threshold affects MMSE scores.

$$\widehat{MMSE} = \beta_0 + \beta_1 \widehat{Gender} + \beta_2 \widehat{Age} + \beta_3 \widehat{Age\,Split} + \epsilon$$

Both gender and age have a beta-coefficient of -04. Age has a significant p-value while gender is nearly significant and should have some weight on the analysis. The way to interpret these cofficents is to consider a man and a woman that are identical in all factors except the woman is a year older. This model suggests that the woman and man would have the same MMSE scores holding all factors equal and a year of aging is about equivalent to the effect of gender.

	coef	std err	t	P>|t|	[0.025	0.975]
const	12.6796	0.375	33.850	0.000	11.945	13.414
Gender_Indicator	-0.0444	0.298	-0.149	0.882	-0.630	0.541
Age	-0.0429	0.044	-0.970	0.332	-0.129	0.044
Age_Split	0.4928	0.491	1.005	0.315	-0.469	1.455

As for answering whether or not there is a significant difference between if a patient has their diagnosis at age 75 or below, the age split variable targets this difference. If a patient has their first diagnosis after the threshold, they are nearly half a MMSE pointer higher holding other factors equal. Also given that the p-value of the split is under 0.5, it suggests that this measure is statistically significant.

An additional model is also tested with an interaction term with age split and gender. The coefficient of the interaction term is negative, but the p-value is only significant with a 90% confidence interval. This suggests that there may be a differentiation in how the segmentation along age of diagnosis may vary across gender.

Conclusion

One of the guiding questions motivating this analysis is to explore how patients that received their first AD diagnosis at age 75 or later differed in terms of their MMSE scores than patients that received a diagnosis of earlier onset AD. In this analysis, it is confirmed that there is a statistically significant difference amongst these groups additionally. However, some of the exploratory analysis with interaction effects suggests that male patients may be affected negatively by later onsets. Future analysis may utilize multilevel models to explore how gender specific effects impact MMSE scores with the aging process of the AD.

Additionally, a piece-weise regression model with a partition to split the groups by threshold may also be advisable to further investigate.