Compositional data analysis to explore the association between joint brain volumetric variation and genetic predisposition to Alzheimer's disease

Patricia Genius[1], *M.Luz Calle*[2], *Raffaele Cacciaglia*[1], *Arcadi Navarro*[1], *Juan Domingo Gispert*[1], *Natalia Vilor-Tejedor*[1]

pgenius@barcelonabeta.org, [1]Barcelonaβeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain. malu.calle@uvic.cat, [2]Biosciences Department, Faculty of Sciences, Technology and Engineering, University of Vic-Central University of Catalonia, Vic, Spain. rcacciaglia@barcelonabeta.org,[1]Barcelonaβeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain. anavarro@barcelonabeta.org, [1]Barcelonaβeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain. jdgispert@barcelonabeta.org,[1]Barcelonaβeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain. nvilor@barcelonabeta.org, [1]Barcelonaβeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain.

Genetic characterization of brain volumes (i.e. imaging genetic studies; IG) is often assessed and analyzed using correlation and univariate modelling. However, given the structural heterogeneity of the brain, it is likely that correlated brain regions may vary together in morphometric characteristics, and genetics may play a role in this joint variation. We propose a new modelling strategy for volumetric analysis in IG studies based on the use of compositional data analysis (CoDA).

The CoDA approach focuses on the relative variation between brain features (components) instead of analyzing them separately. In this study, a CoDA methodology (*coda4microbiome* algorithm) was used, which involves modelling and variable selection (elastic net penalized regression) to identify brain structural variation signatures involving brain features that vary together based on the genetic predisposition to specific neurological conditions. Specifically, in this study, the focus was on Alzheimer's disease (AD) and different disease stages.

The study included middle-aged participants from the ALFA (Alzheimer's and FAmilies) and ADNI (Alzheimer's Disease Neuroimaging Initiative) studies who had magnetic resonance imaging scans, genetic information and cerebrospinal fluid-amyloid status. Individuals were classified into different groups based on the disease-stage and amyloid status: cognitively unimpaired amyloid-beta negative (CU) Aβ- (N=220), CU Aβ+ (N=118), (mild cognitive impaired) MCI Aβ+ (N=230) and AD Aβ+ (N=100). We used the *coda4microbiome* algorithm to analyze the joint structural variation of specific brain components (optimal brain signature) that was most closely associated with higher genetic predisposition to AD (Polygenic Risk score of AD; PRS-AD) (High genetic predisposition to AD: PRS-AD $\geqslant$ quantile 0.8). The elastic net penalty term was defined through a cross-validation procedure. After reparameterization, the brain signature was expressed as a weighted sum of the selected variables in the form of a log-contrast function (Equation 1). The joint change of the brain structural variation signature and the genetic predisposition to AD was assessed through disease-stage stratified logistic regression models adjusted for age and sex.

$$\text{(Equation 1) Signature} = \sum_{j=1}^{K} \hat{\theta}_j \cdot \log(x_j), \text{ where } \sum_{j=0}^{K} \hat{\theta}_j = 0$$

Through the application of the *coda4microbiome* algorithm, we found disease-stage specific joint volumetric variations associated with higher genetic predisposition to AD. These variations were related to specific memory network regions of the brain. Our study highlights the potential of implementing CODA methods to address issues in the neuroimaging research area and provides new insights into the genetics of AD.

Keywords: Compositional Data Analysis; Imaging Genetics; Coda4microbiome.