Longitudinal study of magnetic resonances in early onset Alzheimer's disease

Author: Marta Bassa Lladó
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Roser Sala Llonch

Abstract: Alzheimer's disease (AD) is the most common neurodegenerative type of dementia [1]. It is associated with morphometric changes in the human brain such as smaller regional volumes and larger ventricles compared with healthy controls [2]. In this project, we will use volumetric measures extracted from magnetic resonance images (MRI) of early onset AD (EOAD) patients and normal controls (NC) in a longitudinal design. First, we will use statistical methods to find the regions that differ at baseline. Then we will calculate the rate of change for each subject and each region and we will study group differences in this rate of change. Additionally, we will create an algorithm that provides a graphic visualization of the results of both the baseline and the longitudinal analysis.

I. INTRODUCTION

Alzheimer's disease (AD) is the most common neurodegenerative type of dementia [1]. Depending on the onset of AD we differ late onset AD (LOAD) from Early onset Alzheimer's disease (EOAD), which onsets in patients $\leq$ 65 years of age. EOAD comprises about 5 % of AD and it differs from LOAD in many aspects.[3]

The project involves a cross-sectional and a longitudinal design. For that, we will first compare the subcortical volumes and the thickness of the brain cortex between EOAD patients and normal controls subjects (NC). We will do so because EOAD, as a neurodegenerative disorder, can cause changes of morphometric properties of brain structures. Then, for the longitudinal part we will use MRI data acquired at two different timepoints for each subject. With this study we will be able to detect changes in the characteristics of the brain over time.

First of all, we will define some important concepts. The subcortical volumes are the volumes of different areas of the brain. We will study them by using an automated software from MRI. The cortical thickness is defined as the distance between the white/gray and the pial surfaces. These properties are important for the study of EOAD because they can reveal the causative factors of diseases. In addition, the longitudinal studies measuring volumes or cortical thickness could be further implemented in the clinical context in order to evaluate the efficacy of a variety of treatments. [4]

The aim of this project is to study changes in volume and cortical thickness in a sample of EOAD compared with NC and to create a tool that can help in the graphic visualization of the regions of the brain that are more sensitive to morphological changes during EOAD. In order to do that, given a set of measures obtained from semi-automatic procedures, we will use python to create an algorithm which returns the statistical differences as well as graphical representations of data exported from automatic segmentation of MRI data.

A. THE BRAIN

The human brain is a complex organ which interprets the senses, initiates the body movement, seats the intelligence and controls the behavior. It is composed of the cerebrum, the cerebellum and the brainstem. [5]

The cerebrum is the largest part of the brain and it is composed of left and right hemispheres. It is the source of intellectual activities such as reasoning, interpreting the senses, learning, etc.

Each hemisphere can be divided into anatomical regions, each of whom control different functions.[4] We will compare the volume of each anatomical region between EOAD and NC subjects.

Additionally, we will also compare the thickness of the cortex. The cortex is the layer of tissue that coats the surface of the cerebrum and cerebellum. This thin rind contains mainly gray matter. It has hills and valleys which determine the folded appearance of the cortex. [6] Beneath the cortex there is the white matter, made of long nerve fibers (axons) which communicate different parts of the brain. The measured cortical thickness corresponds to the distance between the pial surface (outer layer in the gray matter) and the white matter.

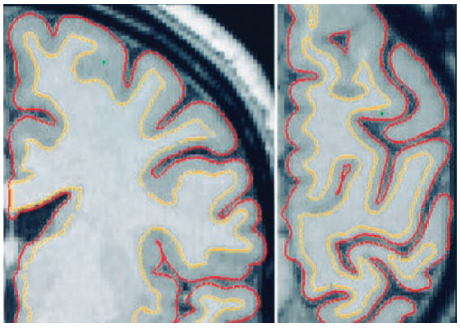

FIG. 1: Slices of the left hemisphere with the gray/white surface (yellow) and the pial surface (red) marked.

B. FREESURFER

Freesurfer is a set of software tools for the study of cortical and subcortical brain anatomy. (http://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferAnalysisPipelineOverview)

The process of manually labeling the different parts of the brain can be done by a trained anatomist but this takes much longer than auto-labeling using Freesurfer.[4] Freesurfer automatically extracts the information required for automating the segmentation procedure. This program uses a probabilistic atlas and compares it with the intensity distributions of the MRI image in order to define the limits of the brain areas. In this atlas, the information regarding statistical properties of anatomical structures is stored in a space in which coordinates have anatomical meaning. [4]

On the other hand, when measuring the cortical thickness from MRI, the Freesurfer's methodology consists in finding the white matter and the pial surface and calculating the thickness as the distance between them [6].

