UPPSALA
UNIVERSITET

COMPARING SUPERVISED AND UNSUPERVISED METHODS FOR OPERATIONALIZING RESILIENCE USING BRAIN AGE GAP

Submitted by

Rashiq Al Tariq

A thesis submitted to the Department of Statistics
in partial fulfillment of the requirements
for a two-year Master of Science degree in Statistics
in the Faculty of Social Sciences

Supervisor

Rauf Ahmad

Spring, 2025

ABSTRACT

Brain maintenance and cognitive reserve describe complementary resilience mechanisms supporting preserved cognition in aging. This study evaluates the Brain Age Gap (BAG) as a biomarker for these resilience mechanisms by comparing supervised and unsupervised grouping methods in 429 individuals with MRI (Magnetic Resonance Imaging), cortical thickness, and Mini-Mental State Examination (MMSE) data at ages 70 and 75. Supervised "theory-based" groups, defined from BAG and MMSE, are analyzed with Linear Mixed Effects models to assess trajectories of BAG and cortical thickness. Unsupervised "data-driven" groups are derived from K-means and Gaussian Mixture Model clustering applied to paired observations of the resilience variables BAG, MMSE, and cortical thickness. Theory-based analysis captures brain maintenance and cognitive reserve but not a distinct non-resilient group. Clustering partly aligns with these constructs but mainly reflects the cohort's overall homogeneity. Results support the potential of BAG as a resilience biomarker, whilst highlighting the need for further refinement of the theory-based approach.

Key words: Resilience mechanisms, Linear mixed effects models, Cluster analysis

Contents

1 Introduction

Resilience refers to the body's capacity to resist or adapt to functional and cognitive decline following health stressors, such as age-related brain pathologies or chronic conditions (Stern et al., 2020; Whitson et al., 2016). As such, it is a broad concept applicable across various aging contexts. This thesis project focuses on resilience- the ability of the brain to maintain cognition and functional abilities despite aging and disease. It implements the novel Resilience in Aging and Dementia framework, proposed by Stern and colleagues in 2023 (Stern et al., 2023), which defined resilience as encompassing any process that supports cognitive preservation in the face of accumulating neuropathology. Such processes include malleable, complementary mechanisms that leverage brain plasticity to cope with changes like amyloid and tau accumulation (Alzheimer's disease characteristics), other proteinopathies, and cerebrovascular disease during the preclinical phase of dementia (Badji et al., 2023; Jack Jr et al., 2024).

In Stern's resilience framework, two major mechanisms are presented: Brain maintenance (**BM**) and Cognitive reserve (**CR**). **BM** is defined as "the relative absence of changes in neural resources or neuropathologic change over time as a determinant of preserved cognition in older age.". Thus, individuals who exhibit brain maintenance are those who display minimal changes in their cognitive function and brain health over time. **CR** is defined as "a property of the brain that allows for cognitive performance that is better than expected given the degree of life-course related brain changes and brain injury or disease". **CR** exemplifies an exceptional ability of the brain to "cope with or compensate for changes in the brain and the consequences of brain injury or disease.". Thus, individuals who exhibit cognitive reserve are those who, despite changes in brain health over time, have been able to maintain cognitive function. **BM** and **CR** are theoretically sound and have clear distinctions. However, the key challenge of how to measure resilience persists. The disadvantage to current methods of measuring resilience is that they fail to capture the core biological dimension of resilience. However, this can be achieved through the utilization of deep learning methods.

The Brain Age Gap (*BAG*) is defined as the difference between an individual's "brain age" and one's chronological age, where brain age is defined as an estimate of the underlying hypothetical age of an individual, also known as biological age. Using brain age as a biomarker for biological aging was first introduced by (Cole and Franke, 2017). In the paper, they discuss the application of machine learning techniques to predict the biological age of an individual using high dimensional neuroimaging data e.g. Magnetic Resonance Imaging (MRI) scans. The paper presents advantages, and criticisms, of utilizing deep learning methods alongside neuroimaging to further research within brain age analysis. The authors emphasize the fact that taking advantage of advancements in artificial intelligence and statistical computation could have positive implications for our ability to develop and evaluate methods for preventing neurological disease.

There have been numerous models developed for the purpose of predicting brain age (Bocancea et al., 2021; Cole et al., 2017; Franke et al., 2010). The estimate of brain age used in this study was derived from the deep learning model presented and developed in-house at Karolinska Institute by (Dartora et al., 2024). The specific type of model employed was a convolutional neural network (CNN) using T1-weighted MRI scans as the input. Model performance was assessed using the Mean Absolute Error (MAE) between the predicted brain age and the chronological age of a sample of cognitively healthy individuals. Since the sample represented a population of individuals without any cognitive impairment, their biological age should theoretically be relatively close to their chronological age, hence an MAE closer to 0 would suggest good predictive performance. However, due to known and unknown brain heterogeneity in ageing, and the lack of a gold-standard measure of biological age, an MAE different from zero is expected. The work of Dartora et al, 2024, had an MAE of 2.67 yrs, showing good model performance compared to existing counterparts.

(Marseglia et al., 2024) were the first to introduce the concept of using the *BAG* as a biomarker for resilience. In the paper, the authors showcase a practical example of how the in-house developed CNN model can be used to estimate the *BAG* and examine potential factors associated with differences in *BAG* values within a population, such as physical activity and sex. For this thesis, the hypothesis to be tested is the proposal of being able to distinguish **BM** and **CR** based on the *BAG* metric. If the *BAG* is a valid measure of resilience, theoretically we could be able to distinguish **BM** and **CR** in a population based on its value. If that were the case, younger-appearing brains would indicate preserved brain structure, which would indicate **BM**. Older-appearing brains, in the absence of significant cognitive decline, would indicate exceptional coping abilities linked to **CR**. If it can be shown that the *BAG* can be used to reliably identify resilience mechanisms within a study population, it would concretize the functionality of using the *BAG* as a measure of resilience whilst also being an indication of the usefulness of utilizing artificial intelligence-based technologies for the purpose of brain age analysis.

1.1 Aims and research questions

In this thesis, the term *operationalize* refers to the process of translating the theoretical constructs of **BM** and **CR** into measurable statistical entities that can be analysed quantitatively. Specifically, this involves using *BAG*, *MCT*, and *MMSE* as observable "resilience variables" from which resilience groupings can be defined. The primary aim of this thesis is to compare supervised and unsupervised statistical strategies for grouping individuals in the context of resilience research, using said resilience variables. The supervised approach is represented by an expert-defined, theory-based grouping of individuals into resilience groups, followed by statistical modelling using linear mixed effects (LME) models. The unsupervised approach is represented by cluster analysis methods, including K-means and Gaussian Mixture

Models (GMM). The goal is to compare the two approaches of identifying the resilience mechanisms of **BM** and **CR** from the Brain Age Gap in a community-based cohort of Swedish septuagenarians, i.e., individuals aged between 70 and 79 (a detailed data description is provided in section 2.1).

Research questions:

1. Do longitudinal LME models provide statistical evidence that the Brain Age Gap can distinguish between Brain Maintenance and Cognitive Reserve?
2. What group structures emerge from unsupervised clustering of the resilience variables?
3. To what extent do the supervised theory-based and unsupervised data-driven approaches produce overlapping classifications?

2 Methods

2.1 Data

The sample was derived from the *Gothenburg H70 Study - Birth Cohort 1944* (Rydberg Sterner et al., 2019), including 429 individuals aged 70-years-old, with available brain magnetic resonance imaging (MRI) examination at baseline (2016-2019) and 5-year follow-up (2021-2024). The dataset contains important information for the purpose of operationalizing **BM** and **CR**, such as longitudinal measures of cognitive ability and brain pathology. Table 1 shows the relevant variables used for the theory-based operationalization and the data-driven clustering along with descriptive statistics. As per the study design, each variable had repeated measures for two distinct time points.

A challenge that can arise when studying resilience in a population is that of age-related bias. If a study population contains individuals of differing ages, these individuals will naturally show different levels of resilience due to their differences in chronological age. Thus, it is difficult to isolate factors, other than aging, that affect the different expressions of resilience in the population. All subjects in this cohort being of the same age at each examination reduces the potential age-related bias that can occur when studying patterns of resilience.

Figure 1 includes a flowchart illustrating the study population selection for analysis for the current study. For the estimate of the *BAG* to be reliable, the MRI scans used to calculate the brain age need to pass a standard quality control; if the MRI scans fail the quality control, they are deemed as unreliable. Thus, individuals with poor quality MRIs were excluded from the analysis. A point of contention surrounding the legitimacy of using the *BAG* as a measure of resilience relates to the fact that one's level of resilience heavily depends on an individual's history of neurological diseases. Neurological diseases directly impact brain health and are likely to lead to worsened states of resilience in older age. The goal

Table 1: Descriptive Statistics of H70 Data (N = 429).

	Mean	sd	Median	Min	Max	Range
BAG_{v1}	0.514	1.492	0.300	-3.500	6.900	10.400
BAG_{v2}	-0.109	2.064	-0.300	-6.400	9.600	16.000
MCT_{v1}	2.395	0.072	2.398	2.109	2.572	0.462
MCT_{v2}	2.382	0.081	2.383	2.100	2.583	0.483
$MMSE_{v1}$	29.186	1.080	30.000	24.000	30.000	6.000
$MMSE_{v2}$	28.818	1.456	29.000	20.000	30.000	10.000

Note. v1 = Baseline examination; v2 = Follow-up examination; BAG = Brain Age Gap; MCT = Mean Cortical Thickness; MMSE = Mini-Mental State Examination

of this study is to disentangle resilience mechanisms displayed in individuals who exhibit normal aging; hence it is of interest to isolate the possible factors and resilience mechanisms related to preserving cognitive function as a part of the normal aging process. Individuals who have previously been diagnosed with neurological diseases such as brain cancer, multiple sclerosis, or dementia already present a severe condition of cognitive impairment potentially caused by factors unrelated to the normal aging process, e.g., genetic predisposition to certain diseases. Therefore, including these individuals will lead to a biased sample that does not represent the target population. Thus, individuals who have been diagnosed with previous neuropathological diseases (Parkinson's, Multiple sclerosis, Hydrocephalus, Brain cancer, Dementia, Epilepsy) are excluded from the analysis. The resulting sample consisted of 429 individuals with the following data at baseline and follow-up: reliable brain MRI examinations, examination of cognitive function, and an absence of neurological disease.

The variables that were used in the following analysis alongside the *BAG* were Mean Cortical Thickness (*MCT*) and Mini-Mental State Examination (*MMSE*). The Brain Age Gap is defined as:

$$Brain\ Age\ Gap\ (BAG) = Predicted\ Brain\ Age\ (PBA) - Chronological\ Age\ (CA)$$

where the PBA is the predicted or estimated biological age of an individual. In the data used, each individuals PBA was estimated from a CNN model based on the ResNet Architecture (Dartora et al., 2024; He et al., 2016) with their individual brain MRI as the input. The BAG was then calculated by subtracting each individuals PBA by their corresponding CA i.e. their actual age. Since all individuals in the data were of the same age at examination, 70 at baseline and 75 at follow-up, no age-bias correction was deemed necessary and therefore not applied. Cortical thickness is a measure of the width of the outer most layer of the brain, the cortex. The *MCT* is the average cortical thickness between the right

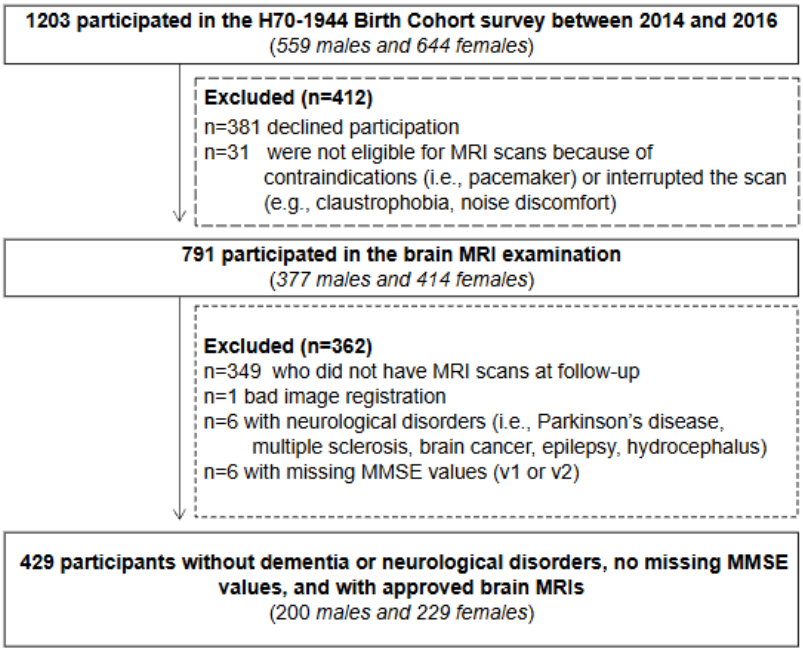


Figure 1: Flowchart of the study selection process

and left hemispheres of an individual's brain, which was obtained using Freesurfer 7.2.0, an open-source neuroimaging toolkit used for processing, analyzing, and visualizing human MRI. The *MMSE* is a test of cognitive function that is measured on a point scale of 0-30, with a value of less than or equal to 23 being an indication of dementia (Woodward and Galea, 2005).

