

Title

Discovering Subtypes with Imaging Signatures in the Motoric Cognitive Risk Syndrome Consortium using Weakly-Supervised Clustering.

Bhargav Teja Nallapu¹, Ali Ezzati², Helena M. Blumen^{3,4}, Kellen K. Petersen⁵, Richard B. Lipton¹, Emmeline Ayers¹, V G Pradeep Kumar⁶, Srikanth Velandai^{7,8}, Richard Beare^{7,9}, Olivier Beauchet^{10,11}, Takehiko Doi¹², Hiroyuki Shimada¹², Sofiya Milman⁴, Sandra Aleksic⁴, Joe Verghese^{1,3}

¹Department of Neurology, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, New York, 10461, USA

²Department of Neurology, University of California, Irvine (UCI), Irvine, 200 S. Manchester Ave. Ste. 206, Orange, California, 92868, USA

³Department of Neurology, Stony Brook University, 101 Nicolls Road, Stony Brook, New York 11794, USA

⁴Departments of Medicine, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, New York, 10461, USA

⁵Department of Neurology, Washington University in St. Louis, 660 S. Euclid Ave, St. Louis, MO 63110, USA

⁶Department of Neurology, Baby Memorial Hospital, Indira Gandhi Road, Kozhikode, Kerala, 673004, India

⁷National Centre for Healthy Ageing, 2 Hastings Rd, Frankston VIC 3199, Australia

⁸School of Translational Medicine, Monash University, 99 Commercial Rd, Melbourne, VIC, 3004, Australia

⁹Developmental Imaging Group, Murdoch Children's Research Institute, 50 Flemington Rd, Parkville VIC 3052, Australia

¹⁰Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), 4565, chemin Queen Mary, Montréal QC H3W 1W5, Canada

¹¹ Department of Medicine, University of Montreal, C. P. 6128, succursale Centre-ville, Montréal, QC, H3C 3J7, Canada

¹²Department of Preventive Gerontology, Center for Gerontology and Social Science, Research Institute, National Centre for Geriatrics and Gerontology, 7-430, Morioka-cho, Obu City, 474-8511, Japan

Corresponding author:

Bhargav T. Nallapu

Email: bhargav.nallapu@einsteinmed.edu

Address:

Department of Neurology

Albert Einstein College of Medicine

Van Etten 3C12

1300 Morris Park Avenue

Bronx, NY 10461

Phone: 718-430-3896

STRUCTURED ABSTRACT

INTRODUCTION: Understanding the heterogeneity of brain structure in individuals with the Motoric Cognitive Risk Syndrome (MCR) may improve the current risk assessments of dementia.

METHODS: We used data from 6 cohorts from the *MCR consortium* (N=1987). A weakly-supervised clustering algorithm called HYDRA was applied to volumetric MRI measures to identify distinct subgroups in the population with gait speeds lower than one standard deviation (1SD) above mean.

RESULTS: Three subgroups (Groups A, B & C) were identified through MRI-based clustering with significant differences in regional brain volumes, gait speeds, and performance on Trail Making (Part-B) and Free and Cued Selective Reminding Tests.

DISCUSSION: Based on structural MRI, our results reflect heterogeneity in the population with moderate and slow gait, including those with MCR. Such a data-driven approach could help pave new pathways toward dementia at-risk stratification and have implications for precision health for patients.

1. BACKGROUND

Dementia is marked by a progressive decline in cognitive and functional. Accessible and informative clinical indicators that help identify individuals at risk of incident dementia could enable targeted, earlier and more effective interventions. Like Mild Cognitive Impairment (MCI), Motoric Cognitive Risk (MCR) syndrome is a pre-dementia state characterized by subjective cognitive complaints coupled with objectively measured impairment.¹ While MCI which is characterized by reduced performance on objective cognitive tests, MCR is characterized by slow gait speed and subjective cognitive impairment². Prior studies indicate that MCR is a significant risk factor for dementia in older adults³. Compared to MCI, MCR does not require neuropsychological assessments, which are generally time-consuming. This makes MCR a more efficient and accessible solution for risk stratification of dementia in community-dwelling adults and across various settings, from primary care to specialized neurology clinics⁴. Both components of MCR, slow gait and subjective cognitive impairment, are also independently predictive of future cognitive decline and incident dementia – but the MCR phenotype has higher predictive validity for dementia than either of the individual components alone^{5,6}.

In older adults, slow gait speed can stem from various factors such as neurological issues, muscle-related conditions, arthritis, or a combination of these factors. Neurodegeneration caused by various dementia etiologies, such as Alzheimer's disease (AD), is considered one of the primary reasons behind declines in both gait speed and cognitive function⁷. The brain regions that are known to mediate executive functions (EFs), such as the frontal and prefrontal-lobe networks, also control gait ability^{8,9}. These brain regions are involved in integrating information from many cortical sensory systems, modulate and produce goal-directed actions and behavior¹⁰. Atrophy in these regions causes both cognitive and gait decline, concurrently with the aging process¹¹⁻¹⁴. Preliminary intervention trials using cognitive training or brain stimulation to enhance EF have also shown improvements in gait velocity^{15,16}. MCR, despite these shared neuroanatomical pathways, describes people who are still cognitively intact but with cognitive complaints

and slowing of gait. The precise neuroanatomical signatures that correspond to the clinical manifestations in MCR, however, are unknown. To study the biological underpinnings of MCR, an MCR-neuroimaging consortium was established, incorporating data from seven cohorts across five different countries, and three continents. that collected structural MRIs, gait speeds, cognitive assessments, and other clinical symptoms. This consortium provides a valuable opportunity to study neurodegeneration patterns as identified through structural MRIs and to connect these findings with MCR, its components, and other clinical symptoms.

There is substantial neuroanatomical heterogeneity in preclinical stages of dementia¹⁷. Several studies have linked MCR to specific patterns of brain atrophy, specifically in prefrontal, supplementary motor, insular and motor cortices^{13,14}. However, very few data-driven studies have explored the variability of neuroanatomical patterns in MCR and its components¹⁴. Therefore, in this study, we adopted a machine learning based data-driven approach to investigate structural brain differences among participants in the MCR consortium using volumetric imaging. We employed a novel semi-supervised clustering approach called HYDRA (Heterogeneity through Discriminative Analysis)¹⁸, to identify subgroups with distinct structural brain patterns within the MCR Consortium. Several other approaches were previously proposed to reveal the inherent disease heterogeneity. But most of these methods either relied on predefined clinical subgroups, ignoring multivariate relationships in the data^{19,20}, or applied clustering directly to brain anatomies, risking the identification of normal inter-individual variability, some of which could be due to confounding factors such as sex, age etc. rather than disease-specific heterogeneity^{21,22}. The HYDRA method enables us to mitigate these challenges and disentangles the heterogeneity in a population, using another reference population. This was achieved by leveraging data from a reference group composed of individuals with faster gait speeds (population FG, those with gait speed faster than 1 standard deviation above the population mean within each cohort). Identifying subgroups with homogenous neuroanatomical patterns in a cohort population with gait variability has a potential to further our understanding of biological underpinnings of MCR, and in developing interventions for dementia.

2. Methods

2.1. Participants

Data from 2,007 older adults in the MCR consortium was examined. The data were obtained from 7 different cohorts and 5 different countries. The cohorts were : (i) the Central Control of Mobility in Aging Study (CCMA) ^{7,23} in the US, (ii) the LonGenity study ²⁴ in the US, (iii) the Einstein Aging Study (EAS) ^{2,25} in the US, (iv) the Tasmanian Study of Cognition and Gait (TASCOG) ^{26,27} in Australia, (v) the National Center for Geriatrics and Gerontology–Study of Geriatric Syndromes (NCGG-SGS) ^{28,29} in Japan, (vi) the Kerala-Einstein Study ³⁰ in India, and (vii) the Gait and Alzheimer’s Interactions Tracking study (GAIT) ^{9,31} in France. All the study procedures in each of the cohorts were approved by the local institutional review boards. All the cohorts excluded individuals with prevalent dementia.

Data from all the cohorts were used in the analyses. We reserved a subset of population from each cohort as the reference population for the clustering model, that had faster gait speeds (FG).

This analysis was approved by the institutional review board of the Albert Einstein College of Medicine, NY. All participating cohorts have received approval from their local ethics committees.

2.2. Study Measures

Data from all the cohorts included population demographics. To harmonize this data and address cohort variability, we followed a standardized process for the preparation of each modality of data.

Gait Speeds: Gait speed (cm/s) was available in all the studies. More detailed description of gait speed measurements and quantification were presented in previous studies^{5,32,33}. Gait speed was normalized within each study, then the individuals with the normalized gait speed greater than 1 standard deviation above the mean of the study population were labelled as FG. The study population was the remaining individuals with gait speeds less than 1 SD above population mean (normal and slower gait speeds). The FG participants from all the studies were used as the reference population for the HYDRA clustering algorithm.