In the context of the current project, a set of tables containing volume measures of the subcortical structures and cortical thickness for the different parcellations (also from a predefined atlas) will be provided. These tables have been obtained from the preprocessing of each individual subject and the visual validation, and manual intervention when needed, of this processing.

II. METHODS

A. Dataset

The data used in this project was acquired at the Alzheimer's and Memory unit of the Hospital Clinic. It comprises a sample of 13 EOAD subjects and 19 NC, each subject acquired two times in a 3T Siemens MRI Scanner with a T1-weighted MPRAGE sequence.

Firstly, a data management will be needed. We will get the data set in a *.csv* format and we will use the panda library from python to create a data frame to work with. For each subject we will have 41 measures of the volumes (one for each region, in mm^3) and 70 measures of the different parts of the cortex.

When comparing volumes, it is important to normalize each region with the total volume. The normalized volume is the volume of each region divided by the total intracranial volume. By doing so, we take into account that the total brain size differs between different subjects.

The table provided, which contains the volume measures, also contains the total intracranial volume for each subject. Therefore, in order to obtain the normalized volume, we will divide each volume measure of each region over the total intracranial volume in each subject.

The demographics are summarized in Table 1. As we can see, there are no significant differences (NS) between the demographics between the two groups. In order to stipulate that, we used a Mann-Whitney U test for the age and the time between scans and a chi-square test for the sex comparison.

	EOAD	NC	p-value
Age, Mean(SD) (years)	61.2 ± 4.8	58.0±4.7	NS
Sex (female/male)	9/4	16/3	NS
Time between scans (years)	1.9 ± 0.3	2.0 ± 0.3	NS

TABLE I: Demographics of the experiment. NS means that there is no significant difference between EOAD subjects and NC subjects in the corresponding variable.

B. Cross-sectional analysis

In this analysis we will compare the volumes of the two groups at a single point of time. This type of analysis is also called baseline analysis. Note that from now on we will present only the results of volumes due to the space limitations, but the same analysis were performed with the cortical thickness measures.

To determine whether the volume of a specific part of the brain statistically differs between NC or EOAD subjects can be seen as to analyze if two sampled groups (volumes of EOAD and NC) are from a single population. This is an example of hypothesis testing. In order to do so, parametric and non-parametric tests could be used.

We have two independently sampled groups and we will determine whether the two groups differ on a single variable. We could use the t-test(parametric test) or the Mann-Whitney U tests (non- parametric). The great difference between these two types of tests is that parametric tests assume a specific distribution to a data set which can be parameterized by a finite number of parameters. [7] The non-parametric tests assume no specific distribution. The use of non parametric tests does not mean that the data completely lack parameters but that the number and nature of the parameters are not fixed in advance. [8]

We will use the Mann-Whitney U test because our N (number of subjects) is not large enough. Therefore, our data sample does not accomplish the parametric assumptions of the t-tests.

The great advantage of the Mann-Whitney U test is that it can be used for small samples of subjects such as the sample of this study. This test can be used for samples from 5 to 20 participants [8] and is analogous to the parametric two-sample t-test, but it compares medians rather than means. [9]

When comparing the two data samples (NC and EOAD) we want to know whether the difference between their central tendency (e.g. mean or median) is statistically significant. The null hypothesis (H_0) of the test is the assumption that both samples were drawn from a population with the same distribution. Therefore, the same population parameters, such as mean or median

[9]. Given the situation where the null hypothesis is rejected, it would indicate that the difference between data samples parameters is significant.

Generally speaking, these tests calculate a test statistic and return a p-value that can be used to interpret the results of the test.

If the p-value is below a significance level (discussed later) then the test says there is enough evidence to reject the null hypothesis. Hence, the samples were drawn from populations with different distributions.