2.2 Operationalization of resilience to identify groups

This study seeks to determine whether it is possible to reliably identify different resilience mechanisms based on the *BAG*. Therefore, methodologically, it is of interest to identify and evaluate distinct groups of individuals based on their expressions of resilience , i.e., resilience groups. To identify and evaluate different resilience groups, two methods will be used: the theory-based and the data-driven. Finally, the results from these two approaches will be compared. Figure 2 depicts the analysis process for the methods applied.

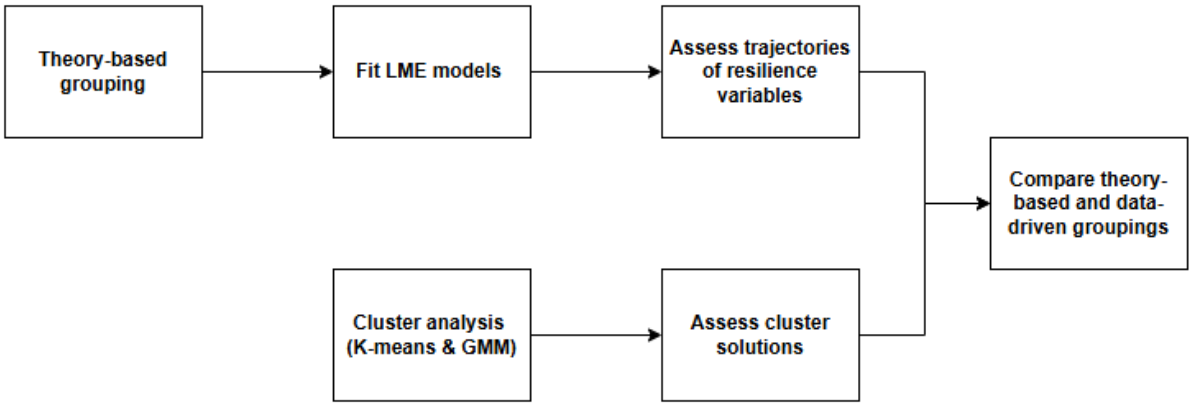


Figure 2: Flowchart of the analysis process

2.2.1 Method 1: Theory-based approach

The theory-based approach is comprised of two main parts:

1. Splitting of data into three subsets each representing a corresponding resilience group.

2. Fitting linear mixed effects models for each subset of data or resilience group

Firstly, individuals were assigned to a specific resilience group based on conditions defined by their examined values of *BAG*, *MCT*, *MMSE*. These conditions were determined based on previous knowledge of the characteristics of brain maintenance and cognitive reserve outlined in the resilience framework by Stern. Hence, the theory-based approach can be seen as a supervised strategy where the response, the resilience variables, are modeled based on domain-specific expert knowledge.

The three subsets of data, and thus the three resilience groups, were labeled as: **BM**, **CR**, and **No resilience**. The **BM** resilience group consisted of individuals who display younger appearing brains. Whether a brain appears younger or older depends on the sign of the *BAG* value associated with each individual. A negative *BAG* implies that one's predicted brain age is less than one's actual age, thereby displaying features of a brain that is younger than expected, indicating an absence of brain pathology. Because BAG values are calculated as predicted age minus chronological age, they are centred around zero by design, which represents the expected value in normal aging. Negative values therefore reflect younger-appearing brains relative to age norms, which aligns conceptually with the definition of brain maintenance. Thus, individuals with $BAG < 0$ were assigned to the **BM** subgroup.

The **CR** resilience group consisted of individuals who have older appearing brains (positive *BAG*) and stable *MMSE* scores between baseline examination (denoted as $v1$) and follow-up examination (denoted as $v2$). For the purpose of this study, a stable change in *MMSE* is determined as a decrease in *MMSE* from baseline to follow-up less than 1.5 (denoted as $\Delta MMSE > -1.5$ where $\Delta MMSE = MMSE_{v2} - MMSE_{v1}$). The threshold of -1.5 points has been commonly used as a cut-off in cohort

studies to identify objective cognitive impairment (Dove et al., 2021; Winblad et al., 2004) and has been proposed as an indicator for minimal clinically important difference (MCID) of reliable cognitive change (Borland et al., 2022). Reported thresholds for clinically meaningful decline in MMSE scores typically range between 1-3 points depending on population characteristics and follow-up length. To test robustness, a sensitivity analysis was performed using $\Delta MMSE = -1.0$ and $\Delta MMSE = -3.0$, see section (A.1) in the Appendix.

An individual having a stable *MMSE* score over time is an indication of preserved cognitive ability. The combination of stable cognitive function despite an older appearing brain due to brain pathology exemplifies the concept of cognitive reserve. Thus, individuals with $BAG > 0$ and $\Delta MMSE > -1.5$ were assigned to the **CR** group. Individuals who did not get assigned to either **BM** or **CR** were assigned to **No resilience**. The conditions for splitting the data are summarized in Table 2.

Table 2: Theory-based grouping criteria for resilience mechanisms

Group	Definition
BM	Individuals with $BAG < 0$
CR	Individuals with $BAG > 0$ and $\Delta MMSE > -1.5$
No resilience	Individuals not assigned to **BM** or **CR**

Once all the individuals were assigned to a corresponding resilience group, the trajectories of *BAG* and *MCT* were examined for each group using linear mixed effects models. The purpose of estimating the trajectories was to assess whether the longitudinal trends of the resilience variables within each resilience group behaved predictably based on the theoretical expectations of the resilience framework.

2.2.2 Linear Mixed Effects Models

Linear models can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y}$ is a $K \times 1$ vector of responses, $\mathbf{X}$ is a $K \times m$ design matrix, $\boldsymbol{\beta}$ is a $m \times 1$ vector of coefficients and $\boldsymbol{\epsilon}$ is a $K \times 1$ vector of error terms. Linear models assume that the data are independent and the role of $\boldsymbol{\beta}$ is simply to capture one source of variation, the "fixed" effects of covariates on the response. Linear Mixed Effects (LME) Models extend the standard linear model by allowing for a second type of variation, random effects of the covariates on the response. The addition of random effects allows units (individuals or clusters) to deviate from the fixed-effects parameters. Thus, LME models consider both between-unit (described by fixed effects) and within-unit (described by individual

or cluster-specific random effects) variation in the data. Fitting an LME model, as opposed to a standard linear model, might be appropriate if data are assumed to exhibit some kind of dependency, for example, due to a hierarchical structure or longitudinal observations. A general expression of LME models can be given as

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j + \boldsymbol{\epsilon}_j, \;\; j = 1, ..., N \tag{2}$$

where $\mathbf{b}_j$ is a vector of random effects, $\mathbf{Z}_j$ is a design matrix for random effects and N are the number of units. In this application of LME models, $\boldsymbol{\epsilon}_i$ and $\mathbf{b}_j$ are assumed to be normally distributed as $\boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{I})$ and $\mathbf{b} \sim N(0, \sigma^2\mathbf{D})$ where σ^2 is the residual error variance, $\mathbf{I}$ is the identity matrix and $\mathbf{D}$ contains the random-effect variance components.

By stacking vectors and block-diagonal matrices

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_N \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{bmatrix}.$$

the compressed form of (2) is given by

$$\mathbf{y} \;=\; \mathbf{X}\boldsymbol{\beta} \;+\; \mathbf{Z}\mathbf{b} \;+\; \boldsymbol{\epsilon}, \tag{3}$$

where

$$\begin{bmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim N\left(\mathbf{0},\; \sigma^2 \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right).$$

and

$$\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, \qquad \mathrm{Var}(\mathbf{y}) = \sigma^2\Big(\mathbf{Z}\mathbf{D}\mathbf{Z}^\top + \mathbf{I}\Big),$$

hence the marginal distribution of $\boldsymbol{y}$ can be expressed as

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta},\; \sigma^2[\mathbf{I} \;+\; \mathbf{Z}\mathbf{D}\mathbf{Z}^\top]) \tag{4}$$

where the covariance component $\sigma^2[\mathbf{I} \;+\; \mathbf{Z}\mathbf{D}\mathbf{Z}^\top])$ considers the two separate sources of variation in the model, the residual error $\boldsymbol{\epsilon}$ and random effects variation $\mathbf{b}$. Estimation of ($\boldsymbol{\beta}$, σ^2, $\mathbf{D}$) are performed by restricted maximum likelihood (RMLE). Given the marginal distribution in (4) the log-likelihood (ignoring constants) is

$$\ell(\boldsymbol{\beta}, \sigma^2, \mathbf{D}) = -\tfrac{1}{2}\left\{n\ln\sigma^2 + \ln|\mathbf{V}| + \tfrac{1}{\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}, \tag{5}$$

where $\mathbf{V} = \mathbf{I} + \mathbf{Z}\mathbf{D}\mathbf{Z}^\top$ and n is the total sample size.

The RML function for the LME model can be obtained by integrating out the fixed effects $\boldsymbol{\beta}$. The restricted log-likelihood is

$$\ell_R(\sigma^2, \mathbf{D}) = -\tfrac{1}{2}\left\{(n-m)\ln\sigma^2 + \ln|\mathbf{V}| + \ln|\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}| + \tfrac{1}{\sigma^2}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})^\top\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})\right\}, \quad (6)$$

where $m = \mathrm{rank}(\mathbf{X})$ and

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{y} \quad (7)$$

is the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$.

Maximization in σ^2 yields the RMLE estimator

$$\widehat{\sigma}^2_{\mathrm{RMLE}} = \frac{1}{n-m}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})^\top\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}). \quad (8)$$

Note that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2_{\mathrm{RMLE}}$ are valid estimators only if $\mathbf{V}$ is known. The variance component matrix $\mathbf{D}$ is then estimated by maximizing $\ell_R(\sigma^2, \mathbf{D})$ with respect to $\mathbf{D}$, typically using iterative algorithms such as Fisher scoring (Demidenko, 2013).

The LME models used in this study were fitted using the **lmer** function from the R package **lme4**, with the variance parameters estimated by restricted maximum likelihood (RMLE) via the BOBYQA algorithm (Powell et al., 2009). The BOBYQA (Bound Optimization BY Quadratic Approximation) algorithm is a derivative-free optimization method and is particularly suited for likelihood maximization in mixed-effects models, where analytic gradients may be difficult to obtain. The fitted LME models were used to estimate the change in the resilience variables between baseline ($Time = 0$) and follow-up ($Time = 1$) within the theory-based resilience groups while allowing for subject-specific (individual) deviations from the population mean trajectory.

In the general mixed model formulation given in (3), the response vector $\mathbf{y}$ corresponds to repeated measures of either *BAG* or *MCT* for each individual. The design matrix $\mathbf{X}$ encodes the fixed intercept and the fixed effect of *Time*, whereas the random-effects design matrix $\mathbf{Z}$ encodes subject-specific deviations (random intercepts or random slopes). Here, the *Time* variable is a binary indicator of measurement occasion (0 for baseline and 1 for follow-up), so that its regression coefficient captures the mean change in the response between baseline and follow-up.