Volumetric MRI: MRI measures of all the individuals across all the cohorts were available, which were collected at the respective study sites and harmonized at a single site using the FreeSurfer pipeline, using the standard parcellation and correction methods, as described elsewhere¹³. For this study, we used the volumetric measures of 41 brain regions (Supplementary Table 1), both the hemispheres added where applicable, and total Intra Cranial Volume (tICV). To account for the individual differences in tICV, within each cohort, the volume of each region (VR) was normalized according to the mean tICV of the population within that cohort. The adjusted volume (VRa) of a brain region of an individual was calculated as

$$VRa = (VR/tICV) * \text{mean}(tICV)$$

Neuropsychological evaluations: For the post-hoc analyses of objective cognitive scores of different subgroups obtained in our analyses, we used available neuropsychological scores from different studies, normalized per study, wherever applicable. The following were the neuropsychological tests that were used :

- Trail Making Test, Part B (TMT-B)³⁴; available in CCMA, LonGenity, EAS, NCGG-SGS (evaluated using tablet version³⁵), KES and GAIT
- Stroop Color Word test³⁶, consisting of subtests measuring time required to name the colors seen (COLOR), read the given words (WORD) and name the color of the printed word (COLOR_WORD); we used Interference (Stroop-INT) which is the difference between the third subtask and the first task (COLOR_WORD - COLOR), available in EAS, TASCOG and GAIT
- Free and Cued Selective Reminding Test (FCSRT)³⁷, a recall test that uses either words or images. Scores includes the sum of free recall (FR) alone (range 0-48) and combined with cued recall as total recall (TR), and the sum of FR and TR (FR96: range 0-96); available in CCMA, EAS and LonGenity

Diagnosis of MCR: MCR is characterized by slow gait accompanied by cognitive complaints. In our study, slow gait was defined as gait speed lesser than one standard deviation below the age and sex-specific means in each cohort³⁸. Gait speed in older adults from different cohorts and countries can vary due to differences in factors such as height, sex, race and ethnicity, education, and socioeconomic status.^{39,40} Hence, using a universal single gait speed cut score, although used in some prior studies to define normal or abnormal performance, is problematic as it introduces greater heterogeneity between different populations. Different cognitive complaint questionnaires were used across the 7 cohorts, sometimes with difference in their versions. For our study, subjective cognitive complaint was determined using the memory item from Geriatric Depression Scale (GDS) uniformly in all cohorts⁴¹ or the Instrumental Activities of Daily Living (I-ADL)⁴², where available in cohorts CCMA and KES.

2.3. Analytic Approach

Figure 1 depicts the overall analytic pipeline of our study. The primary goal of this study was to identify subgroups with neuroanatomical heterogeneity among individuals with slower gait speeds using a machine learning clustering algorithm. There were as many as 41 brain region volumes available through preprocessing of structural MRIs via the automated FreeSurfer pipeline. Some of the brain region volumes are highly correlated with each other and using all of them together in a clustering model may lead to skewness and bias in the clustering results and making them less representative of the true underlying patterns in the data. Moreover, in high-dimensional spaces with many correlated features, there's an increased risk of finding weak correlations that may result in misleading subgroup formations. Therefore, we used Factor Analyses (FA) to reduce the number of input variables to the clustering model that results in heterogeneous subgroups of population.

Factor Analyses: For the Factor Analyses, we used maximum likelihood estimation method and varimax rotation with the 41 regions available as input variables to obtain a substantially smaller number of latent factors. The relationship between each variable (region) and the latent factors is described by a matrix of weights, or factor loadings, generated from an FA model. We chose one variable within each factor that had the highest loading to make the feature set ($\text{Regions}_{\text{Factor}}$) for the subsequent clustering model. Therefore, before performing factor analysis, we thoroughly evaluated the “factorability” of the combined all-cohort dataset. Bartlett’s test of sphericity⁴³, which checks whether or not the observed variables intercorrelate, returned a significant value of $\chi^2 = 48722.9$, $p < 0.0001$. If the test found statistically insignificant, the dataset would not have been suitable for factor analysis. We also performed the Kaiser-Meyer-Olkin (KMO) test which assesses the suitability of data for FA by measuring the sampling adequacy for each variable using the proportion of variance⁴⁴. The overall KMO value of our dataset was 0.94 indicating a more than adequate sampling in the data (KMO value below 0.6 is considered insufficient to proceed with FA). Thus, we proceeded with the Factor Analyses of the dataset

with the available 41 volumetric MRI variables. We must specify the number of factors desired to the Factor Analyzer method being used. We performed an exploratory FA with several possible number of factors up to 25 and chose the number of factors for the model using the Kaiser criterion⁴⁴, based on the eigenvalues of the factors greater than 1, which represent the amount of variance explained by each factor.

Weakly-Supervised Clustering: Traditional clustering methods in machine learning are referred as ‘unsupervised’ clustering where only specific measures from the desired population are used as input for the model that results in the subgroups (clusters) of that population. The challenge with unsupervised clustering approaches is that they are not capable of accounting for ‘healthy state’ or ‘disease state’, and clustering is only performed on the specified measure (e.g., imaging biomarkers). To mitigate this problem and to obtain the subgroups in our study population relative to the FG group (i.e. *healthy state* in our population), we used a novel clustering variant, a weakly-supervised algorithm called HYDRA (Heterogeneity through Discriminative Analysis), that also considers the characteristics of population in FG group with faster gait speeds as reference, such that the population to be clustered can be simultaneously separated from the controls (FG) while quantifying the heterogeneity within the population through their association to the sub-classifiers that separate them from the reference group¹⁸. As input features, we used only those regions that had the highest loading within each factor in the result of FA in the previous step. Age and sex were added as covariates to the clustering algorithm.

Statistical Analyses: The descriptive characteristics of the entire population, each cohort, and subgroups resulted from clustering were examined. The demographic variables such as age, sex, education, and ethnicity were included in the descriptive characteristics. An Analysis of Variance (ANOVA) was conducted for the quantitative demographics such as age and education. For categorical variables such as ethnicity and sex, Fisher’s exact test was performed. For the volumetric MRI measures, gait speeds and

cognitive scores, an ANCOVA analysis was performed, which included age and sex as covariates to identify differences across the subgroups. For the MRI measures, we first compared a set of regions known to be associated with MCR based on previous studies¹³. We also compared the volumes of regions used as input for the clustering model. Similar analyses were performed for gait speeds and cognitive scores across the subgroups. In case of significant differences found in the ANCOVA analyses, pairwise comparisons were performed accounted by a post-hoc Tukey's honest significant difference (HSD) test ($p < 0.05$).

All the data preprocessing, Factor Analyses and HYDRA clustering were performed using Python with scikit-learn library for FA and mlni package for HYDRA clustering¹⁸. The statistical analyses were conducted using the R statistical software, version 4.2.2 for MacOS.

3. Results

Population Characteristics

A total of 1,987 participants were part of the final clustering model after excluding those with missing data or outliers in their volumetric MRI measures (mean age, 71.73 ± 6.87 years; 47.7% female). There were 294 participants in the reference FG group (mean age, 69.30 ± 5.52 years; 47.3% female) and 1693 participants in the remaining study population (mean age, 72.15 ± 6.99 years; 47.8% female). The FG group had younger population compared to the study population ($p < 0.001$, effect size of -0.52). Both the groups differed in the mean years of education of the population ($p = 0.03$, effect size of 0.14). There were no sex-differences found between the groups. There was no significant difference found in both the groups in terms of cohort membership. The reference group FG differed from the study population in the proportion of participants with MCR diagnosis, as defined by 'slow gait', depending on threshold specific to the cohort, and subjective cognitive complaint as determined by either GDS or ADL (3.3% in the FG group, 13.3% in the study population, $p < 0.001$) (see [Table 1](#)).

Factor Analyses

We found the number of factors to be 8 and conducted FA with the same. **Supplementary Table 1** summarizes the factor loadings of each variable in all the 8 factors. For simplicity, we presented only those variables which had loadings of at least 0.4 in any factor. We chose one variable (region) from each factor that had the highest loading as the input for clustering model. The eight regions that were chosen to be the input features for the clustering model, $Regions_{Factor}$, were the following - Accumbensarea, Caudate, Banks of the Superior Temporal Sulcus (banksSTS), Caudal Anterior Cingulate, Entorhinal Cortex, Pars Triangularis, Pericalcarine, and Precentral Gyrus.

HYDRA Clustering Analyses

We obtained 3 subgroups within the study population from HYDRA, with an adjusted rand index (ARI) of 0.1. We performed statistical analyses to assess the differences across the 3 subgroups, Group A, Group B and Group C, between each of them as well as with respect to the FG group. **Table 1** includes descriptive characteristics of the population in each subgroup obtained. Of the 1693 individuals, 585 were grouped under Group A (mean age, 72.07 ± 6.80 years; 47.4% female), 455 were grouped under Group B (mean age \pm SD, 70.94 ± 6.77 years; 50.3% female), and 653 under Group C (mean age, 73.06 ± 7.18 years; 46.4% female). The mean gait speed (\pm SD) in Group A was $104.91 (\pm 17.05)$, $103.79 (\pm 18.28)$ in Group B and $100.04 (\pm 21.34)$ in Group C.

Heterogeneity in Brain Patterns in the MCR Consortium participants

Post-hoc pairwise group comparisons revealed significant differences across the subgroups for most of the regions (**Table 2 and Figure 2, Factors**). Among all the subgroups, Group A had the highest volumes of Accumbensarea (mean z-score of 0.481 ± 0.140), followed by the group FG (0.434 ± 0.126), Group B (0.425 ± 0.106) and Group C (0.329 ± 0.106). Group A had the highest volumes of Caudal Anterior

Cingulate (mean z-score of 0.513 ± 0.117), followed by the Group B (0.478 ± 0.113), FG (0.465 ± 0.126), and Group C (0.392 ± 0.118). Except the Groups FG & B, all the pairs of groups differed significantly in their volumes of Accumbensarea and Caudal Anterior Cingulate ($p < 0.0001$). All the pairs of subgroups differed significantly in their volumes of banksSTS, parstriangularis, pericalcarine and precentral gyrus. While Group A had the highest volume of banksSTS and precentral gyrus, Group B had the highest volume of parstriangularis and pericalcarine. Group B & FG did not differ significantly in their precentral gyrus volumes.