Regarding the significance level (α), many statistical tests use $\alpha = 0.05$ as level of significance of the p-value. But this is the level of significance for only one test. In this case, we are testing the different parts of the brain simultaneously. In general, the more simultaneous inferences we make at one time, the smaller the probability that all inferences are correct. [10]

For an independent experiment of hypothesis testing (as it would be testing just one region of the brain) the probability of false positive is:

$$P = \alpha \tag{1}$$

While the probability of correct inference is then:

$$P = 1 - \alpha \tag{2}$$

But if instead of just one experiment we make n simultaneous experiments this probability is reduced to:

$$P = (1 - \alpha)^n \tag{3}$$

At the same time, the probability of making one incorrect inference (at least) , P_{ii}, is now:

$$P_{ii} = 1 - (1 - \alpha)^n \tag{4}$$

So, the more simultaneous comparisons we make, the more P_{ii} increases.

We will correct this by applying Bonferroni (BFN) and the false discovery rate (FDR) methods, two multiple comparison procedures. These methods provide a larger critical value (α) than the default value.

Consider n hypothesis tests and $p_i ... p_n$ the p-values for these tests. Using the BFN method we will reject the H_0 hypothesis if:

$$p_i < \frac{\alpha}{n} \tag{5}$$

Where α is the threshold used in independent experiments ($\alpha = 0.05$).

On the other hand, FDR is designed to control the proportion of false positives among the rejected hypothesis (i.e. The null hypothesis are actually true in this tests). In the Benjamini-Hochberg (BH) method of FDR we will define:

$$l_i = \frac{i\alpha}{C_n n} \quad \text{and} \quad R = max\Big\{i : P_i < l_i\Big\} \tag{6}$$

Where C_n is defined to be 1 if the p-values are independent and $C_n = \sum_{i=1}^{m}(1/i)$ otherwise. We will call T the BH threshold defined as $T = p_R$ and we will reject the null hypothesis for which $p_i < T$. [7]

Generally, FDR is a more powerful method for correcting multiple comparisons than the BFN method but in this experiment, using the data provided, there is no difference when applying BFN or FDR methods. Both methods are already implemented in the python *statsmodels* package.

C. Longitudinal analysis

Once the parts of the brain which statistically differ between EOAD patients and NC have been determined, we want to analyze the evolution of this parts during time. As explained before, two MRI have been taken to each individual participant in the experiment, with two years between each MRI. For each part of the brain we will create a new variable called symmetrized percent change (SPC). If the volume of the region in the first MRI is Vol_1 and the volume of the region in the second acquisition is Vol_2, then SPC is defined as:

$$SPC = \frac{Vol_2 - Vol_1}{(Vol_1 + Vol_2)/2} \tag{7}$$

SPC is the rate of change in volume with respect of the average volume between the two time-points. We will compute SPC for all the regions of all the subjects. Then, we will use the same test used in the cross-sectional analysis with the assembly of multiple comparison with the SPC data.

III. RESULTS

In this paper, we will only show the results of the analysis of the volumes due to the lack of available space. The results obtained in the cortical thickness study agree with the AD literature. [11]

A. Cross-sectional results

In FIG. 2 one can observe all the regions of the brain. From the total of 41 regions, we indicate the regions with significant differences between EOAD and NC. Two criteria have been considered to mark these regions:

By using a criteria color, we have used the p-values obtained in the Mann-Whitney U test and with a level of significance $\alpha = 0.05$. All the regions which had a p-value below 0.05 have been marked as significant.

In the second criteria we added the correction of multiple comparisons using the FDR method. This is a more rigorous criteria than the first one because, as mentioned