Two specifications of the model were considered: a random-intercept model and a random intercept and slope model. In the random-intercept model, only subject-specific intercepts are included, so that $\mathbf{D}$ reduces to the scalar variance component $D_{\text{Intercept}}$. In the random intercept and slope model, both subject-specific intercepts and subject-specific slopes with respect to *Time* are included, with variance components

$$\begin{bmatrix} D_{\text{Intercept}} & 0 \\ 0 & D_{\text{slope}} \end{bmatrix},$$

where the correlation between the random effects was set to 0. The correlation between the random effects was set to zero to avoid identifiability issues. With only two time points per person, each individual contributes just one degree of freedom for change, which makes it statistically impossible to estimate both random intercept and random slope variances along with their correlation. This is because there would be more random-effect parameters to estimate than observations available in the data. Setting the correlation to zero circumvents this lack of information while still allowing for individual variability in both baseline levels and change over time.

For subject $i = 1, \ldots, n$ at occasion $j = 1, 2$, the two models were:

$$Y_{ij} = (\beta_0 + b_{0i}) + \boldsymbol{\beta}_G^\top \text{Group}_i + \beta_T \text{Time}_{ij} + \boldsymbol{\beta}_{G\times T}^\top (\text{Group}_i \times \text{Time}_{ij}) + \epsilon_{ij}, \tag{9}$$

$$Y_{ij} = (\beta_0 + b_{0i}) + \boldsymbol{\beta}_G^\top \text{Group}_i + (\beta_T + b_{1i}) \text{Time}_{ij} + \boldsymbol{\beta}_{G\times T}^\top (\text{Group}_i \times \text{Time}_{ij}) + \epsilon_{ij}, \tag{10}$$

where Group_i is a vector of dummy indicators for the three resilience groups (**BM**, **CR**, and **No resilience**), and $\boldsymbol{\beta}_G$ and $\boldsymbol{\beta}_{G\times T}$ are vectors of the corresponding group and interaction effects. The purpose of the interaction effects in the model is to formally test differences in baseline values as well as change over time between resilience groups. In (9), b_{0i} is the random intercept for subject i; in (10), b_{0i} and b_{1i} are the random intercept and random slope for subject i, assumed normally distributed with mean zero, uncorrelated, and with variances $\sigma^2_{b_0}$ and $\sigma^2_{b_1}$. The residuals ϵ_{ij} are assumed independent and normally distributed with mean zero and variance σ^2.

The random-intercept model estimates the mean trajectories of *BAG* and *MCT* while allowing each individual to deviate in their baseline level. The random intercept and slope model extends the random intercept model with the addition of allowing individuals to have their own unique linear trajectory between baseline and follow-up. These estimates were examined to determine whether the global trajectories of *BAG* and *MCT* behaved in line with what one would expect for the different resilience groups and whether the subject-specific random effects showed notable individual variation in the data.

2.2.3 Method 2: Data-driven approach

The data-driven approach aimed to identify distinct groupings in the data based on *BAG*, *MCT*, and *MMSE*. This aim was fulfilled by the use of two cluster analysis methods; K-means clustering and Gaussian Mixture Model clustering. The two observed examinations of each variable (baseline - $v1$ and follow-up - $v2$) were treated as separate observed variables, resulting in 6 variables for the analysis: BAG_{v1}, BAG_{v2}, MCT_{v1}, MCT_{v2}, $MMSE_{v1}$, $MMSE_{v2}$. In contrast to the theory-based approach, these

groups were not decided based on predetermined conditions; instead, the goal was to assess whether groupings based on the resilience variables emerge naturally in the data. If the two methods provide similar results in terms of grouping, it would imply that the proposed grouping conditions provided by the theory-based approach are supported in the data by the fact that similar groupings can be identified not only when predefined. The data-driven approach can be seen as the unsupervised strategy counterpart to that of the supervised theory-based approach.

2.2.4 K-means

In the data-driven approach, each individual $i = 1, \ldots, n$ is represented by a six-dimensional feature vector

$$x_i = \big(\textit{BAG}_{i,v1},\ \textit{BAG}_{i,v2},\ \textit{MCT}_{i,v1},\ \textit{MCT}_{i,v2},\ \textit{MMSE}_{i,v1},\ \textit{MMSE}_{i,v2}\big)^\top \in \mathbb{R}^6,$$

where, for example, $\textit{BAG}_{i,v1}$ denotes the Brain Age Gap at baseline for individual i. Thus the dataset is $\{x_1, x_2, \ldots, x_n\}$ with $n = 429$ individuals. The goal of K-means clustering is to partition the dataset into K disjoint clusters $C_1, \ldots, C_K$ by minimizing the within-cluster sum of squares (WSS),

$$J = \sum_{j=1}^{K} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2, \tag{11}$$

where $\mu_j \in \mathbb{R}^6$ is the centroid of cluster C_j and $\|\cdot\|$ denotes the Euclidean norm.

The standard K-means algorithm proceeds iteratively: starting from randomly selected initial centroids $\mu_1(0), \ldots, \mu_K(0)$, each datapoint is assigned to its nearest centroid, and the cluster centroids are then recomputed as

$$\mu_j(t+1) = \frac{1}{|C_j(t)|} \sum_{x_i \in C_j(t)} x_i,$$

where $C_j(t)$ is the set of datapoints assigned to cluster j at iteration t. This reassignment-update cycle continues until convergence, i.e. until cluster assignments no longer change. The algorithm is summarized in Algorithm 1.

Algorithm 1 K-means

1. Select K-points as the initial centroids $\mu_1(0), \mu_2(0), ..., \mu_k(0)$
2. Assign each data point to its closest centroid by minimizing $J = \sum_{j=1}^{K} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2$
3. Recompute the centroids of the K-clusters as $\frac{1}{|C_j(t)|} \Sigma x_i$
4. Repeat steps 2 and 3 until the centroids do not change.

Before initiating the K-means algorithm the number of clusters needs to be specified. Determining the optimal number of clusters is not always straightforward, as most diagnostics used to set this number are accompanied by a degree of subjectivity. One common way to decide the number of clusters is by performing a Scree test, which involves examining a so-called Scree plot. For the Scree-plot, the K-means algorithm is performed with an increasing number of clusters k. For each one of these initializations of the K-means algorithm, the WSS is calculated. Then, the WSS for each specification of k is plotted against k. The number of clusters is decided by inspecting where the WSS no longer decreases substantially for a subsequent increase in k, or in other words, where an "Elbow" in the visualization is formed. Another method for determining the number of clusters is performing a "Silhouette analysis". For each observation i, the silhouette width

$$s(i) = (b(i) - a(i)) \,/\, max\{a(i), b(i)\} \tag{12}$$

was computed, where $a(i)$ denotes the average distance to all other points in the same cluster and $b(i)$ the smallest average distance to any other cluster. The mean silhouette width $\bar{s}(k)$ was then averaged over all observations. The number of clusters k retained was chosen as the value of k that maximized $\bar{s}(k)$. Finally, a formal test was also used to determine the optimal number of clusters, the Gap statistic (Tibshirani et al., 2001), which compares the pooled WSS ($\log(W_k)$) with the expected WSS ($E\{\log(W_k)\}$) derived from a uniform distribution i.e. reference-distribution for different numbers of clusters k. The Gap statistic is defined as

$$\mathrm{Gap}(k) = E\{\log(W_k)\} - \log(W_k) \tag{13}$$

The optimal number of clusters is chosen as the smallest k such that $\mathrm{Gap}(k) \geq \mathrm{Gap}(k+1) - SD(\mathrm{Gap}(k+1))$.

The purpose of applying K-means in this analysis was to identify clusters of individuals with distinct profiles of the resilience variables. Since *BAG* and *MCT* are continuous, and *MMSE* is commonly treated as approximately continuous in epidemiological and neurobiological research (Koch et al., 2005; Maltais et al., 2015), all six variables were suitable for Euclidean distance-based clustering. Prior to clustering, all variables were standardized to have mean zero and unit variance to ensure that differences in scale did not disproportionately influence the results. The resulting cluster assignments were subsequently compared with the theory-based resilience groups in order to assess the degree of overlap between the supervised and unsupervised approaches. Since the resilience variables include repeated measures at two time points, they are likely to be correlated. Standard K-means does not account for correlation, so it was used here primarily as a heuristic baseline method, with model-based clustering (outlined in section 2.2.5) providing a more flexible alternative by modelling full covariance structures.

The K-means algorithm for clustering was employed using the kmeans function included in the **stats** package in R. The kmeans function includes a parameter specifiying the number of random initializations (`nstart`) of the algorithm. For example, with `nstart` = 5, the full K-means algorithm is run independently five times with different initial centroids. The final solution is then chosen as the one with the smallest WSS out of the five initializations. An increased number of random starts decreases the risk of ending up with a solution with a poor local minimum and increases the chance of finding the "best" clustering solution. The number of random initializations was set to 50 to ensure valid cluster results.

2.2.5 Finite Mixture Models

Model-based clustering is an alternative to heuristic clustering methods that provides the opportunity to utilize probability models to partition the data. Finite Mixture Models (FMM) are a type of probability model that, within the context of clustering, assumes that each cluster corresponds to an independent probability distribution. In other words, FMMs assume that data are comprised of a mixture of probability distributions. In this framework, determining the initial number of clusters and clustering method become a model selection issue that is typical for formal statistical analysis. FMM clustering was thus utilized as a probabilistic counterpart to the heuristic K-means clustering. The reason for including two different data-driven methods for the clustering was to examine whether differences in the characteristics of each method would result in substantial discrepancies in terms of cluster assignments.

A common variant of model-based clustering is Gaussian Mixture Model (GMM) clustering. As before with K-means, each individual $i = 1, \ldots, n$ is represented by the six-dimensional feature vector

$$x_i = \big(\mathit{BAG}_{i,v1},\ \mathit{BAG}_{i,v2},\ \mathit{MCT}_{i,v1},\ \mathit{MCT}_{i,v2},\ \mathit{MMSE}_{i,v1},\ \mathit{MMSE}_{i,v2}\big)^\top \in \mathbb{R}^6,$$

so that the dataset consists of $\{x_1, \ldots, x_n\}$ with $n = 429$ individuals where independence between individuals is assumed. The modeling goal is to identify latent clusters of individuals who share similar profiles across the resilience variables. The GMM clustering was performed using the **mclust** package in R which provides the necessary functions for clustering and model diagnostics.

Under a Gaussian Mixture Model with G components (clusters), the marginal density of each datapoint is

$$f(x_i) = \sum_{k=1}^{G} \pi_k \, \phi(x_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{14}$$

where π_k are the mixing proportions i.e., the probability of an observation belonging to component k ($\pi_k > 0$, $\sum_{k=1}^{G} \pi_k = 1$), and $\phi(\,\cdot\,; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ denotes the d-dimensional Gaussian density with mean

vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. The full likelihood for the sample is therefore

$$L(\{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^{G}) = \prod_{i=1}^{n} \sum_{k=1}^{G} \pi_k \, \phi(x_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{15}$$

The model parameters $(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ are estimated by maximum likelihood using the Expectation-Maximization (EM) algorithm (Fraley and Raftery, 2002).

The covariance structure of the model determines the geometric shape of the clusters. For example, $\boldsymbol{\Sigma}_k = \lambda \mathbf{I}$, where $\lambda > 0$ is a scalar variance parameter and $\mathbf{I}$ is the identity matrix, implies spherical clusters of equal size. If $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$, the covariance is constant across clusters, so they share the same geometric shape but may differ in location. In contrast, an unrestricted $\boldsymbol{\Sigma}_k$ allows each cluster to have its own shape, orientation, and volume. The **mclust** package implements a broad range of such covariance parameterizations, as detailed in (Fraley and Raftery, 2007, p.3).

When performing cluster analysis, it is of interest to decide upon an optimal "best" model. The best model according to the **mclust** implementation is determined by fitting several models, through maximum likelihood parameter estimation using the EM algorithm, with different numbers of clusters and covariance structures. In this study, models with all available covariance structures were fitted, and the optimal number of components G and covariance parameterization were selected according to the Bayesian Information Criterion (BIC). For this implementation the BIC is defined as

$$\text{BIC} = 2\ell(\hat{\theta}) - p \log(n), \tag{16}$$

where $\ell(\hat{\theta})$ is the maximized log-likelihood of the model, p is the number of estimated parameters, and n is the sample size. The first term rewards models that fit the data well, while the second term penalizes model complexity. This trade-off is essential in the context of Gaussian Mixture Models, since the likelihood always increases with additional mixture components. Without such a penalty, model selection would systematically favor more clusters. Compared to other criteria such as AIC, the BIC imposes a stronger penalty on additional parameters and is therefore particularly suitable for clustering applications where avoiding overfitting is crucial. The model specification with the highest BIC was chosen as the optimal GMM solution. The resulting GMM solution provided a data-driven clustering that could be compared to the theory-based resilience groups to assess the degree of overlap and divergence

2.2.6 Adjusted Rand Index (ARI)

To evaluate the stability of the data-driven clustering solutions, the Adjusted Rand Index (ARI) was used (Hubert and Arabie, 1985). The ARI is a chance-corrected measure of similarity between two partitions of the same set of n observations and is widely used to assess clustering reproducibility.