Next, we combined the volumes of brain regions into respective cortical, subcortical and other regions that were generally implied in MCR and dementia - C/S ROIs – Frontal, Temporal, Parietal, Occipital, Hippocampus, Cerebellum and Subcortical (Basal Ganglia nuclei, Amygdala, and Thalamus). We repeated the ANCOVA analyses and post-hoc pairwise comparisons to examine the differences across the subgroups. Differences in the C/S ROIs between the subgroups is reported in **Table 3; Figure 2, Cortical**). Group B did not show any significant differences in the volumes of C/S ROIs compared to that of FG, except in the Occipital lobe. Group A showed significant differences in all the C/S ROIs volumes compared to those of Group B, except in the Occipital lobe. Group A had the highest volumes across all the C/S ROIs while Group C had the least. Group C showed significant differences in the all the C/S ROIs volumes compared to those of FG, except in the Subcortical regions. Group C showed significant differences in the all the C/S ROIs volumes compared to those of Group B, except in the Cerebellum and Subcortical regions. Finally, Group C and Group A differed across all the C/S ROIs volumes.

Heterogeneity in Gait Speeds and Cognition in the MCR Consortium participants

ANCOVA analyses showed that the subgroups differed in their gait speeds. A post-hoc analysis was performed to evaluate subgroups differences. While Group B did not show a significant difference with either Group A or Group C, the mean gait speed of Group A was higher than that of Group C (-

0.192±0.739 vs -0.364±0.864, $p=0.003$). In the ANCOVA analyses with each cognitive score as the outcome, all subgroups differed in their TMT-B and FR96 scores but not in their Stroop-INT scores. For TMT-B, the FG group had better scores (mean±SD: -0.354±0.892) compared to Group B (-0.066±0.906, $p=0.03$), Group A (-0.044±0.905, $p=0.05$), and Group C (0.174±1.153, $p<0.001$). Within the subgroups, only Groups A and C showed a significant difference ($p=0.02$). With respect to FR96 scores, the FG group had higher scores (0.599±0.716) compared to Group C (-0.125±1.136, $p=0.02$). Group A also had higher scores (0.389±0.778) compared to Group C ($p=0.01$). The groups did not show significant differences in their Stroop-Int scores. **Table 4** and **Figure 3** summarize the ANCOVA and post-hoc pairwise comparison results for gait speed, and cognitive scores across the subgroups.

4. Discussion

The MCR Consortium comprises over 2,000 MRIs from older adults without dementia from multiple locations across different continents. It provides an invaluable repository of neuroimaging and other measures to further our understanding of MCR as a crucial risk factor for dementia. Using a novel data-driven machine learning approach, we focused on exploring the heterogeneity in the population with fast and normative gait speeds (slower than 1 SD above the population mean), while accounting for the distinct neuroanatomical patterns that separate each subgroup from the population with fast gait speeds. Among the participants without dementia and gait speeds in the normal to slow range, we found three subgroups which significantly differed in the patterns of brain regional volumes, gait speeds, and cognitive performance.

Group A had slower gait speeds compared to the reference fast gait speed (FG) group and largest volumes of cortical and subcortical regions. Group B had intermediate gait speeds and cortical volumes similar to the reference group. Group C had the slowest gait speeds, the smallest overall regional brain volumes in

most brain regions and the highest prevalence of MCR. The subgroups also differed in their performance in Trail Making Test (Part B) and Free & Cued Selective Reminder Test.

Abnormal gait has been shown to be a reliable predictor of non-Alzheimer dementia^{45,46}. Since slowing of gait speed precedes cognitive decline^{32,33}, understanding the heterogeneity related to gait speed patterns can be very valuable in characterizing predementia syndromes such as MCR. MCR, which is characterized by slow gait and subjective cognitive impairment, depends on varying thresholds to determine slow gait ranging from 44.4-101.9 cm/s among men and 36.9-97.4 cm/s among women as observed across different study locations⁵. In our analyses, instead of studying abnormal gait using fixed thresholds, we considered a broader group of older population without dementia with gait speeds that are fast or normative (less than 1 SD above population mean). We were interested in studying neuroanatomical patterns within this population in relation to their gait speed and in exploring possible differences in their cognition. As hypothesized, we found homogenous subgroups that showed significant differences in their mean gait speeds accompanied by corresponding distinct neuroanatomical patterns as shown by differences in their brain region volumes. The subgroup with the slowest mean gait speed (Group C) had the least volumes across the prefrontal, temporal, and parietal cortex regions, hippocampus and cerebellar cortex than any other subgroups, resembling a pathological pattern of an MCR subtype accompanied by olfactory dysfunction that is more likely to be associated with Alzheimer and Lewy body dementias⁴⁷. Our results were consistent with the previous MCR studies that showed higher association of lower cortical gray matter volumes and total hippocampal volumes with incident MCR^{8,9,13}. Similar results were observed across the regions that were used in the clustering model, except for the Caudate nucleus, for which no significant difference was found across any of the groups. Interestingly, Group A, which had a slower mean gait speed compared to the reference fast gait group, had higher brain region volumes across prefrontal, hippocampal, temporal, parietal regions. We could not find a specific neuroanatomical pattern that could explain why Group A may have slower gait speeds than the reference group FG while having greater brain region volumes. This finding might be explainable by other

comorbidities. However, due to the lack of detailed data on these comorbidities, we were not able to study them in the context of the current study.

Prior studies indicate that MCR syndrome is associated with worse performance in different cognitive domains such as attention and memory, as well as global cognition⁴⁸. In a study of 314 nondemented older adults⁴⁹, it was shown that different MCR subtypes show differences in cognitive profiles, with different gait attributes such as speed, stride length and swing being differently associated with global cognition and memory. In this study, we evaluated possible subgroup differences in their cognitive performance as measured by different tests. We compared the performances of the individuals across all the subgroups in TMT-B, FCSRT (as measured by FR96) and Stroop Task - Interference. All the subgroups showed worse performance in TMT-B and FCSRT compared to that of the faster gait speed group. All the subgroup pairs, except Groups A and B, differed in their TMT-B scores, underscoring the previously established links between executive function and gait speed, accompanied by cognitive complaints^{8,15}. Whereas only the slowest gait subgroup, Group C, differed in its mean FR96 scores compared to FG and Group A. None of the subgroups differed in their cued recall scores and Stroop-Interference scores, which is likely attributable to the fact that the study populations were all dementia-free at baseline visit. Our findings underscore various results from previous studies that highlighted the negative association between gait speed and impairment in executive function and global cognition^{48,50,51}.

We used the data from a large multi-center consortium of studies involving community-dwelling non-demented adults. However, a few limitations should be noted. Due to cross-sectional analyses, it is not possible to establish any causal relationships between identified structural brain changes among groups and clinical outcomes. A future direction for us is to study if these subgroups that differ in their neuroanatomic and gait speed patterns, would show different trajectories in their gait speed slowdown or cognitive decline. Previous studies that both cortical volume and thickness correlate with cognitive function and neurodegeneration⁵², while particularly in the case of MCR it was shown that cortical

thickness detected neurodegeneration easier than cortical volumes¹³. However, it was shown that cortical volume can be more sensitive to detecting changes in brain structure associated with aging and neurodegenerative processes⁵³ and numerous studies have successfully used cortical volumes in machine learning models for various neurological conditions^{54,55}. Therefore, we used cortical volumes as the inputs for our model. The effects of using cortical thickness on the resulting subgroups should be further explored. We used a novel method of combining factor analyses with weakly-supervised clustering, which might not be necessarily the most fitting approach to study the relation between gait speeds and neuroanatomical heterogeneity. However, this approach handles the problem of collinearity in brain region volumes very well. We must highlight that the results of the analyses of subgroup differences in their cognitive performance may not be generalizable since different cognitive tests and different versions have been used across the cohorts and some of the cognitive tests were not available for some cohorts. In combination with longitudinal measures and harmonized cognitive measures, novel machine learning methods such as HYDRA or the most recent generative methods⁵⁶ that are specialized in studying heterogeneity in imaging, have a huge potential to explore disease-subtypes and disease-staging in gait abnormalities and MCR syndrome.

The significance of our work lies in its potential to encourage further research related to previously unexplored heterogeneity within the at-risk population with slower gait speeds. By identifying distinct subgroups based on brain volume patterns as well as other biomarkers, we can potentially improve our understanding of the underlying pathophysiology associated with slow gait and its relation to MCR syndrome and subsequent progression to dementia. This approach might help paving pathways toward patient stratification at early asymptomatic stages and have implications for precision health. As the multi-site study consortiums such as MCR Imaging Consortium are invaluable to the research of aging and dementia, our methodology succeeds in handling the challenges of such complex datasets, offering a robust framework for analyzing neuroimaging data across diverse populations. Ultimately, our work

contributes to the growing body of knowledge on MCR and may pave the way for more nuanced approaches to early dementia detection and prevention.