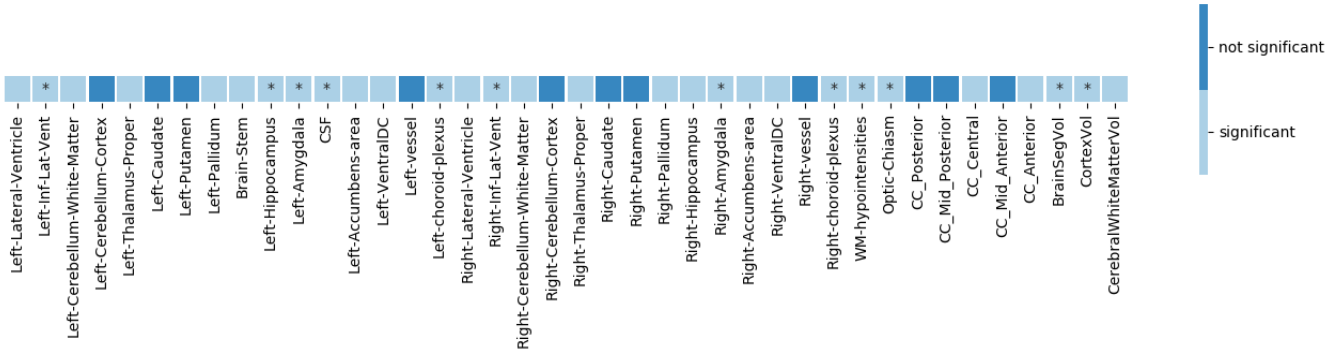

FIG. 2: Parts of the brain marked as if they are statistically different between EOAD and NC (using the initial p-values). Regions with significant difference using the FDR method have been marked with a $\star$ sign.

before, by this we take into account the simultaneous comparisons. In FIG. 2 one can observe the regions statistically different using this criteria with a " $*$ " mark.

It can be seen that the results follow the expected behaviour. We have a lower amount of rejected null hypothesis when using the first criteria than when using the second one. These will be the regions of interest for this study.

B. SPC results

In FIG. 3 we show the results after applying the Mann-Whitney U test with the assemble of multiple comparisons at the SPC values. Here, for simplicity, only the significant parts are shown.

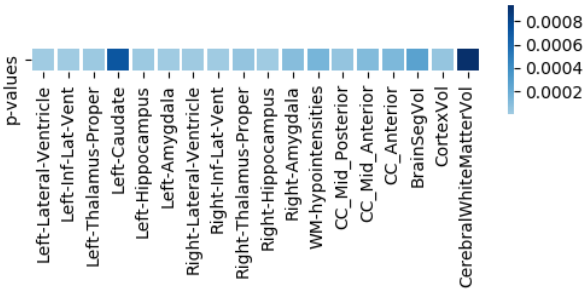

FIG. 3: Areas in which the SPC value differs between EOAD and NC using the FDR method.

It is important to conceive the different regions highlighted in FIG. 2 as statistically different between EOAD and NC are not the same as those observed in FIG. 3. If we study the SPC variable for all regions, we are taking into account the evolution and time factor which are really important in the definition of the disease. Hence, a brain region that does not show statistical differences between EOAD and NC in a cross-sectional study may be significant during disease evolution because its rate of change over time is statistically significant. This shows the importance of longitudinal studies which can reveal important aspects in the development of EOAD. In order to study this in detail, we will now focus on two different brain regions: the left hippocampus and the left lateral ventricle.

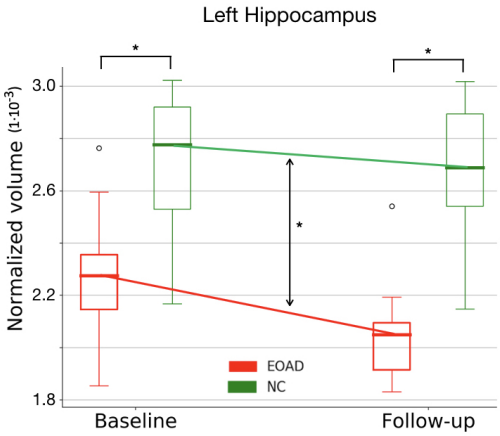

FIG. 4: Representation of the volumes of the left hippocampus for the EOAD and the NC group in the baseline and the follow-up acquisitions. The normalized volume is calculated as the relative volume with respect to the total intracranial volume. Marked whit $\star$ when it shows difference between EOAD and NC.

The hippocampus shows statistical difference in both the cross-sectional and the longitudinal studies. The volume of this part of the brain decreases in time reflecting the loss of gray matter. The EOAD subjects decrease is more remarkable than the NC subjects decrease. This shows one of the effects of the disease in the brain.

On the other hand, the left lateral ventricle only shows statistical difference in the longitudinal study and its volume increases in time. The depletion of gray matter causes the accumulation of more cerebrospinal fluid which is located in the ventricle, causing the increase of this part of the brain.