In this study, the reference clustering solution was first obtained using the full dataset for both K-means and GMM. Bootstrap resampling was then performed ($B = 200$), and each clustering method was re-fit to the bootstrap sample. The resulting cluster assignments were compared to the reference clustering on the overlapping individuals using the ARI, which measures the proportion of subject pairs that are classified together (or apart) in both partitions, adjusted for the agreement expected by chance. (Random assignment)

Let $U = \{U_1, \ldots, U_r\}$ and $V = \{V_1, \ldots, V_s\}$ be two partitions (e.g., reference and bootstrap solutions) of the same n individuals, with n_{ij} denoting the number of individuals assigned to both cluster U_i and cluster V_j. These counts can be represented in a contingency table:

	V_1	V_2	$\ldots$	Row sums
U_1	n_{11}	n_{12}	$\ldots$	$n_{1\cdot}$
U_2	n_{21}	n_{22}	$\ldots$	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
U_r	n_{r1}	n_{r2}	$\ldots$	$n_{r\cdot}$
Column sums	$n_{\cdot 1}$	$n_{\cdot 2}$	$\ldots$	n

Here, $n_{i\cdot} = \sum_j n_{ij}$ and $n_{\cdot j} = \sum_i n_{ij}$. The ARI is then given by:

$$\text{ARI} = \frac{\sum_{i,j} \binom{n_{ij}}{2} - \left[\sum_i \binom{n_{i\cdot}}{2} \sum_j \binom{n_{\cdot j}}{2}\right] \Big/ \binom{n}{2}}{\left[\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2}\right] \Big/ 2 - \left[\sum_i \binom{n_{i\cdot}}{2} \sum_j \binom{n_{\cdot j}}{2}\right] \Big/ \binom{n}{2}} \tag{17}$$

The ARI ranges from -1 (less agreement than expected by chance) to 1 (perfect agreement), with 0 indicating chance-level agreement.

3 Results & Discussion

The following sections report the results obtained from the analysis of the theory-based and data-driven approaches.

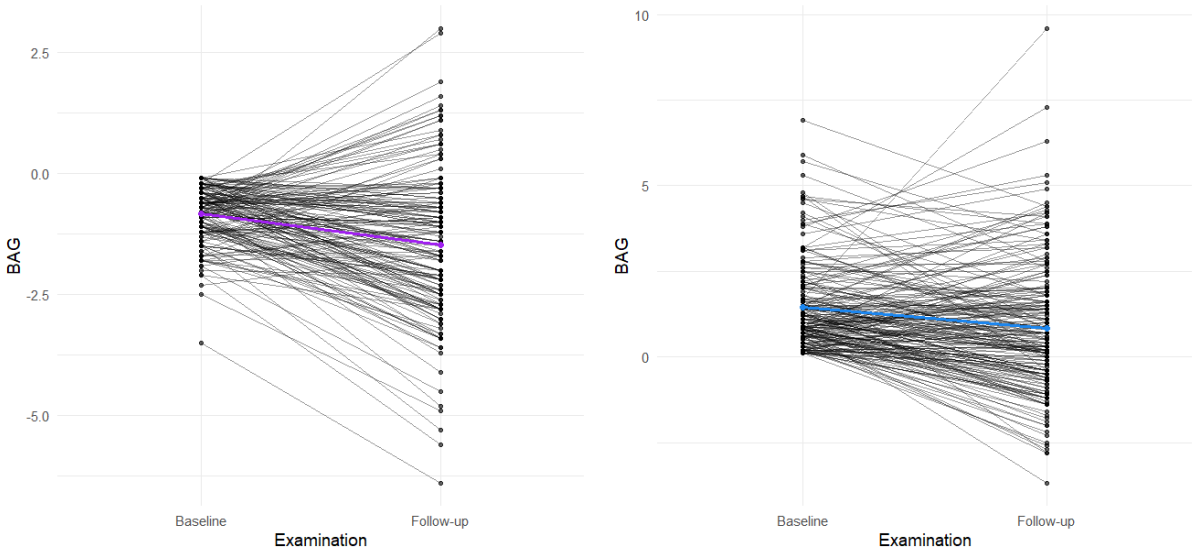
3.1 Theory-based approach

With the theory-based conditions formulated in section 2.2.1, the data was split into three distinct resilience groups. Table 3 contains the number of individuals within each resilience group. The size of each group is 168, 207, 54 for the **BM**, **CR** and **No resilience** groups respectively.

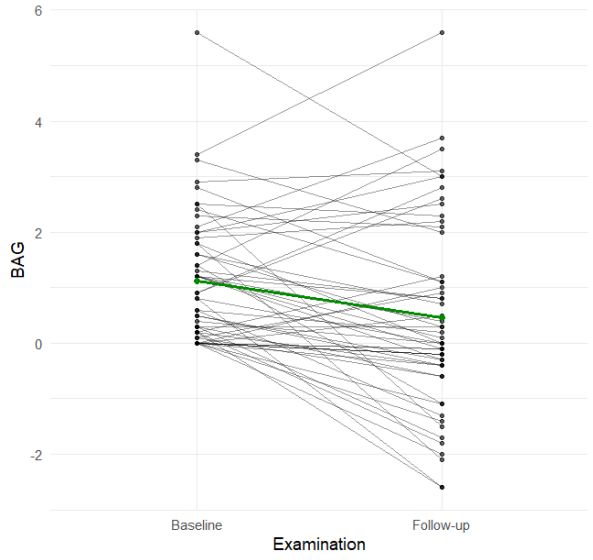
Figure 3 depicts the change in *BAG* in the three resilience groups from baseline to follow-up. Though there is individual variation in the trajectories, the average change is declining among all resilience groups.

Table 3: Number of individuals in each resilience group

BM	CR	No resilience
168	207	54



(a) **BM** (b) **CR**

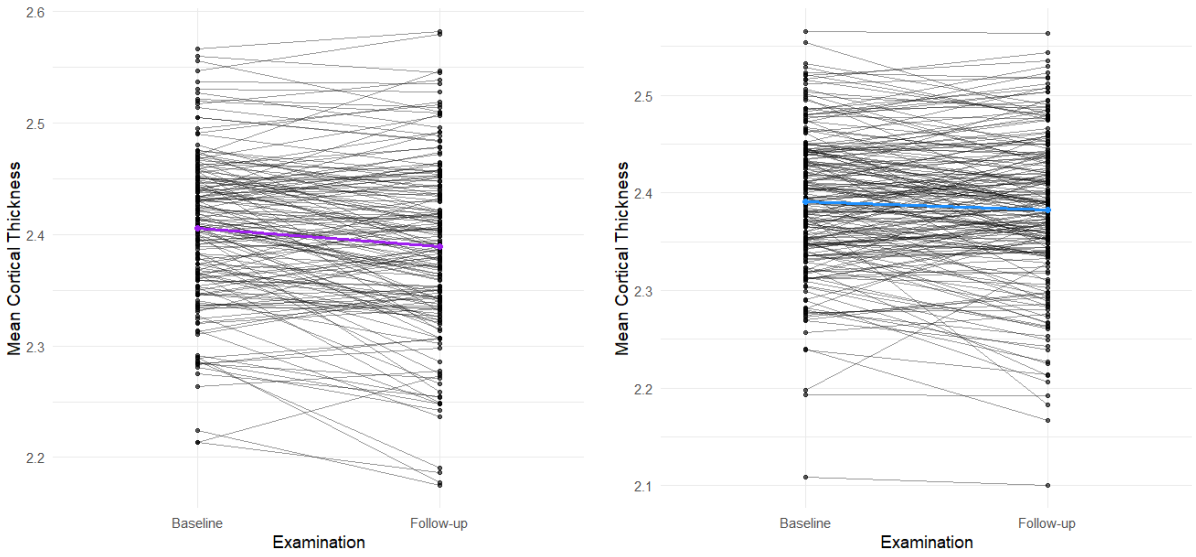


(c) **No-resilience**

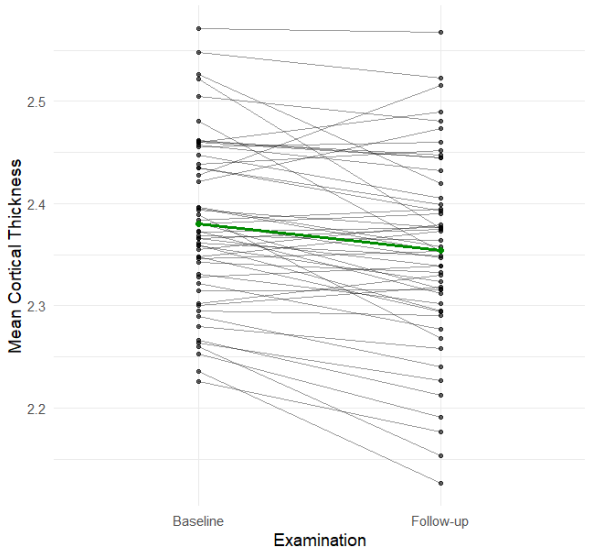
Figure 3: Change in Brain Age Gap (BAG) from baseline to 5-year follow-up. Black lines = individual trajectories; Colored line = group average trajectory.

Figure 4 depicts the individual and average trajectories of *MCT* within each of the resilience groups. All of the resilience groups show stable average trajectories of *MCT* from baseline to follow-up, whilst there are some deviating individual trajectories with slightly increasing or decreasing values of *MCT*

between examinations.



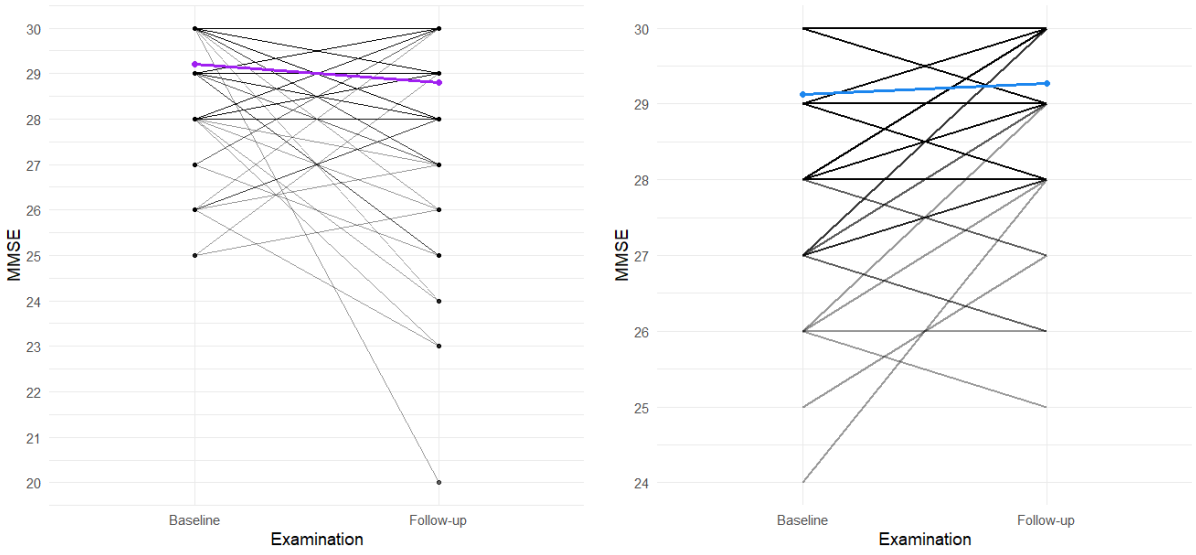
(a) **BM** (b) **CR**



(c) **No-resilience**

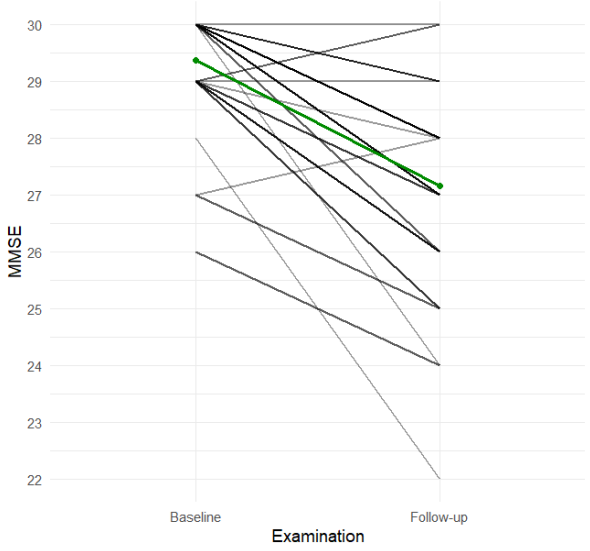
Figure 4: Change in Mean Cortical Thickness (MCT) from baseline to 5-year follow-up. Black lines = individual trajectories; Colored line = group average trajectory.