References

1. Campbell NL, Unverzagt F, LaMantia MA, Khan BA, Boustani MA. Risk Factors for the Progression of Mild Cognitive Impairment to Dementia. *Clin Geriatr Med*. 2013;29(4):873-893. doi:10.1016/j.cger.2013.07.009
2. Verghese J, Wang C, Lipton RB, Holtzer R. Motoric cognitive risk syndrome and the risk of dementia. *J Gerontol A Biol Sci Med Sci*. Apr 2013;68(4):412-8. doi:10.1093/gerona/gls191
3. Meiner Z, Ayers E, Verghese J. Motoric Cognitive Risk Syndrome: A Risk Factor for Cognitive Impairment and Dementia in Different Populations. *Ann Geriatr Med Res*. Mar 2020;24(1):3-14. doi:10.4235/agmr.20.0001
4. Belleville S, Fouquet C, Hudon C, Zomahoun HTV, Croteau J, Disease-Quebec CftEIoAs. Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. *Neuropsychol Rev*. 2017;27:328-353.
5. Verghese J, Ayers E, Barzilai N, et al. Motoric cognitive risk syndrome. *Neurology*. 2014;83(24):2278-2284. doi:doi:10.1212/WNL.0000000000001084
6. Semba RD, Tian Q, Carlson MC, Xue Q-L, Ferrucci L. Motoric cognitive risk syndrome: Integration of two early harbingers of dementia in older adults. *Ageing Research Reviews*. 2020/03/01/ 2020;58:101022. doi:<https://doi.org/10.1016/j.arr.2020.101022>
7. Holtzer R, Mahoney J, Verghese J. Intraindividual variability in executive functions but not speed of processing or conflict resolution predicts performance differences in gait speed in older adults. *J Gerontol A Biol Sci Med Sci*. Aug 2014;69(8):980-6. doi:10.1093/gerona/glt180
8. Ezzati A, Katz MJ, Lipton ML, Lipton RB, Verghese J. The association of brain structure with gait velocity in older adults: a quantitative volumetric analysis of brain MRI. *Neuroradiology*. Aug 2015;57(8):851-61. doi:10.1007/s00234-015-1536-2
9. Beauchet O, Allali G, Annweiler C, Verghese J. Association of Motoric Cognitive Risk Syndrome With Brain Volumes: Results From the GAIT Study. *J Gerontol A Biol Sci Med Sci*. Aug 2016;71(8):1081-8. doi:10.1093/gerona/glw012
10. Holtzer R, Wang C, Lipton R, Verghese J. The protective effects of executive functions and episodic memory on gait speed decline in aging defined in the context of cognitive reserve. *J Am Geriatr Soc*. Nov 2012;60(11):2093-8. doi:10.1111/j.1532-5415.2012.04193.x
11. Martin KL, Blizzard L, Wood AG, et al. Cognitive function, gait, and gait variability in older people: a population-based study. *J Gerontol A Biol Sci Med Sci*. Jun 2013;68(6):726-32. doi:10.1093/gerona/gls224
12. Nagamalla V, Verghese J, Ayers E, et al. Distinct Patterns of Brain Atrophy in amnesic Mild Cognitive Impairment and Motoric Cognitive Risk Syndromes. *Neurodegener Dis*. Aug 5 2024;doi:10.1159/000540512
13. Blumen HM, Schwartz E, Allali G, et al. Cortical Thickness, Volume, and Surface Area in the Motoric Cognitive Risk Syndrome. *J Alzheimers Dis*. 2021;81(2):651-665. doi:10.3233/jad-201576
14. Blumen HM, Allali G, Beauchet O, Lipton RB, Verghese J. A Gray Matter Volume Covariance Network Associated with the Motoric Cognitive Risk Syndrome: A Multicohort MRI Study. *J Gerontol A Biol Sci Med Sci*. May 16 2019;74(6):884-889. doi:10.1093/gerona/gly158
15. Marusic U, Verghese J, Mahoney JR. Cognitive-Based Interventions to Improve Mobility: A Systematic Review and Meta-analysis. *J Am Med Dir Assoc*. Jun 2018;19(6):484-491.e3. doi:10.1016/j.jamda.2018.02.002

16. Wrightson JG, Twomey R, Ross EZ, Smeeton NJ. The effect of transcranial direct current stimulation on task processing and prioritisation during dual-task gait. *Exp Brain Res*. May 2015;233(5):1575-83. doi:10.1007/s00221-015-4232-x
17. Ezzati A, Zammit AR, Habeck C, Hall CB, Lipton RB, for the Alzheimer's Disease Neuroimaging I. Detecting biological heterogeneity patterns in ADNI amnesic mild cognitive impairment based on volumetric MRI. *Brain Imaging Behav*. 2020/10/01 2020;14(5):1792-1804. doi:10.1007/s11682-019-00115-6
18. Varol E, Sotiras A, Davatzikos C. HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *Neuroimage*. 2017/01/15/ 2017;145:346-364. doi:<https://doi.org/10.1016/j.neuroimage.2016.02.041>
19. Whitwell JL, Dickson DW, Murray ME, et al. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: a case-control study. *The Lancet Neurology*. 2012;11(10):868-877. doi:10.1016/S1474-4422(12)70200-4
20. Zhang T, Koutsouleris N, Meisenzahl E, Davatzikos C. Heterogeneity of Structural Brain Changes in Subtypes of Schizophrenia Revealed Using Magnetic Resonance Imaging Pattern Analysis. *Schizophr Bull*. 2014;41(1):74-84. doi:10.1093/schbul/sbu136
21. Whitwell JL, Petersen RC, Negash S, et al. Patterns of Atrophy Differ Among Specific Subtypes of Mild Cognitive Impairment. *Archives of Neurology*. 2007;64(8):1130-1138. doi:10.1001/archneur.64.8.1130
22. Noh Y, Jeon S, Lee JM, et al. Anatomical heterogeneity of Alzheimer disease. *Neurology*. 2014;83(21):1936-1944. doi:doi:10.1212/WNL.0000000000001003
23. Blumen HM, Holtzer R, Brown LL, Gazes Y, Verghese J. Behavioral and neural correlates of imagined walking and walking-while-talking in the elderly. *Hum Brain Mapp*. Aug 2014;35(8):4090-104. doi:10.1002/hbm.22461
24. Rajpathak SN, Liu Y, Ben-David O, et al. Lifestyle factors of people with exceptional longevity. *J Am Geriatr Soc*. Aug 2011;59(8):1509-12. doi:10.1111/j.1532-5415.2011.03498.x
25. Verghese J, Wang C, Lipton RB, Holtzer R, Xue X. Quantitative gait dysfunction and risk of cognitive decline and dementia. *J Neurol Neurosurg Psychiatry*. Sep 2007;78(9):929-35. doi:10.1136/jnnp.2006.106914
26. Callisaya ML, Beare R, Phan TG, Chen J, Srikanth VK. Global and regional associations of smaller cerebral gray and white matter volumes with gait in older people. *PLoS One*. 2014;9(1):e84909. doi:10.1371/journal.pone.0084909
27. Srikanth V, Phan TG, Chen J, Beare R, Stapleton JM, Reutens DC. The location of white matter lesions and gait--a voxel-based study. *Ann Neurol*. Feb 2010;67(2):265-9. doi:10.1002/ana.21826
28. Doi T, Blumen HM, Verghese J, et al. Gray matter volume and dual-task gait performance in mild cognitive impairment. *Brain Imaging Behav*. Jun 2017;11(3):887-898. doi:10.1007/s11682-016-9562-1
29. Shimada H, Tsutsumimoto K, Lee S, et al. Driving continuity in cognitively impaired older drivers. *Geriatr Gerontol Int*. Apr 2016;16(4):508-14. doi:10.1111/ggi.12504
30. Wang N, Allali G, Kesavadas C, et al. Cerebral Small Vessel Disease and Motoric Cognitive Risk Syndrome: Results from the Kerala-Einstein Study. *Journal of Alzheimer's Disease*. 2016;50:699-707. doi:10.3233/JAD-150523
31. Beauchet O, Allali G, Launay C, Herrmann FR, Annweiler C. Gait variability at fast-pace walking speed: a biomarker of mild cognitive impairment? *J Nutr Health Aging*. Mar 2013;17(3):235-9. doi:10.1007/s12603-012-0394-4