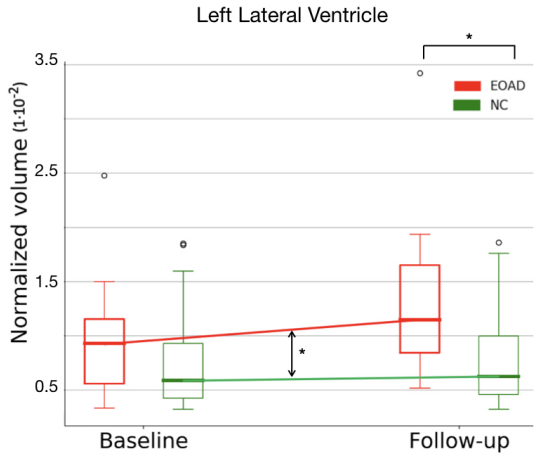

FIG. 5: Volume values of the left lateral Ventricle for the EOAD and the NC group in the baseline and the follow-up acquisitions.The normalized volume is calculated as the relative volume with respect to the total intracranial volume. Marked whit ⋆ when it shows difference between EOAD and NC.

As we can see in FIG. 4 and 5, the healthy group of subjects also show brain differences between the baseline and the follow-up which reflects that normal aging is also associated with hippocampal atrophy and ventricular dilatation [12]. But the NC subjects have a rate of progression over time slower than the EOAD group.

IV. CONCLUSION

We studied the volume of the brain and cortical thickness differences between two groups of subjects. We created an algorithm that provides graphic visualization of the regions of the brain, differs between this two groups and identifies the important regions that should be take into account in the detection and treatment of EOAD.

In order to compare the regions between the two groups of subjects, we considered the use of parametric tests but we used the non-parametric Mann-Whitney U test. This test is more accurate taking into account our data-set. As we hypothesized, this test is appropriate in the velocity and providing the results. That made easier to correct the multiple comparison problem.

After comparing the differences in a baseline, or cross-sectional, analysis we did the same tests but comparing a longitudinal variable. The longitudinal analysis showed different significant parts of the brain than the cross-sectional analysis. From that, we concluded that it is necessary to study both the evolution and time-point changes in the brain due to EOAD. The observation of both analysis side by side help us understand the process of pathology.

It is worth mentioning that this is a preliminary study. Due to the shortage of data sample, we can not generalize our results but we created the correct tool that would provide the significant parts of the brain in the EOAD evolution when used with a larger data sample.

Acknowledgments

I would like to thank my advisor Dra. Roser Sala Llonch for her guidance, support and help through the making of this essay and despite all the circumstances of this semester.

References

[1] Xi Chen Zhu, Lan Tan, Hui Fu Wang, Teng Jiang, Lei Cao, Chong Wang, Jun Wang, Chen Chen Tan, Xiang Fei Meng, and Jin Tai Yu. Rate of early onset Alzheimer's disease: A systematic review and meta-analysis. *Annals of Translational Medicine*, 3(3), 2015.

[2] Matthew C Evans, Josephine Barnes, Casper Nielsen, Lois G Kim, Shona L Clegg, Melanie Blair, Kelvin K Leung, Abdel Douiri, Richard G Boyes, Sebastien Ourselin, and Nick C Fox. Volume changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. *Eur Radiol*, 20:674–682, 2010.

[3] Mario F Mendez. Early-Onset Alzheimer's Disease. 35(2):263–281, 2018.

[4] Bruce Fischl, David H. Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, Albert Montillo, Nikos Makris, Bruce Rosen, and Anders M. Dale. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

[5] M. Sachs. Anatomy of the brain. *Soins; la revue de reference infirmiere*, 27(2):3–8, 1982.

[6] B. Fischl and A. M. Dale. Measuring the thickness of the human cerebral cortex. *NeuroImage*, 9(6 PART II), 1999.

[7] Gonzalo Hechavarría, Rodney; López. *All of statistics*, volume 53. 2013.

[8] Nadim Nachar. The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. Technical Report 1, 2008.

[9] Jason Brownlee. How to Calculate Nonparametric Statistical Hypothesis Tests in Python.

[10] Chris Chatfield. *Statistical Analysis and Data Display*, volume 168. 2005.

[11] Hanna Cho, Seun Jeon, Sue J Kang, Jong-Min Lee, Jae-Hong Lee, Geon Ha Kim, Ji Soo Shin, Chi Hun Kim, Young Noh, Kiho Im, Sung Tae Kim, Juhee Chin, Sang Won Seo, and Duk L Na. Longitudinal changes of cortical thickness in early-versus late-onset Alzheimer's disease. 2013.

[12] Liana G Apostolova, Amity E Green, Sona Babakchanian, Kristy S Hwang, Yi-Yu Chou, Arthur W Toga, and Paul M Thompson. Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment and Alzheimer's disease.