Figure 5 depicts the individual and average trajectories of *MMSE* within each of the resilience groups. The **BM** group seems to have a stable average trajectory of *MMSE* between baseline and follow-up, although there are some individuals within this group who experienced substantial decreases in *MMSE* values between examinations. The **CR** group also has a stable average trajectory that is slightly positive. The **No resilience** group has the largest change in *MMSE* between examinations and is decreasing. This is expected for this resilience group since the theoretical condition for being assigned as **No resilience** included having a decrease in *MMSE* between examinations greater than 1.5.



(a) **BM**

(b) **CR**



(c) **No-resilience**

Figure 5: Change in Mini-Mental State Examination (MMSE) from baseline to 5-year follow-up. Black lines = individual trajectories; Colored line = group average trajectory.

Table 4 presents the results of the model comparison between the random intercept model and the extended random intercept and slope model. A likelihood ratio test (LRT) was used to compare the two nested models for the two response variables (*BAG* and *MCT*). The tests yielded statistically significant p-values ($p < 2.2 \times 10^{-16}$ for *BAG* and $p < 1.34 \times 10^{-5}$ for *MCT*). This indicates that including random slopes for *Time* significantly improves the model fit relative to a random intercept only specification.

Table 4: Model comparison : **RI** vs **RI+RS**

Outcome	**Model**	**AIC**	**BIC**	**Deviance**	**LRT** p
BAG	RI	2982.3	3020.3	2966.3	–
BAG	RI+RS	2838.1	2880.9	2820.1	$< 2.2 \times 10^{-16}$
MCT	RI	-2480.8	-2442.8	-2496.8	–
MCT	RI+RS	-2497.8	-2455.0	-2515.8	1.34×10^{-5}

Note. RI = Random intercept, RS = Random slope, LRT = Likelihood ratio test.

Table 5: Fixed effects for LME models with *BAG* as response

	Estimate	**95% CI**	**df**	t	p
RI model					
Intercept	-0.827	[-1.050, -0.610]	722.0	-7.46	$<$ .001
Intercept (**CR**)	2.272	[1.975, 2.570]	722.0	15.22	$<$ .001
Intercept (**No resilience**)	1.948	[1.505, 2.392]	722.0	8.66	$<$ .001
Time	-0.638	[-0.872, -0.404]	426.0	-5.36	$<$ .001
CR $\times$ *Time*	0.038	[-0.277, 0.352]	426.0	0.24	.814
No resilience $\times$ *Time*	-0.025	[-0.499, 0.449]	426.0	-0.10	.918
RI + RS model					
Intercept	-0.827	[-0.984, -0.671]	426.0	-10.41	$<$ .001
Intercept (**CR**)	2.272	[2.059, 2.484]	426.0	21.25	$<$ .001
Intercept (**No resilience**)	1.948	[1.636, 2.259]	426.0	12.09	$<$ .001
Time	-0.638	[-0.872, -0.404]	426.0	-5.36	$<$ .001
CR $\times$ *Time*	0.038	[-0.277, 0.352]	426.0	0.24	.814
No resilience $\times$ *Time*	-0.025	[-0.499, 0.449]	426.0	-0.10	.918

Note. RI = Random intercept, RS = Random Slope. **BM** is the reference group. CIs are 95% Wald intervals. Intercept (**CR**) and Intercept (**No resilience**) is the difference in baseline from the reference group (**BM**).

Table 6: Fixed effects for LME models with *MCT* as response

RI model					
	Estimate	**95% CI**	**df**	t	p
Intercept	2.4053	[2.3936, 2.4171]	504.1	410.47	$< .001$
Intercept (**CR**)	-0.0143	[-0.0298, 0.0012]	504.1	-1.82	.070
Intercept (**No resilience**)	-0.0248	[-0.0482, -0.0015]	504.1	-2.09	.037
Time	-0.0158	[-0.0225, -0.0091]	426.0	-4.64	$< .001$
CR $\times$ *Time*	0.0073	[-0.0018, 0.0163]	426.0	1.58	.115
No resilience $\times$ *Time*	-0.0109	[-0.0245, 0.0027]	426.0	-1.58	.115
RI + RS model					
Intercept	2.4053	[2.3940, 2.4165]	426.0	436.50	$< .001$
Intercept (**CR**)	-0.0143	[-0.0289, 0.0003]	426.0	-1.93	.054
Intercept (**No resilience**)	-0.0248	[-0.0468, -0.0029]	426.0	-2.22	.027
Time	-0.0158	[-0.0225, -0.0091]	426.0	-4.64	$< .001$
CR $\times$ *Time*	0.0073	[-0.0018, 0.0163]	426.0	1.58	.115
No resilience $\times$ *Time*	-0.0109	[-0.0245, 0.0027]	426.0	-1.58	.115

Note. RI = Random intercept, RS = Random Slope. **BM** is the reference group. CIs are 95% Wald intervals. Intercept (**CR**) and Intercept (**No resilience**) is the difference in baseline from the reference group (**BM**).

Table 7: Random effects variance and standard deviation for LME models

Response	**Model**	**Component**	**Variance**	**SD**
BAG	RI	Intercept	0.8766	0.9362
	RI+RS	Intercept	0.8766	0.9362
		Slope (*Time*)	2.0108	1.4180
MCT	RI	Intercept	0.0047923	0.06923
	RI+RS	Intercept	0.0047923	0.06923
		Slope (*Time*)	0.0013353	0.03654

Note. RI = Random intercept, RS = Random slope.

Table 5 includes the fixed-effects regression estimates for the LME models, with *BAG* as the response and with **BM** as the reference group for the group and interaction effects. The baseline value (intercept) for the **BM** group is significantly below 0 ($\hat{\beta} = -0.827, 95\%$ CI$[-1.050,\ -0.610], p < .001$). Compared with **BM**, the **CR** group showed substantially higher baseline *BAG* ($\hat{\beta} = 2.27$, 95% CI[1.98, 2.57], $p < .001$), as did the **No resilience** group ($\hat{\beta} = 1.95,\ 95\%$ CI$[1.51,\ 2.39],\ p < .001$). There was a significant effect of *Time* within the **BM** group ($\hat{\beta} = -0.638, 95\%$ CI$[-0.872, -0.404],\ p < 0.001$), indicating a significant decrease in *BAG* from baseline to follow-up. The group-by-time interaction terms were not significant for either **CR** ($p = 0.814$) nor **No resilience** ($p = 0.918$), suggesting no evidence that the trajectories differed between groups. The change in *BAG* was negative for all resilience groups (-0.638, -0.601, -0.663), indicating an average decline in *BAG* between baseline and follow-up examinations regardless of resilience group.

Table 6 presents the fixed-effect regression estimates from the LME models with *MCT* as the response, using **BM** as the reference group for the group and interaction effects. At baseline, the estimated mean *MCT* for the **BM** group was ($\hat{\beta} = 2.4053,\ 95\%$ CI$[2.3936,\ 2.4171],\ p < .001$). The **CR** group did not differ significantly from **BM** at baseline ($\hat{\beta} = -0.0143,\ 95\%$ CI$[-0.0298,\ 0.0012],\ p = .070$), and the **No resilience** group showed a slightly lower baseline mean *MCT* compared to **BM** ($\hat{\beta} = -0.0248,\ 95\%$ CI$[-0.0482,\ -0.0015],\ p = .037$). There was a significant effect of *Time*, indicating a significant decline in *MCT* between baseline and follow-up for **BM** ($\hat{\beta} = -0.0158$, 95% CI[-0.0225, -0.0091], $p < .001$). The interaction effects were non-significant (**CR**×*Time*: $\hat{\beta} = 0.0073$, 95% CI[-0.0018, 0.0163], $p = .115$; **No resilience**×*Time*: $\hat{\beta} = -0.0109,\ 95\%$ CI$[-0.0245,\ 0.0027],\ p = .115$), indicating no evidence that the rate of decline differed significantly between groups. Overall, the results suggest very small differences in baseline *MCT* values between resilience groups and a small but statistically significant decline in *MCT* from baseline to follow-up.

Table 7 includes the random effect estimates for the random intercept and the random intercept and slopes models for both response variables. These random effects reflect the subject-specific variation for the entire data set. For the models with *BAG* as the response, the random intercept is (0.9362) indicating that the variation in baseline for individuals in the data is close to, but less than, 1 year in terms of BAG. The random slopes standard deviation is (1.4180), indicating notable subject-specific variation in the trajectory of BAG from baseline to follow-up, corresponding to slightly more than 1 year in terms of BAG. For the models with MCT as the response, the standard deviations for the random intercepts are (0.06923) and (0.03654) for the random slopes. In a clinical context these values are small, hence, these results indicate that there are not any substantial subject-specific differences in terms of baseline values of *MCT* or trajectory of *MCT* between examinations for the individuals in this data.

Overall, these findings support the potential of the theory-based operationalization of **BM** and **CR**

using the *BAG*. This can be seen by the examination of the trajectories of the resilience variables within each resilience group, where the *BAG* and *MCT* within these two groups behaved in the way we would expect them to according to the conceptual framework by Stern and colleagues. For both the **BM** and **CR** groups, the *BAG* showed a significant decrease between baseline and follow-up, indicating that these individuals exhibit resilience since their brain age is becoming younger relative to their chronological age over time.

Furthermore, the trajectories of *MCT* within the **BM** group were found to be stable over time, confirming the status of the individuals in this group as "Brain maintainers". This was verified by the results of the LME estimates. Although the effect of *time* on mean cortical thickness was statistically significant, the absolute magnitude of the change is very small and unlikely to be clinically meaningful. Furthermore, resilience is not expected to eliminate age-related decline entirely, but rather to attenuate it. It is expected that even healthy individuals would show some degree of decline in cortical thickness between ages 70 to 75. Thus, a small and stable detectable decline in cortical thickness is consistent with resilience.

The **CR** group also exhibited stable *MCT* trajectory. In addition, the **CR** group fulfills the requirement of greater than expected cognitive function despite older appearing brain, since the individuals assigned to this resilience group were those who did not have a decrease in *MMSE* between examinations that exceeded the specified threshold of -1.5. Thus, the results capture the complementary roles of brain maintenance and cognitive reserve: **BM** reflects structural stability of the brain while **CR** reflects preserved function of the brain in the presence of structural disadvantage.

As for the **No resilience** group, they had, by design, a clinically meaningful degree of decline of *MMSE* between baseline and follow-up. Per this design, the expectation would be that these individuals would show a trajectory of *BAG* indicating poor development in terms of brain health i.e. a positive slope. However, the *BAG* in the **No resilience** group, like the other resilience groups, showed an average decline in the *BAG* between baseline and follow-up. These results suggest that individuals in this data, regardless of resilience group assignment, on average appear to have healthier looking brains at follow-up relative to baseline. Two possible reasons as to why this is the observed effect relate to the characteristics of the data used. The data used consists of septuagenarians with reliable MRI scans at baseline and follow-up examinations. To undergo an MRI scan can be seen as a relatively strenuous procedure, especially for elderly individuals, since you are constrained in a tight space for a prolonged period of time. This fact would naturally make it so that those subjects who have successfully undergone an MRI scan at both 70 and 75 years old would likely be healthier than those who could not. In addition, MRI scans cannot be performed for individuals who have prostheses or pacemakers. Thus, it is probably the case that this data is biased towards generally healthier individuals, resulting in trajectories of *BAG* that show more

favorable brain development regardless of resilience group assignment. It could also be the case that cross-sectional values of *BAG* might not be sufficient in their ability to express all forms of resilience. If this is the case, it would be of interest to further develop the theory-based conditions for resilience group assignment to include longitudinal aspects of *BAG* as a measure of resilience.