32. Mielke MM, Roberts RO, Savica R, et al. Assessing the temporal relationship between cognition and gait: slow gait predicts cognitive decline in the Mayo Clinic Study of Aging. *J Gerontol A Biol Sci Med Sci*. Aug 2013;68(8):929-37. doi:10.1093/gerona/gls256
33. Buracchio T, Dodge HH, Howieson D, Wasserman D, Kaye J. The trajectory of gait speed preceding mild cognitive impairment. *Arch Neurol*. Aug 2010;67(8):980-6. doi:10.1001/archneurol.2010.159
34. Army U. Army individual test battery. *Manual of directions and scoring*. 1944;
35. Makizako H, Shimada H, Park H, et al. Evaluation of multidimensional neurocognitive function using a tablet personal computer: test-retest reliability and validity in community-dwelling older adults. *Geriatr Gerontol Int*. Oct 2013;13(4):860-6. doi:10.1111/ggi.12014
36. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol*. 1935;18(6):643.
37. Grober E, Buschke H. Genuine memory deficits in dementia. *Developmental Neuropsychology*. 1987/01/01 1987;3(1):13-36. doi:10.1080/87565648709540361
38. Verghese J, Annweiler C, Ayers E, et al. Motoric cognitive risk syndrome: multicountry prevalence and dementia risk. *Neurology*. Aug 19 2014;83(8):718-26. doi:10.1212/wnl.0000000000000717
39. Boulifard DA, Ayers E, Verghese J. Home-Based Gait Speed Assessment: Normative Data and Racial/Ethnic Correlates Among Older Adults. *J Am Med Dir Assoc*. Oct 2019;20(10):1224-1229. doi:10.1016/j.jamda.2019.06.002
40. Capistrant BD, Glymour MM, Berkman LF. Assessing mobility difficulties for cross-national comparisons: results from the World Health Organization Study on Global AGEing and Adult Health. *J Am Geriatr Soc*. Feb 2014;62(2):329-35. doi:10.1111/jgs.12633
41. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res*. 1982;17(1):37-49. doi:10.1016/0022-3956(82)90033-4
42. Depp CA, Jeste DV. Definitions and predictors of successful aging: a comprehensive review of larger quantitative studies. *Am J Geriatr Psychiatry*. Jan 2006;14(1):6-20. doi:10.1097/01.JGP.0000192501.03069.bc
43. BARTLETT MS. TESTS OF SIGNIFICANCE IN FACTOR ANALYSIS. *British Journal of Statistical Psychology*. 1950;3(2):77-85. doi:<https://doi.org/10.1111/j.2044-8317.1950.tb00285.x>
44. Kaiser HF. A second generation little jiffy. *Psychometrika*. 1970/12/01 1970;35(4):401-415. doi:10.1007/BF02291817
45. Verghese J, Lipton RB, Hall CB, Kuslansky G, Katz MJ, Buschke H. Abnormality of gait as a predictor of non-Alzheimer's dementia. *N Engl J Med*. Nov 28 2002;347(22):1761-8. doi:10.1056/NEJMoa020441
46. Kerminen H, Marzetti E, D'Angelo E. Biological and Physical Performance Markers for Early Detection of Cognitive Impairment in Older Adults. *Journal of Clinical Medicine*. 2024;13(3):806.
47. Kravatz NL, Ayers E, Bennett DA, Verghese J. Olfactory Dysfunction and Incidence of Motoric Cognitive Risk Syndrome: A Prospective Clinical-Pathologic Study. *Neurology*. Oct 24 2022;99(17):e1886-e1896. doi:10.1212/wnl.0000000000201030
48. Maguire FJ, Killane I, Creagh AP, Donoghue O, Kenny RA, Reilly RB. Baseline Association of Motoric Cognitive Risk Syndrome With Sustained Attention, Memory, and Global Cognition. *J Am Med Dir Assoc*. Jan 2018;19(1):53-58. doi:10.1016/j.jamda.2017.07.016

49. Allali G, Ayers EI, Verghese J. Motoric Cognitive Risk Syndrome Subtypes and Cognitive Profiles. *J Gerontol A Biol Sci Med Sci*. Mar 2016;71(3):378-84. doi:10.1093/gerona/glv092
50. Sekhon H, Allali G, Launay CP, Chabot J, Beauchet O. The spectrum of pre-dementia stages: cognitive profile of motoric cognitive risk syndrome and relationship with mild cognitive impairment. *Eur J Neurol*. Aug 2017;24(8):1047-1054. doi:10.1111/ene.13331
51. Shim H, Kim M, Won CW. Motoric cognitive risk syndrome is associated with processing speed and executive function, but not delayed free recall memory: The Korean frailty and aging cohort study (KFACS). *Arch Gerontol Geriatr*. Mar-Apr 2020;87:103990. doi:10.1016/j.archger.2019.103990
52. Popuri K, Ma D, Wang L, Beg MF. Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: Independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. *Hum Brain Mapp*. Oct 1 2020;41(14):4127-4147. doi:10.1002/hbm.25115
53. Kim M-J, Hong E, Yum M-S, Lee Y-J, Kim J, Ko T-S. Deep learning-based, fully automated, pediatric brain segmentation. *Scientific Reports*. 2024/02/22 2024;14(1):4344. doi:10.1038/s41598-024-54663-z
54. Mateos-Pérez JM, Dadar M, Lacalle-Aurioles M, Iturria-Medina Y, Zeighami Y, Evans AC. Structural neuroimaging as clinical predictor: A review of machine learning applications. *NeuroImage: Clinical*. 2018/01/01/ 2018;20:506-522. doi:<https://doi.org/10.1016/j.nicl.2018.08.019>
55. Jang J-W, Kim J, Park S-W, et al. Machine learning-based automatic estimation of cortical atrophy using brain computed tomography images. *Scientific Reports*. 2022/08/30 2022;12(1):14740. doi:10.1038/s41598-022-18696-6
56. Yang Z, Nasrallah IM, Shou H, et al. A deep learning framework identifies dimensional representations of Alzheimer's Disease from brain structure. *Nature Communications*. 2021/12/03 2021;12(1):7065. doi:10.1038/s41467-021-26703-z

Acknowledgements

The authors gratefully acknowledge the contributions of the participants, families, and study staff from all the studies in the consortium that provided the data for this research. We also thank the funding agencies and institutions that have supported these studies.

Conflicts of Interests

HMB serves as a consultant for Neural+. RBL is the Edwin S. Lowe Professor of Neurology at the Albert Einstein College of Medicine in New York. He also receives support from the Migraine Research Foundation and the National Headache Foundation and research grants from TEVA, Satsuma and Amgen. He serves on the editorial board of Neurology, senior advisor to Headache, and associate editor to Cephalalgia. He has reviewed for the NIA and NINDS, holds stock and stock options in Axon, Biohaven Holdings, CoolTech and Manistee; serves as consultant, advisory board member, or has received honoraria from: Abbvie (Allergan), American Academy of Neurology, American Headache Society, Amgen, Avanir, Axon, Axsome, Biohaven, Biovision, Boston Scientific, Dr. Reddy's (Promius), Electrocore, Eli Lilly, eNeura Therapeutics, Equinox, GlaxoSmithKline, Grifols, Lundbeck (Alder), Manistee, Merck, Pernix, Pfizer, Satsuma, Supernus, Teva, Trigemina, Vector, Vedanta. He receives royalties from Wolff's Headache 7th and 8th Edition, Oxford Press University, 2009, Wiley and Informa. JV serves as an Advisory Committee Member for MedRhythms, and a holds a uncompensated voluntary position at CatchU. Authors BTN, AE, KKP, EA, VGPK, SV, RB, OB, TD, HS, SM, and SA do not have any conflicts of interest to disclose.

Funding Sources:

This study was supported by NIH/NIA grants: 1R56AG057548-01, R01AG057548-01A1 and 2R01AG039330 (JV), NIH RO1 DK129320-01 2021-2026 (RB), NIA R01AG062659-01A1 (HB), NIA K23 AG063993, AG080635, AG003949, the Alzheimer's Association (SG-24-988292 ISAVRAD); Cure Alzheimer's Fund, the Leonard and Sylvia Marx Foundation (AE, RL), NIH/NIA 1K76AG083274 (SA). None of the funding sources had any role in the conduct of the analysis, interpretation of data and preparation of the article.

Consent Statement:

All human subjects across all the study locations provided informed consent.

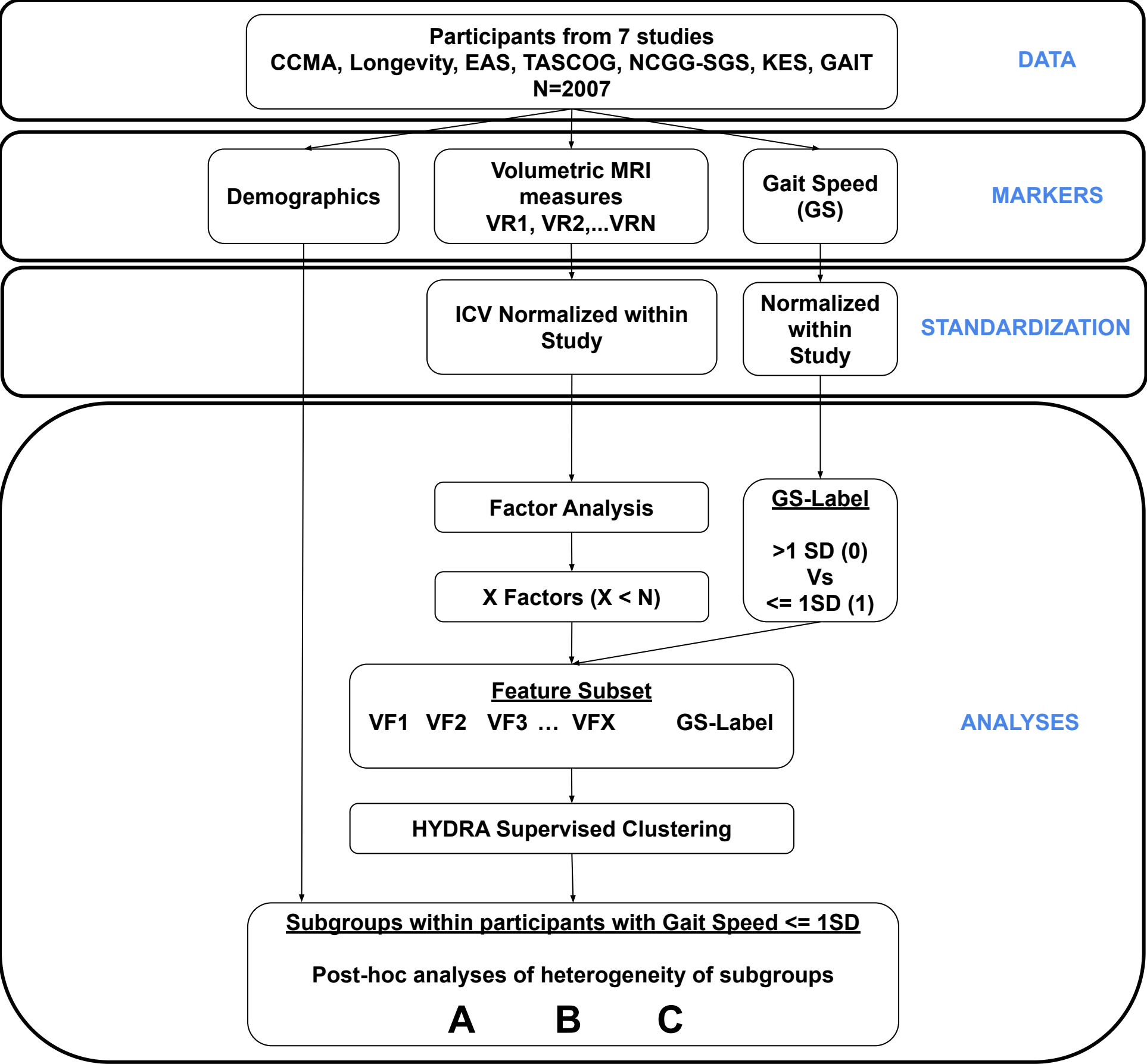
Keywords:

gait; Dementia; MCR; cognitive complaints; volumetric imaging; Machine Learning; weakly-supervised clustering