3.2 Data-driven approach

For the data-driven approach, two cluster models, K-means and GMM, were chosen. Both models were initialized with BAG, MCT, and MMSE, with baseline and follow-up values for these measures treated as separate variables.

3.2.1 K-means

Figure 6 depicts the change in within cluster sum of squares (WSS) across different specifications of the number of clusters k in the k-means solution. From the figure, there does not seem to be a clear elbow being formed at any point, indicating that there is no obvious point where the addition of another cluster no longer substantially lowers the WSS. Figure 7 depicts the average silhouette width for different specifications of the number of clusters k in the k-means solution. In this case, the cluster solution with k = 2 received the highest average silhouette width (approximately 0.39), indicating that a two-cluster solution might be optimal. Figure 8 depicts the estimated gap statistic for different numbers of clusters k. From the figure it can be seen that $Gap(1) \approx 0.72$ and that $Gap(2) \approx 0.705$, furthermore $SD(Gap(2)) \approx 0.015$ meaning that $Gap(1) \geq Gap(2) - SE(Gap(2))$. Based on the gap statistic, the optimal number of clusters is 1, suggesting that the data is comprised of one single mass as opposed to several distinct clusters.

The scree-plot and gap statistic results indicate that clustering might not be suitable for this data. However, with this in mind, it is still of interest to apply some kind of clustering solution in order to examine any potential groupings in the data that could correspond to the clinical constructs of the theory-based approach. Therefore, as per the silhouette analysis, a 2 cluster solution will be applied with the K-means algorithm. It is important to note that because these diagnostics suggest a weak or no cluster structure, the identified clusters will have to be interpreted as tentative or "weak" clusters.

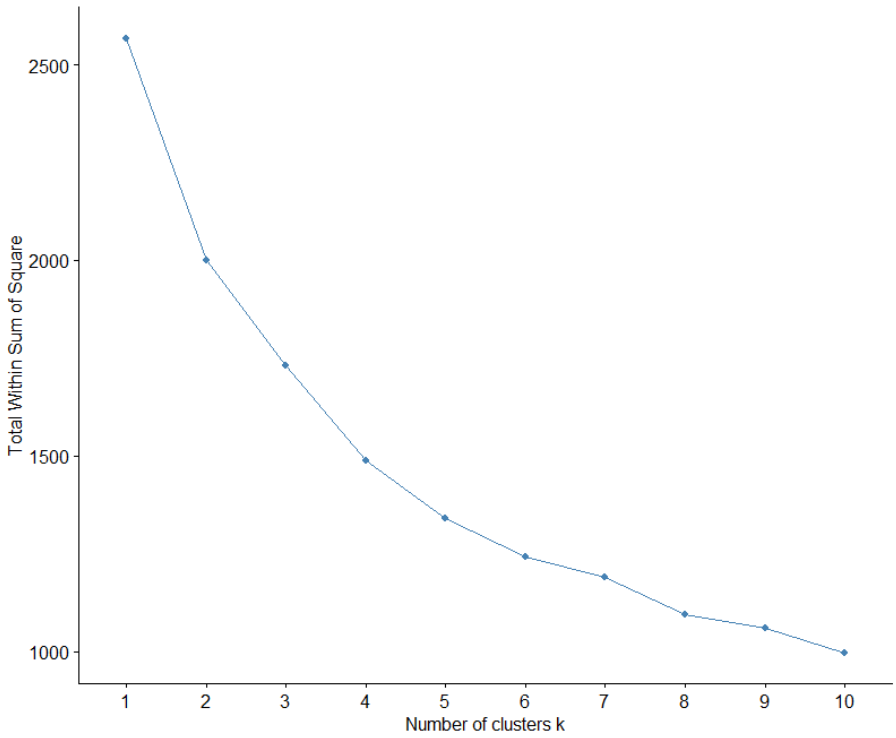


Figure 6: Change in WSS across k clusters

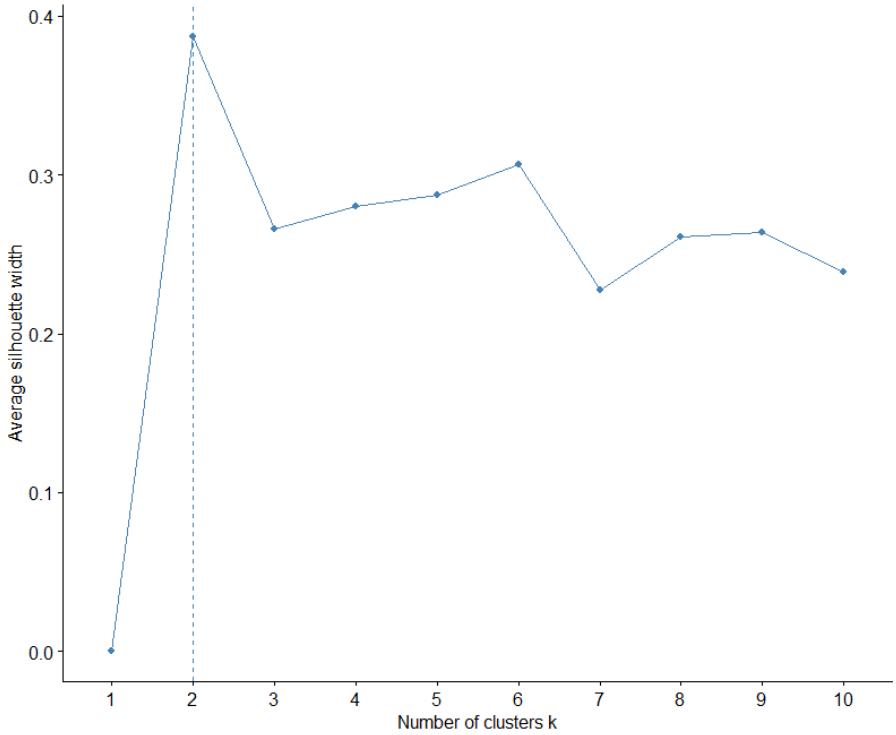


Figure 7: Average Silhouette Width across k clusters

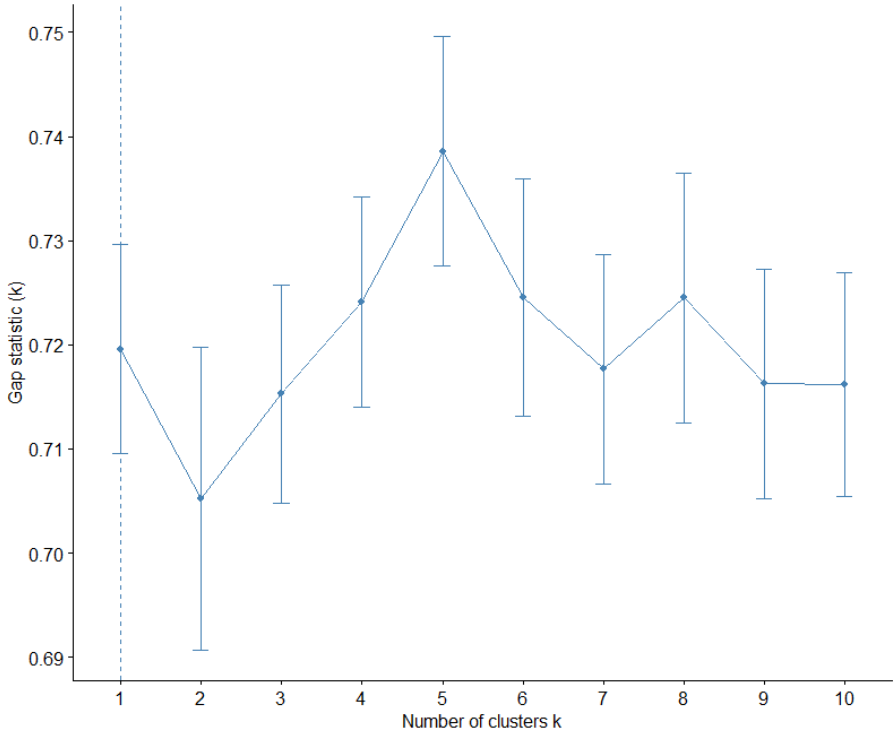


Figure 8: Gap statistics across k clusters

Table 8 contains the cluster means for the k-means solution with two clusters, and Figure 9 depicts the cross-tabulation between the theory-based resilience groups and the two accompanying clusters. From the cross-tabulation, it can be seen that the majority of individuals categorized as either **BM** or **CR** belong to cluster 2. Cluster 1 has a low count of individuals who are **BM** or **No resilience**, whilst having a moderate to high count of individuals who are **CR**. Both of the clusters have a relatively even split of individuals classified as **No resilience**.

It is of interest to examine the characteristics of each cluster to assess whether the potential overlap between data-driven cluster assignments and theory-based group categorization is sensible. The cluster means show stable *MCT* and *MMSE* for both clusters, while the values of *BAG* differ substantially between clusters. Cluster 1 has *BAG* values above 1 for both values of *BAG*, indicating that individuals belonging to this cluster are characterized by having older-appearing brains on average. Whilst cluster 2 has *BAG* values within $|1|$ for both values of BAG, indicating that individuals belonging to this cluster are characterized by having normal-aging brains on average. Cluster 1 has lower values of MMSE (28.8 for $v1$ & 28.1 for $v2$) relative to cluster 2 (29.4 for $v1$ & 29.1 for $v2$) but not enough of a difference to be of clinical significance.

Table 8: Cluster-wise means from the k-means solution

Cluster	**n**	**BAG**$_{v1}$	**BAG**$_{v2}$	**MMSE**$_{v1}$	**MMSE**$_{v2}$	**MCT**$_{v1}$	**MCT**$_{v2}$
1	129	1.40	1.30	28.8	28.1	2.32	2.30
2	300	0.14	-0.71	29.4	29.1	2.43	2.41

Note. BAG = Brain Age Gap; MCT = Mean Cortical Thickness; MMSE = Mini-Mental State Examination

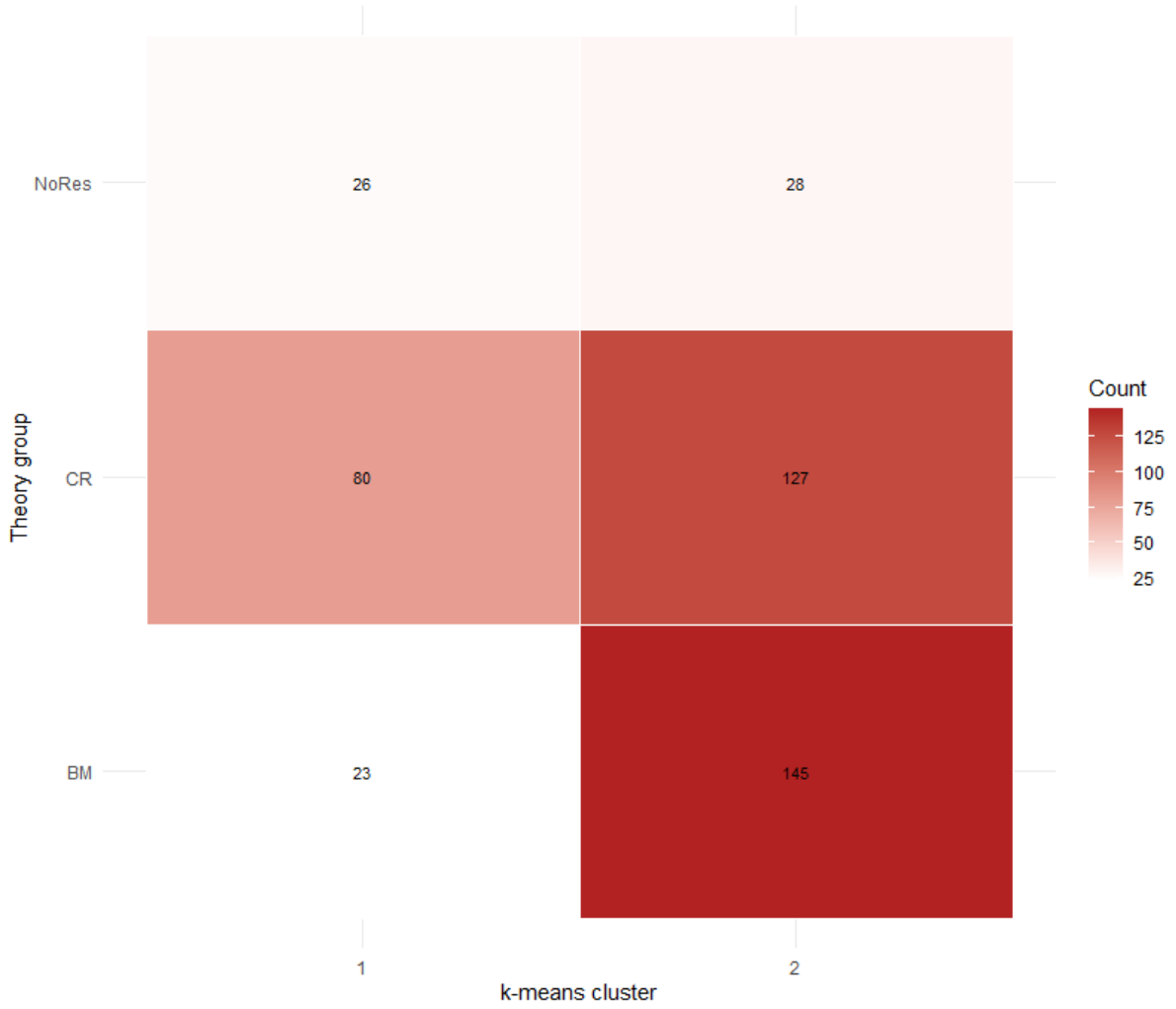


Figure 9: Cross-tabulation between theory-based resilience groups and k-means clusters

Note. NoRes = No resilience

For the K-means clustering, the two-cluster solution resulted in clusters that had characteristics like what one would expect for the two resilience mechanisms of **BM** and **CR**. Cluster 1 is characterized by older-appearing brains, stable *MCT* and *MMSE*. From Figure 9 it can be seen that 62% of the individuals in this cluster also belong to the **CR** group. This degree of overlap makes sense given that this cluster and the **CR** group are characterized by older-appearing brains and stable cognitive function. Cluster 2 is

characterized by normal aging, where the mean values of *BAG* were within |1| for both time points. This characteristic is reflected by the theoretical operationalization in the fact that this cluster contains the majority individuals from the **CR** and **BM** groups, meaning that this cluster contains both individuals with older and younger-appearing brains. The results show that this cluster solution can distinguish between individuals who have substantially older-appearing brains to those with normal-aging brains, but does not distinguish between individuals who have slightly older-appearing brains and younger-appearing brains. These results indicate that there is some sensible overlap between the theory-based operationalization of resilience and K-means clustering when it comes to the resilience mechanisms of **BM** and **CR**. However, there does not seem to be any distinct group that would sensibly correspond to a **No resilience** group, as the individuals from this group are quite evenly spread out among the clusters. It is important to take into account that these cluster results are tentative due to the suggested weak cluster structure in the data indicated by the diagnostics used to determine the number of clusters.

3.2.2 GMM

Figure 10 depicts the change in BIC for an increasing number of components for specifications of GMMs with different covariance structures. From this figure we can determine which model specification yields the best fit. In this case, the model with the highest BIC, and thus the model that will be used for the GMM cluster analysis, is specified with an EEV covariance structure and five total number of components (clusters). EEV stands for Equal volume, Equal shape, and Variable orientation. This covariance structure implies that the clusters formed from this solution have equal overall dispersion (volume), equal eigenvalue ratios (shape), and cluster-specific orientation (clusters can point in different directions) across clusters.

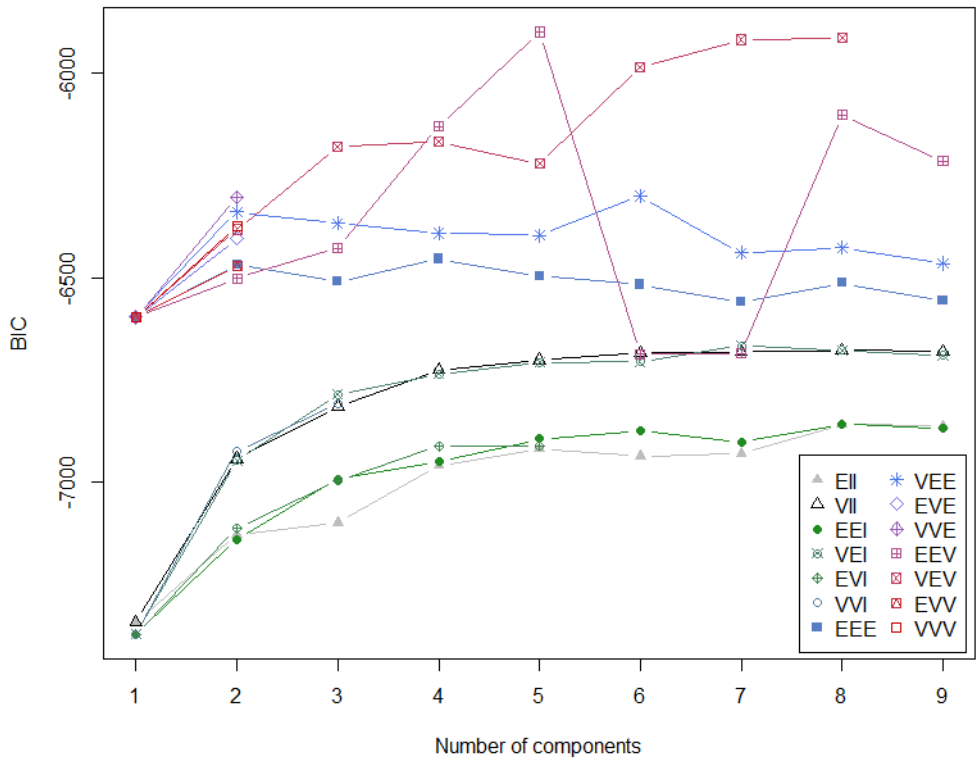


Figure 10: GMM Model Selection - BIC comparison between covariance structure specifications

Table 9 contains the cluster means of the resilience variables for the optimal GMM solution with five clusters and Figure 11 depicts a cross-tabulation between the resilience group assignments and the cluster assignments of the GMM solution. From the cross-tabulation, the cluster assignments for the GMM solution can be seen to be quite spread out across the resilience groups. For the cluster means, clusters 3,4,5 contain a substantial majority of individuals included in the data (128, 46, 214 respectively). These three clusters are also characterized by having, on average, positive *BAG* values (0.47, 0.40, 0.42) at baseline and negative *BAG* values (-0.10, -0.29, -0.25) at follow-up. However, these values are within the range of normal aging for both timepoints. As for the *MMSE* and *MCT*, all clusters have clinically stable differences between baseline and follow-up examination.

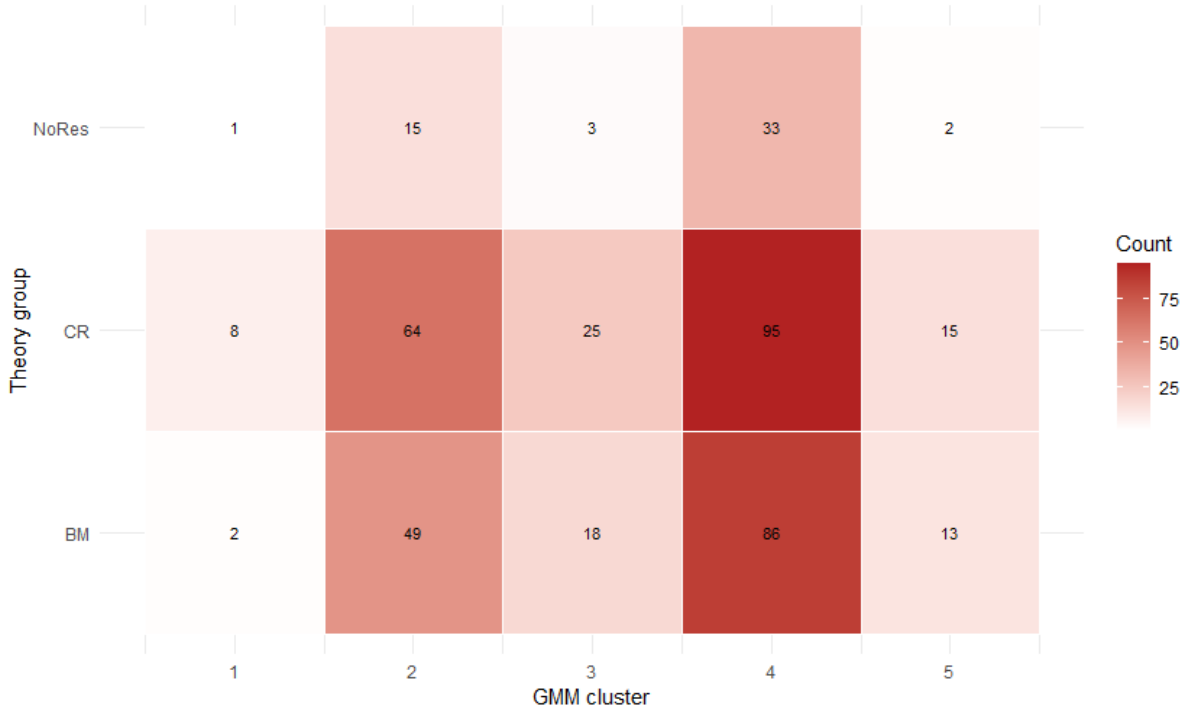


Figure 11: Cross-tabulation between theory-based resilience groups and GMM clusters

Note. NoRes = No resilience

Table 9: Cluster-wise means from the GMM solution

Cluster	**n**	**BAG**$_{v1}$	**BAG**$_{v2}$	**MMSE**$_{v1}$	**MMSE**$_{v2}$	**MCT**$_{v1}$	**MCT**$_{v2}$
1	11	3.24	2.86	27.8	28.4	2.32	2.25
2	128	0.47	-0.10	29.0	28.9	2.40	2.39
3	46	0.40	-0.29	27.6	28.4	2.40	2.39
4	214	0.42	-0.25	30.0	29.1	2.40	2.38
5	30	0.55	0.06	27.2	27.1	2.38	2.37

Note. BAG = Brain Age Gap; MCT = Mean Cortical Thickness; MMSE = Mini-Mental State Examination

Table 10: Bootstrap stability of data-driven clustering solutions

Method	B	**Mean ARI**	**Median ARI**	**SD**
k-means	200	0.729	0.745	0.159
GMM	200	0.780	0.858	0.174

Note B = number of bootstrap resamples.

Table 10 presents the results of the bootstrap stability analysis for the data-driven clustering solutions. Stability was assessed by fitting each clustering method to $B = 200$ bootstrap resamples and comparing the resulting cluster assignments to the reference clustering from the full dataset using the Adjusted Rand Index (ARI). Both k-means and the GMM clustering showed relatively high stability across bootstrap resamples, with mean ARIs of 0.73 and 0.78, respectively. This indicates that both clustering methods produced highly similar solutions across resampled datasets, suggesting that the identified clustering structure is reproducible under resampling-based permutations.

Compared to the K-means solutions, the GMM clustering resulted in a more abstract cluster solution in terms of clinical interpretation. Initially, the five clusters do not seem to have as distinct groupings that would correspond to the theory-based operationalization of the resilience mechanisms of **BM** and **CR**. Cluster 2, 3 and 4 contain the majority of individuals in the data and have very similar mean characteristics in terms of the resilience variables, where the *BAG* values are all positive at baseline and negative at follow-up. The MCT and MMSE for these three clusters are stable between timepoints as well. There are distinctions in the mean values of *MMSE* between these clusters, with cluster 4 having the highest mean *MMSE* values at both timepoints ($v1$ = 30.0, $v2$ = 29.1) and cluster 2 having the lowest ($v1$ = 27.6, $v2$ = 28.4). However, these differences in *MMSE* values are not clinically significant. Upon further examination, these clusters seem to reflect the overall trend in the data, where individuals, on average, seem to have slightly higher *BAG* at baseline relative to follow-up, thereby capturing the observed effect of overall normal aging for the majority of individuals present in the data (as seen in figure 3). If one were to merge these three clusters they could sensibly correspond to the second cluster identified with the K-means algorithm since both are characterized by normal-aging brains. Additionally, cluster 1 has relatively high cluster means for the BAG values ($v1$ = 3.24, $v2$ = 2.86) with stable MMSE ($v1$ = 27.8, $v2$ = 28.4) but only contains 11 individuals. This cluster could correspond to a select group of individuals who exhibit a high degree of coping ability linked to **CR**.