Figure 1: Study Plan - Plan of our study. CCMA: Central Control of Mobility in Aging Study in the US, the LonGenity study in the US, EAS : Einstein Aging Study in the US, TASCOG : Tasmanian Study of Cognition and Gait in Australia, NCGG-SGS: the National Center for Geriatrics and Gerontology–Study of Geriatric Syndromes in Japan, KES: Kerala-Einstein Study in India, and GAIT: the Gait and Alzheimer’s Interactions Tracking study; MRI: Magnetic Resonance Imaging; ICV: Intracranial Volume; SD: Standard Deviation;

Figure 2: Subgroup differences in MRI Volumes by regions - Group differences in MRI Volumes.: PreCen: Precentral Gyrus, EC=Entorhinal Cortex, ParsTri=Parstriangularis, BanksSTS=Banks of Superior Temporal Sulcus, PeriCal=Pericalcarine, Cau=Caudate Nucleus, CauACC=Caudal Anterior Cingulate, AccArea=Nucleus Accumbens, FRO=Frontal Cortex, PAR=Parietal Cortex, TEM=Temporal Lobe, OCC=Occipital Lobe, HIPP=Hippocampus, CEREB: Cerebellar Cortex, SUBCORT=Subcortical regions (Basal Ganglia nuclei, Amygdala, Thalamus)

Figure 3: Subgroup differences in MRI Volumes by regions - Group differences in Gait speed and Cognition. TMT-B : Trail Making Test, Part B, log-transformed, z-score normalized, available in CCMA, LonGenity, EAS, NCGG-SGS, Kes and GAIT. FR96: Free and Cued Selective Reminding Test (FCSRT), a recall test that uses either words or images. Scores includes the sum of free recall (FR) alone (range 0-48) and combined with cued recall as total recall (TR), FR96: the sum of FR and TR (range 0-96); available in CCMA, EAS and LonGenity; Stroop-INT : Stroop Color Word test, consisting of subtests measuring time required to name the colors seen (COLOR), read the given words (WORD) and name the color of the printed word (COLOR_WORD); Stroop-INT: the difference between the third subtask and the first task (COLOR_WORD - COLOR), available in EAS, TASCOG and GAIT.



Participants from 7 studies
CCMA, Longevity, EAS, TASCOG, NCGG-SGS, KES, GAIT
N=2007

DATA

Demographics

Volumetric MRI
measures
VR1, VR2,...VRN

Gait Speed
(GS)

MARKERS

ICV Normalized within
Study

Normalized
within
Study

STANDARDIZATION

Factor Analysis

X Factors (X < N)

GS-Label

>1 SD (0)
Vs
<= 1SD (1)

Feature Subset
VF1 VF2 VF3 ... VFX GS-Label

ANALYSES

HYDRA Supervised Clustering

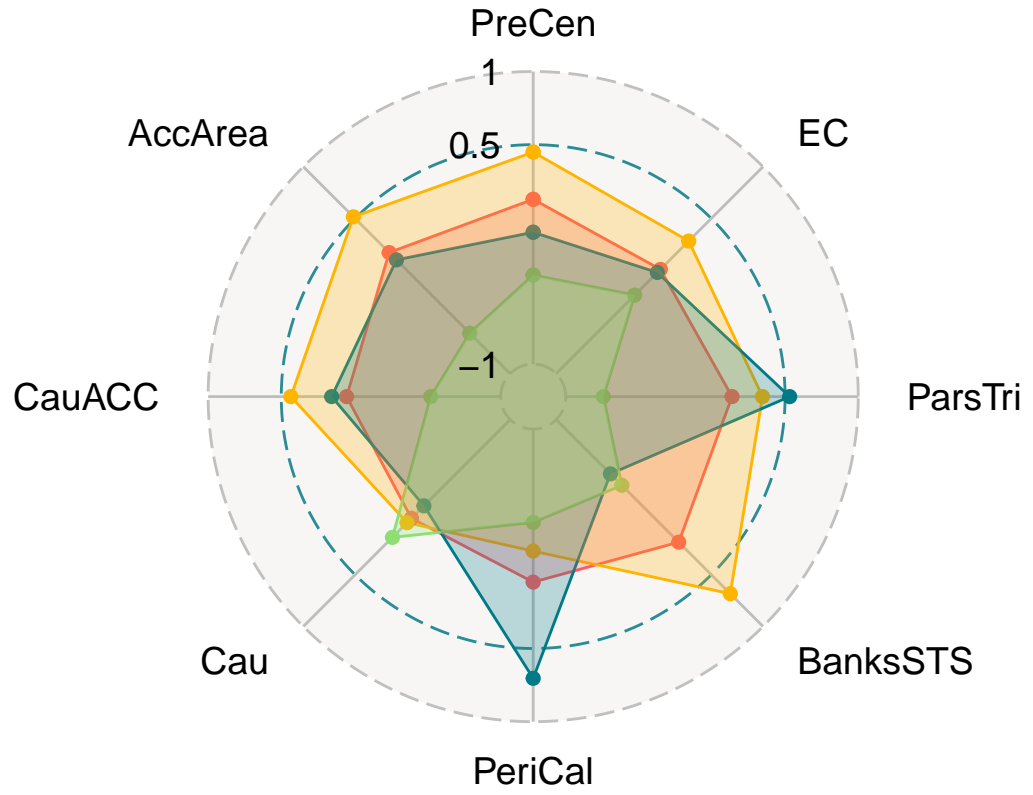
Subgroups within participants with Gait Speed <= 1SD

Post-hoc analyses of heterogeneity of subgroups

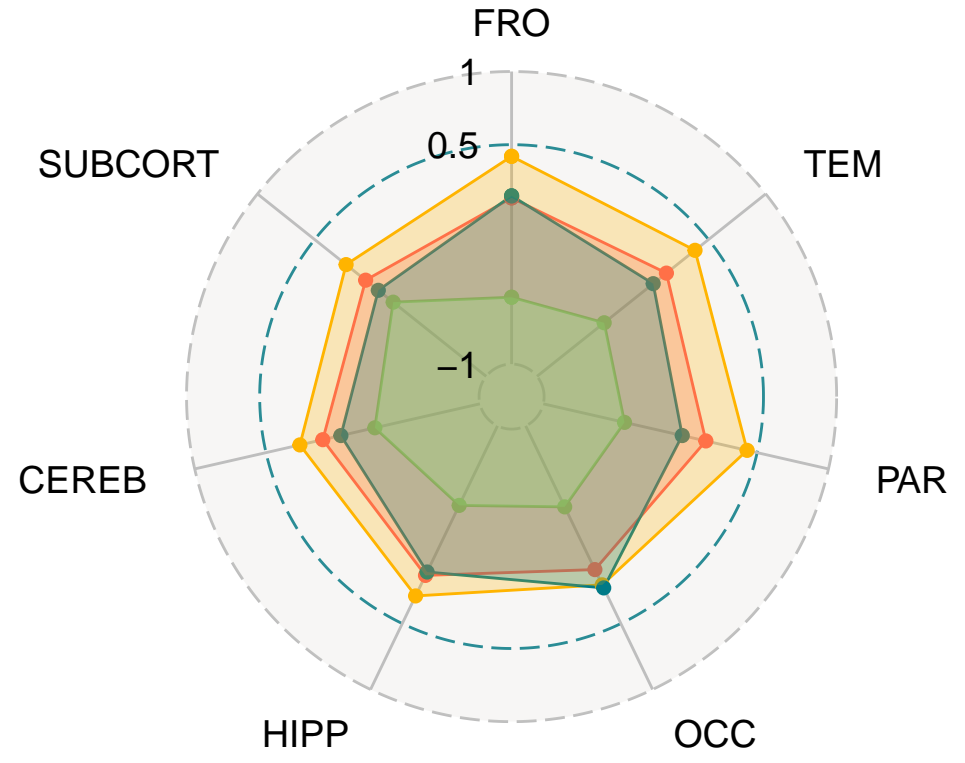
A B C

Atrophy in Volumetric MRI

A. ROIs from Factors



B. Cortical ROIs

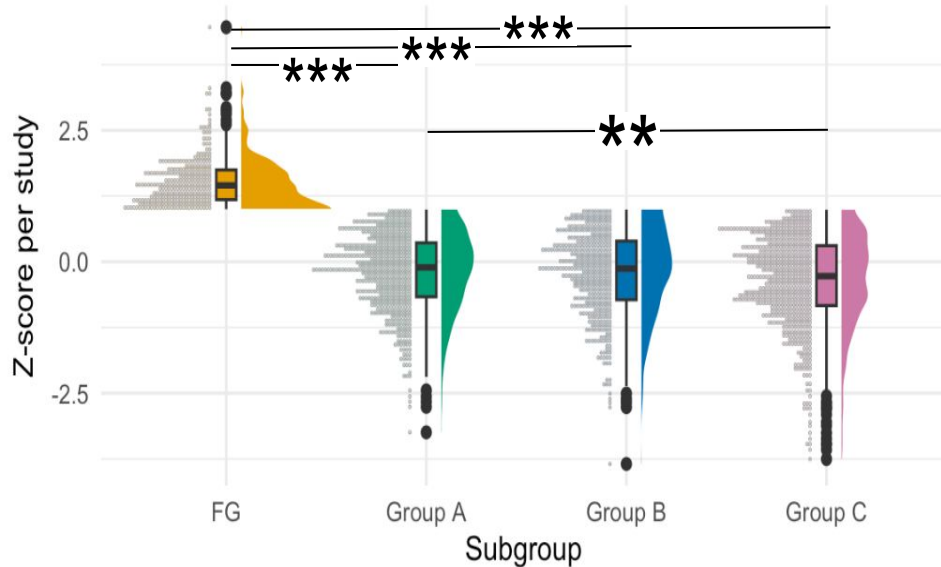


FG Group A Group B Group C

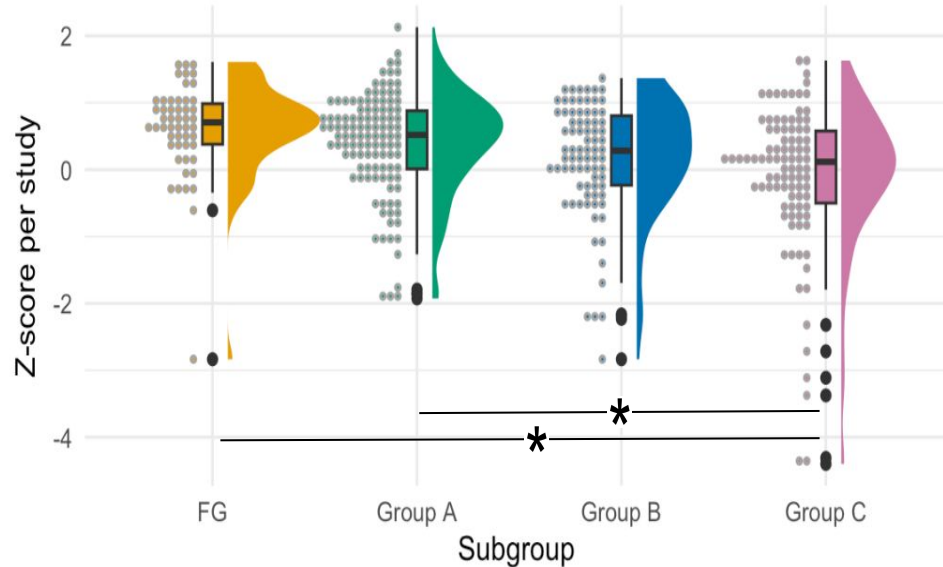
Group differences in Gait Speed & Cognition

group FG Group A Group B Group C

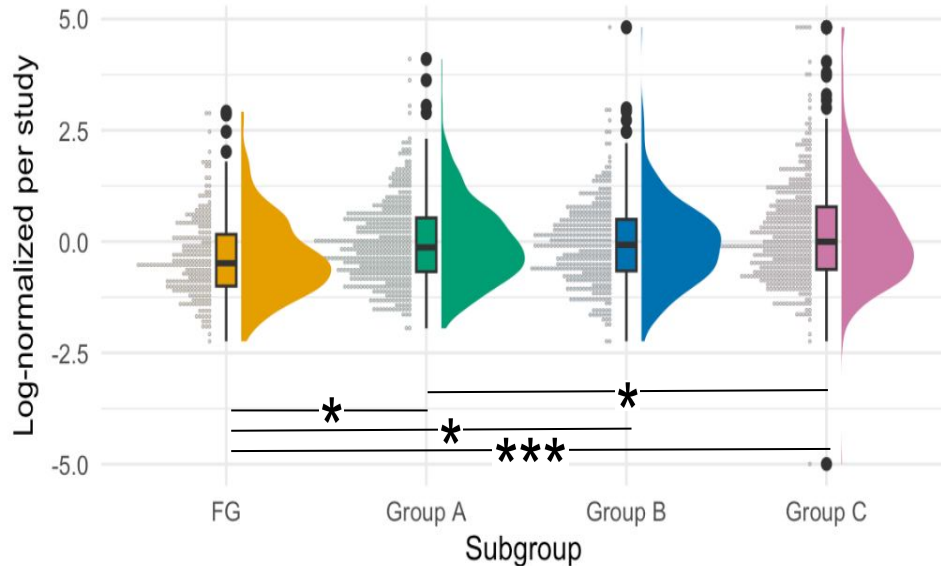
Gait Speed



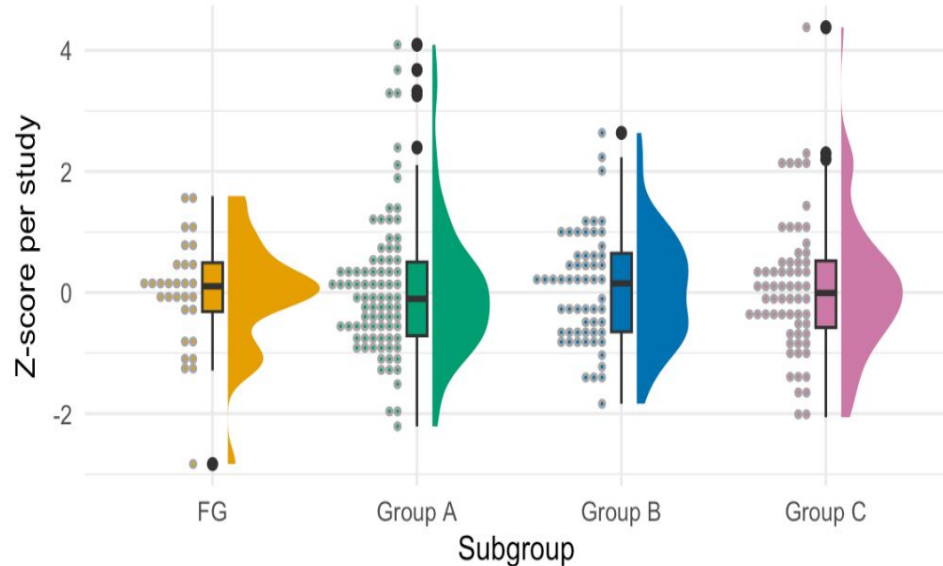
Free and Cued Recall - FR96



Trail Maling Test - B



Stroop Test - Interference



Title

Discovering Subtypes with Imaging Signatures in the Motoric Cognitive Risk Syndrome Consortium using Weakly-Supervised Clustering.

Bhargav Teja Nallapu¹, Ali Ezzati², Helena M. Blumen^{3,4}, Kellen K. Petersen⁵, Richard B. Lipton¹, Emmeline Ayers¹, V G Pradeep Kumar⁶, Srikanth Velandai^{7,8}, Richard Beare^{7,9}, Olivier Beauchet^{10,11}, Takehiko Doi¹², Hiroyuki Shimada¹², Sofiya Milman⁴, Sandra Aleksic⁴, Joe Verghese^{1,3}

¹Department of Neurology, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, New York, 10461, USA

²Department of Neurology, University of California, Irvine (UCI), Irvine, 200 S. Manchester Ave. Ste. 206, Orange, California, 92868, USA

³Department of Neurology, Stony Brook University, 101 Nicolls Road, Stony Brook, New York 11794, USA

⁴Departments of Medicine, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, New York, 10461, USA

⁵Department of Neurology, Washington University in St. Louis, 660 S. Euclid Ave, St. Louis, MO 63110, USA

⁶Department of Neurology, Baby Memorial Hospital, Indira Gandhi Road, Kozhikode, Kerala, 673004, India

⁷National Centre for Healthy Ageing, 2 Hastings Rd, Frankston VIC 3199, Australia

⁸School of Translational Medicine, Monash University, 99 Commercial Rd, Melbourne, VIC, 3004, Australia

⁹Developmental Imaging Group, Murdoch Children's Research Institute, 50 Flemington Rd, Parkville VIC 3052, Australia

¹⁰Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), 4565, chemin Queen Mary, Montréal QC H3W 1W5, Canada

¹¹ Department of Medicine, University of Montreal, C. P. 6128, succursale Centre-ville, Montréal, QC, H3C 3J7, Canada

¹²Department of Preventive Gerontology, Center for Gerontology and Social Science, Research Institute, National Centre for Geriatrics and Gerontology, 7-430, Morioka-cho, Obu City, 474-8511, Japan

Corresponding author:

Bhargav T. Nallapu

Email: bhargav.nallapu@einsteinmed.edu

Address:

Department of Neurology

Albert Einstein College of Medicine

Van Etten 3C12

1300 Morris Park Avenue

Bronx, NY 10461

Phone: 718-430-3896

Table 1: Study Characteristics

	All participants	FG	Group A	Group B	Group C	P value
N	1987	294	585	455	653	
Age, mean (SD), y	71.73(±6.87)	69.30(±5.52)	72.07(±6.80)	70.94(±6.77)	73.06(±7.18)	<0.001
Sex, Female (%)	948(47.7%)	139(47.3%)	277(47.4%)	229(50.3%)	303(46.4%)	0.625
Education, mean (SD), y[#]	12.30(±3.50)	12.73(±3.70)	12.27(±3.51)	12.39(±3.22)	12.08(±3.56)	0.08
Race/Ethnicity, N(%)						0.016
Asian	1146(57.7%)	176(59.9%)	329(56.2%)	282(62.0%)	359(55.0%)	
White	769(38.7%)	112(38.1%)	225(38.5%)	160(35.2%)	272(41.7%)	
Other*	59(3.6%)	2(2.0%)	27(5.3%)	9(2.8%)	21(3.3%)	
Gait Speed, cm/s	108.42(±22.85)	141.21(±11.52)	104.91(±17.05)	103.79(±18.28)	100.04(±21.34)	<0.001
MCR, Positive**, N(%)	117(6.6%,N=1778)	-	38(7.4%,N=515)	26(6.4%,N=405)	53(9%,N=588)	<0.001

Table 1: FG: Fast Gait, a subgroup of participants with gait speed greater than 1 standard deviation above mean; Groups A,B&C : gait speed lower than 1 SD above cohort specific mean; Subgroups obtained by the clustering model.

p-values reported for significant differences between FG and Subgroups;

[#] Education was not available for the GAIT cohort.