Besides the applied clinical interpretation of the results obtained, it is essential to examine the results received from the data-driven approach from a statistical perspective. As previously shown, the two clustering methods, K-means and GMM, received substantially different results in terms of the number of clusters but had similarities in cluster characteristics. The clustering from the K-means method resulted

in clusters that better aligned with the theoretical framework of resilience whilst the GMM method resulted in more abstract clustering. However, if some of the clusters from the GMM method would be merged, based on characteristics of *BAG* and stable trajectories of *MMSE* and *MCT*, they would reflect similar groupings to that found by the K-means method. The reasons for the discrepancy in cluster numbers are most likely related to the distinct properties and assumptions of these two methods. K-means clustering assumes spherical clusters with equal variances. Consequently, it often identifies compact and convex shaped clusters. Due to its rigid assumptions K-means can fail to uncover more nuanced cluster structures, potentially overlooking subtle variations in the data. In contrast, GMM clustering is a probabilistic approach capable of modeling clusters with varying sizes, shapes, and orientations. This flexibility allows GMM to capture more complex data structures, such as ellipsoidal clusters. This property likely explains the higher number of clusters identified by the GMM. However, within the scope of this study, the increased flexibility of the GMM method did not result in a superior clustering solution when compared to the K-means method. Instead, it seemingly captured negligible statistical differences, particularly in terms of *MMSE*, that lack meaningful clinical implications.

4 Conclusion

The aim of this study was to compare supervised and unsupervised statistical approaches for analyzing resilience, as well as to operationalize the resilience mechanisms of **BM** and **CR** using the *BAG* to evaluate the adequacy of *BAG* as a potential measure of resilience. One strength of this study is that all participants were of the same chronological age at both baseline and follow-up. This makes it possible to interpret the *BAG* without having to adjust for age-related bias, which can otherwise confound brain age estimates.

However, several important limitations must be acknowledged. First, the data is likely biased towards healthier individuals, limiting the contrast between resilient and non-resilient patterns. This could partly reflect selective dropout between examinations, but no information on dropout cause (e.g., mortality) was available, which restricts interpretability. Second, the data contained only two time points, which constrains the ability to reliably evaluate longitudinal trajectories and to capture more nuanced and non-linear changes in *BAG*, *MCT*, and *MMSE*. Third, some statistical assumptions were made for pragmatic reasons given the limited data. For example, constraining the correlation between random effects to zero in the LME models to ensure model identifiability. Furthermore, the decision to treat *MMSE* as semi-continuous was based on common practice in the existing literature, however, the validity of this approach should be formally evaluated. Finally, the clustering analyses provided only weak evidence for natural grouping structure in the data, as indicated by the scree-plot and the gap statistic favouring a one-cluster solution. This suggests that the data are relatively homogeneous and not strongly

partitioned into distinct subgroups.

Within these constraints, the supervised theory-based approach showed some ability to differentiate two theoretically defined resilience mechanisms-brain maintenance and cognitive reserve-but did not clearly identify a distinct non-resilient group. This tentatively supports the potential utility of *BAG* as a resilience-related marker, while highlighting the need to refine the theory-based grouping approach, particularly by incorporating repeated measures of *BAG*. The data-driven approaches produced only partial alignment with the theory-based groups, and given the overall weak clustering structure, these results should be interpreted cautiously. The clearer separation seen in the k-means solution may partly reflect its stronger assumptions of spherical clusters, whereas the GMM's flexible covariance structure may have captured more subtle and diffuse patterns that do not map cleanly onto theoretical constructs.

Overall, these findings should be regarded as preliminary. Future studies using larger and more heterogeneous samples, with more than two time points, are needed to confirm whether the proposed group distinctions reflect meaningful resilience mechanisms and to investigate potential life-course, sociodemographic, and genetic factors that may influence individual trajectories of resilience.

A Appendix

A.1 Sensitivity analysis - MMSE thresholds

Table 11: Resilience group sizes under alternative MMSE thresholds

MMSE cut	BM	CR	NoRes
-1.0	168	152	109
-1.5	168	207	54
-3.0	168	232	29

Note. Group definitions use BAG at baseline (**BM**: $BAG<0$; **CR**: $BAG>0$ & ΔMMSE $>$ cut; otherwise **No resilience**).

Table 12: Fixed effects for BAG (RI model) across MMSE thresholds

	Estimate	95% CI	p
MMSE cut = -1.0			
Time	-0.638	$[-0.872, -0.404]$	$p = 1.31 \times 10^{-7}$
CR $\times$ *Time*	-0.056	$[-0.395, 0.283]$	$p = .746$
No resilience $\times$ *Time*	0.137	$[-0.235, 0.510]$	$p = .470$
MMSE cut = -1.5			
Time	-0.638	$[-0.872, -0.404]$	$p = 1.35 \times 10^{-7}$
CR $\times$ *Time*	0.038	$[-0.277, 0.352]$	$p = .814$
No resilience $\times$ *Time*	-0.025	$[-0.499, 0.449]$	$p = .918$
MMSE cut = -3.0			
Time	-0.638	$[-0.872, -0.404]$	$p = 1.33 \times 10^{-7}$
CR $\times$ *Time*	0.0003	$[-0.307, 0.307]$	$p = .999$
No resilience $\times$ *Time*	0.224	$[-0.385, 0.834]$	$p = .470$

Note. RI = random intercept. **BM** is the reference group. CIs are 95% Wald intervals.

Across MMSE thresholds of -1.0, -1.5, and -3.0, the distribution of participants across resilience groups changed as expected (greater **CR** and fewer **No resilience** with more permissive cuts), but the fixed-effect results for *BAG* were highly consistent (Tables 11–13). In all specifications, the BM group (reference) showed a significant decline over time (Time: $\hat{\beta} = -0.638$, 95% CI $[-0.872, -0.404]$, $p \approx 10^{-7}$), and neither the **CR**$\times$Time nor the **No resilience**$\times$Time interactions reached significance under

Table 13: Fixed effects for BAG (RI+RS, uncorrelated) and model comparison

	Estimate	95% *CI*	p
MMSE cut = -1.0			
Time	-0.638	$[-0.872, -0.404]$	$p = 1.31 \times 10^{-7}$
CR $\times$ *Time*	-0.056	$[-0.395, 0.283]$	$p = .746$
No resilience $\times$ *Time*	0.137	$[-0.235, 0.510]$	$p = .470$
MMSE cut = -1.5			
Time	-0.638	$[-0.872, -0.404]$	$p = 1.35 \times 10^{-7}$
CR $\times$ *Time*	0.038	$[-0.277, 0.352]$	$p = .814$
No resilience $\times$ *Time*	-0.025	$[-0.499, 0.449]$	$p = .918$
MMSE cut = -3.0			
Time	-0.638	$[-0.872, -0.404]$	$p = 1.33 \times 10^{-7}$
CR $\times$ *Time*	0.0003	$[-0.307, 0.307]$	$p = .999$
No resilience $\times$ *Time*	0.224	$[-0.385, 0.834]$	$p = .470$

Note. RI+RS = random intercepts + random. **BM** is the reference group. CIs are 95% Wald intervals.

any threshold, indicating no detectable differences in slope between groups. Overall, the longitudinal *BAG* results are robust to the choice of MMSE threshold used to define resilience groups.

References

Badji, A., Youwakim, J., Cooper, A., Westman, E., & Marseglia, A. (2023). Vascular cognitive impairment–past, present, and future challenges. *Ageing Research Reviews*, *90*, 102042.

Bocancea, D. I., van Loenhoud, A. C., Groot, C., Barkhof, F., van der Flier, W. M., & Ossenkoppele, R. (2021). Measuring resilience and resistance in aging and alzheimer disease using residual methods: A systematic review and meta-analysis. *Neurology*, *97*(10), 474–488.

Borland, E., Edgar, C., Stomrud, E., Cullen, N., Hansson, O., & Palmqvist, S. (2022). Clinically relevant changes for cognitive outcomes in preclinical and prodromal cognitive stages: Implications for clinical alzheimer trials. *Neurology*, *99*(11), e1142–e1153.

Cole, J. H., & Franke, K. (2017). Predicting age using neuroimaging: Innovative brain ageing biomarkers. *Trends in neurosciences*, *40*(12), 681–690.

Cole, J. H., Poudel, R. P., Tsagkrasoulis, D., Caan, M. W., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, *163*, 115–124.

Dartora, C., Marseglia, A., Mårtensson, G., Rukh, G., Dang, J., Muehlboeck, J.-S., Wahlund, L.-O., Moreno, R., Barroso, J., Ferreira, D., et al. (2024). A deep learning model for brain age prediction using minimally preprocessed t1w images as input. *Frontiers in Aging Neuroscience*, *15*, 1303036.

Demidenko, E. (2013). *Mixed models: Theory and applications with r*. John Wiley & Sons.

Dove, A., Shang, Y., Xu, W., Grande, G., Laukka, E. J., Fratiglioni, L., & Marseglia, A. (2021). The impact of diabetes on cognitive impairment and its progression to dementia. *Alzheimer's & Dementia*, *17*(11), 1769–1778.

Fraley, C., & Raftery, A. (2007). Model-based methods of classification: Using the mclust software in chemometrics. *Journal of Statistical Software*, *18*, 1–13.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, *97*(458), 611–631.

Franke, K., Ziegler, G., Klöppel, S., Gaser, C., Initiative, A. D. N., et al. (2010). Estimating the age of healthy subjects from t1-weighted mri scans using kernel methods: Exploring the influence of various parameters. *Neuroimage*, *50*(3), 883–892.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, *2*(1), 193–218.

Jack Jr, C. R., Andrews, J. S., Beach, T. G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., et al. (2024). Revised criteria for diagnosis and staging of alzheimer's disease: Alzheimer's association workgroup. *Alzheimer's & Dementia*, *20*(8), 5143–5169.

Koch, H. J., Gürtler, K., & Szecsey, A. (2005). Correlation of mini-mental-state-examination (mmse), syndrom-kurztest (skt) and clock test (ct) scores in patients with cognitive impairment assessed by means of multiple regression and response surface analysis. *Archives of gerontology and geriatrics*, *40*(1), 7–14.

Maltais, J.-R., Gagnon, G., Garant, M.-P., & Trudel, J.-F. (2015). Correlation between age and mmse in schizophrenia. *International Psychogeriatrics*, *27*(11), 1769–1775.

Marseglia, A., Dartora, C., Samuelsson, J., Poulakis, K., Mohanty, R., Shams, S., Lindberg, O., Rydén, L., Sterner, T. R., Skoog, J., et al. (2024). Biological brain age and resilience in cognitively unimpaired 70-year-old individuals. *Alzheimer's & Dementia*.

Powell, M. J., et al. (2009). The bobyqa algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, *26*, 26–46.

Rydberg Sterner, T., Ahlner, F., Blennow, K., Dahlin-Ivanoff, S., Falk, H., Havstam Johansson, L., Hoff, M., Holm, M., Hörder, H., Jacobsson, T., et al. (2019). The gothenburg h70 birth cohort study 2014–16: Design, methods and study population. *European journal of epidemiology*, *34*, 191–209.

Stern, Y., Albert, M., Barnes, C. A., Cabeza, R., Pascual-Leone, A., & Rapp, P. R. (2023). A framework for concepts of reserve and resilience in aging. *Neurobiology of aging*, *124*, 100–103.

Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., Ewers, M., Franzmeier, N., Kempermann, G., Kremen, W. S., et al. (2020). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's & Dementia*, *16*(9), 1305–1311.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the royal statistical society: series b (statistical methodology)*, *63*(2), 411–423.

Whitson, H. E., Duan-Porter, W., Schmader, K. E., Morey, M. C., Cohen, H. J., & Colón-Emeric, C. S. (2016). Physical resilience in older adults: Systematic review and development of an emerging construct. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *71*(4), 489–495.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., et al. (2004). Mild cognitive impairment–beyond controver-

sies, towards a consensus: Report of the international working group on mild cognitive impairment. *Journal of internal medicine*, *256*(3), 240–246.

Woodward, M., & Galea, M. (2005). Mini-mental state examination (mmse). *Aust J Physiother*, *51*(3), 198.