*Other in race/ethnicity included primarily Black (3.0% of total population), Hispanic Black, Hispanic White and others.

**MCR positive is defined as slow gait (as defined by a cohort specific threshold) and any cognitive complaint in Geriatric Depression Scale (GDS) or Assisted Daily Living (ADL) questionnaire.

Table 2: Subgroup differences in Brain Region volumes - Factors

Measure	FG	Group A	Group B	Group C	Pairwise Comparisons (Tukey)					
					Group A - FG	Group B - FG	Group C - FG	Group A - Group B	Group C - Group A	Group C - Group B
Regions_{Factor}										
Accumbensarea	0.434 (±0.126)	0.481 (±0.140)	0.425 (±0.106)	0.329 (±0.106)	***	0.92	***	***	***	***
Caudate	0.277 (±0.121)	0.283 (±0.108)	0.262 (±0.102)	0.301 (±0.151)	1	0.22	0.14	0.05	0.08	***
bankssts	0.488 (±0.129)	0.551 (±0.107)	0.404 (±0.101)	0.418 (±0.110)	***	***	***	***	***	***
caudalanteriorcingulate	0.465 (±0.126)	0.513 (±0.117)	0.478 (±0.113)	0.392 (±0.118)	***	0.34	***	***	***	***
entorhinal	0.508 (±0.097)	0.536 (±0.096)	0.506 (±0.090)	0.484 (±0.105)	***	0.96	0.55	***	***	0.14
parstriangularis	0.576 (±0.106)	0.599 (±0.103)	0.620 (±0.099)	0.476 (±0.080)	***	***	***	0.02*	***	***
pericalcarine	0.436 (±0.168)	0.403 (±0.134)	0.540 (±0.146)	0.372 (±0.140)	0.04*	***	***	***	0.003**	***
precentral	0.528 (±0.109)	0.565 (±0.106)	0.502 (±0.108)	0.468 (±0.113)	***	0.06	***	***	***	0.008**

Table 2: FG: Population with fast gait, gait speed greater than 1SD above population-mean; **Groups A,B&C:** result of HYDRA Clustering of remaining population (gait speed slower than 1 SD above population-mean); **P-value:** Significance Pr(>F) in ANCOVA corrected for age and sex; **Pairwise comparisons:** Multiple Comparisons of Means with Tukey Contrasts.

Regions_{Factor} Brain regions chosen from each of the factors from Factor Analyses, with one region with the highest loading within a factor. Precentral: Precentral Gyrus, entorhinal: Entorhinal Cortex, BanksSTS=Banks of Superior Temporal Sulcus, caudate: Caudate Nucleus
 *** P-values < 0.001

Table 3: Subgroup differences in Brain Region volumes – Cortical and Subcortical Regions

Measure	FG	Group A	Group B	Group C	Pairwise Comparisons (Tukey)					
					Group A - FG	Group B - FG	Group C - FG	Group A - Group B	Group C - Group A	Group C - Group B
C/S ROIs										
Frontal	0.300 (±0.030)	0.309 (±0.028)	0.300 (±0.027)	0.279 (±0.030)	***	0.47	***	***	***	***
Temporal	0.320 (±0.041)	0.330 (±0.038)	0.315 (±0.041)	0.296 (±0.042)	***	0.99	***	***	***	***
Parietal	0.312 (±0.032)	0.322 (±0.031)	0.306 (±0.029)	0.293 (±0.034)	***	0.47	***	***	***	***
Occipital	0.354 (±0.045)	0.359 (±0.042)	0.360 (±0.040)	0.333 (±0.043)	***	0.002**	***	0.78	***	***
Hippocampus	0.371 (±0.058)	0.381 (±0.057)	0.370 (±0.061)	0.339 (±0.059)	***	0.71	***	***	***	***
Cerebellum	0.404 (±0.046)	0.411 (±0.047)	0.398 (±0.046)	0.386 (±0.049)	***	0.79	0.01*	***	***	0.09
Subcortical[§]	0.321 (±0.033)	0.327 (±0.030)	0.317 (±0.033)	0.312 (±0.038)	***	0.86	0.66	***	***	0.98

Table 3: FG: Population with fast gait, gait speed greater than 1SD above population-mean; **Groups A,B&C:** result of HYDRA Clustering of the remaining population (gait speed slower than 1 SD above population-mean); **P-value:** Significance Pr(>F) in ANCOVA corrected for age and sex; **Pairwise comparisons:** Multiple Comparisons of Means with Tukey Contrasts.

C/S ROIs : Brain regions grouped into major cortical regions. [§]**Subcortical :** Includes Basal Ganglia nuclei, Amygdala and Thalamus.

*** P-values < 0.001

Table 4: Pairwise Subgroup Differences in Gait speed and Cognition

Measure	FG	Group A	Group B	Group C	P-value	Group A - FG	Group B - FG	Group C - FG	Group A - Group B	Group C - Group A	Group C - Group B
Gait Speed (Z-score)	1.531 (±0.466)	- 0.192 (±0.739)	-0.223 (±0.793)	-0.364 (±0.864)	***	***	***	***	0.58	0.003**	0.22
Cognition											
TMT-B	-0.354 (±0.892)	- 0.044 (±0.905)	-0.066 (±0.906)	0.174 (±1.153)	***	0.05*	0.03*	***	0.98	0.02*	0.09
FR96	0.599 (±0.716)	=	0.148 (±0.871)	-0.125 (±1.136)	0.005**	0.95	0.23	0.02*	0.31	0.01*	0.61
Stroop-INT	-0.025 (±0.918)	0.069 (±1.175)	0.061 (±0.923)	0.076 (±1.122)	0.99	-	-	-	-	-	-

Table 4: FG: Population with fast gait, gait speed greater than 1SD above population-mean; **Groups A,B&C:** result of HYDRA Clustering of the remaining population (gait speed slower than 1 SD above population-mean); **P-value:** Significance Pr(>F) in ANCOVA corrected for age and sex; **Pairwise comparisons:** Multiple Comparisons of Means with Tukey Contrasts.
TMT-B: Trail Making Test, Part B, log-transformed, z-score normalized, available in CCMA, LonGenity, EAS, NCGG-SGS, Kes and GAIT
Stroop-INT: Stroop Color Word test, consisting of subtests measuring time required to name the colors seen (**COLOR**), read the given words (**WORD**) and name the color of the printed word (**COLOR_WORD**); **Stroop-INT:** the difference between the third subtask and the first task (**COLOR_WORD - COLOR**), available in EAS, TASCOCG and GAIT
FR96: Free and Cued Selective Reminding Test (FCSRT), a recall test that uses either words or images. Scores includes the sum of free recall (**FR**) alone (range 0-48) and combined with cued recall as total recall (**TR**), **FR96:** the sum of **FR** and **TR** (range 0-96); available in CCMA, EAS and LonGenity
 *** P-values < 0.0001

Table 5: MCR Consortium Participant Subgroup Characteristics

Group	Gait speed, Imaging, and Cognition
Fast Gait (FG) Gait Speed > mean+1SD	Reference group, used by the clustering algorithm Best scores in FCSRT and TMT-B, No cortical or subcortical atrophy
Study Population Gait Speed < mean+1SD	
Group A	Fastest gait, largest volumes across almost all ROIs
Group B	Intermediate gait speed, similar cortical volumes as reference group
Group C	Slowest gait, lowest cortical volumes, highest prevalence of MCR

Table 5: FG: Population with fast gait, gait speed greater than 1SD above population-mean; **Groups A,B&C:** result of HYDRA Clustering of the remaining population (gait speed slower than 1 SD above population-mean); **P-value:** Significance $Pr(>F)$ in ANCOVA corrected for age and sex; **Pairwise comparisons:** Multiple Comparisons of Means with Tukey Contrasts. **SD:** Standard Deviation
TMT-B : Trail Making Test, Part B, log-transformed, z-score normalized, available in CCMA, LonGenity, EAS, NCGG-SGS, Kes and GAIT
FCSRT: Free and Cued Selective Reminding Test (FCSRT), a recall test that uses either words or images. Scores includes the sum of free recall (**FR**) alone (range 0-48) and combined with cued recall as total recall (**TR**), **FR96:** the sum of **FR** and **TR** (range 0-96); available in CCMA, EAS and LonGenity